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

NICOLAU OYHENARD DOS SANTOS

O3PO: a domain ontology for semantic interoperability for petroleum production plants

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

Advisor: Profa. Dra. Mara Abel

Porto Alegre
April 2023

CIP — CATALOGING-IN-PUBLICATION

Santos, Nicolau Oyhenard dos

O3PO: a domain ontology for semantic interoperability for petroleum production plants / Nicolau Oyhenard dos Santos. – Porto Alegre: PPGC da UFRGS, 2023.

77 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2023. Advisor: Mara Abel.

1. Applied ontology. 2. Semantic interoperability. 3. Petroleum. 4. Production plants. I. Abel, Mara. II. Título.

ACKNOWLEDGEMENTS

I am deeply grateful to my wife, Stefani, for her unwavering support and patience throughout this journey. She has been the foundation that keeps me grounded. My heartfelt appreciation also goes to my family for their constant encouragement.

The culmination of this work is the product of three years of diligent research, accompanied by uncertainties, challenges, and dedication. I express my profound gratitude to my supervisor and advisor, Prof. Mara Abel. Additionally, I extend my thanks to my colleague, Fabrício Rodrigues, for his invaluable input and insights during this time. I am grateful to all members of the PeTwin research group at UFRGS for their assistance and direction.

This research has been partially funded by CAPES (Funding Code 001), CNPq, and the PETwin Project (supported by FINEP and the LIBRA Consortium).

ABSTRACT

Integrating data from multiple sources remains a persistent challenge in various business industries, including the petroleum sector. The upstream petroleum sector's technological advancements have resulted in silos of information from various service providers, leading to a waste of time trying to locate data within disparate databases. To address this challenge, we present the Offshore Petroleum Production Plant Ontology (O3PO), a domain ontology designed to represent entities in offshore petroleum production plants. We have provided a clear and well-defined reference vocabulary to help professionals in the engineering and information technology fields label and link production plant monitoring, simulation measurements, and facilities. This ontology is based on a comprehensive compilation of industry-specific requirements, such as use cases and competency questions. Our ontology builds upon middle-level ontologies such as GeoCore and the Industry Ontology Foundry's (IOF) core ontology while adhering to the BFO top-level ontology. We have also sourced and integrated additional resources, including industry-specific glossaries and relevant ontologies, to construct this comprehensive domain ontology. This resulted in a robust domain ontology that offers universal concepts, defined classes, and relationships that can be applied in various domains. The validity of our domain ontology has been demonstrated through data analysis from an offshore oil field in Brazil, where we have successfully demonstrated its practical applications in a real-world setting.

Keywords: Applied ontology. Semantic interoperability. Petroleum. Production plants.

RESUMO

A integração de dados de múltiplas fontes continua sendo um desafio persistente em várias indústrias, incluindo o setor de petróleo. Os avanços tecnológicos na indústria de *upstream* resultaram em silos de informações de vários prestadores de serviços, levando a uma perda de tempo tentando localizar dados dentro de bancos de dados díspares. Para enfrentar este desafio, apresentamos a *Offshore Petroleum Production Plant Ontology* (O3PO), uma ontologia de domínio projetada para representar entidades em plantas de produção de petróleo offshore. Fornecemos um vocabulário de referência claro e bem definido para ajudar os profissionais das áreas de engenharia e tecnologia da informação a rotular e ligar o monitoramento de plantas de produção, medições de simulação e instalações. Esta ontologia é baseada em uma compilação abrangente de requisitos específicos da indústria, tais como casos de uso e questões de competência. Nossa ontologia se baseia em ontologias de nível intermediário, tais como GeoCore e a ontologia *core* da *Industrial Ontologies Foundry* (IOF) enquanto aderimos à ontologia de topo, BFO. Também temos recursos adicionais, incluindo glossários específicos da indústria e ontologias relevantes, para construir esta ontologia de domínio abrangente. Isto resultou em uma ontologia de domínio robusta que oferece conceitos universais, classes definidas e relações que podem ser aplicadas em vários domínios. A validade de nossa ontologia de domínio foi demonstrada através da análise de dados de um campo petrolífero offshore no Brasil, onde demonstramos com sucesso suas aplicações práticas em um cenário real.

Palavras-chave: Ontologia aplicada. Interoperabilidade semântica. Petróleo. Plantas de Produção.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Data management in the petroleum sector faces multiple challenges due to numerous businesses offering specialized services and using proprietary software. This creates a complex environment for managing field data across the supply chain. Consequently, substantial efforts have been made to establish industry standards addressing the digitalization needs of Industry 4.0. Some efforts in this way are the industry glossaries like the Professional Petroleum Data Management (PPDM) “What is a Well?” (Professional Petroleum Data Management Association (PPDM), 2014), and “What is a Completion?” (Professional Petroleum Data Management Association (PPDM), 2018). Also, there are initiatives to provide a semantic framework, such as the ontology and reference library of ISO 15926-4 (International Organization for Standardization (ISO), 2019), and a syntactic framework, such as the integrated data platform from the Open Subsurface Data Universe Forum (OSDU) (WHITLEY et al., 2020), data standards of Energistics Consortium, such as PRODML and RESQML standards, and the equipment specifications from CFIHOS (International Association of Oil & Gas Producers (IOGP), 2020a). Past projects have focused on centralizing data accessibility through two approaches: (1) defining a standard software architecture for data storage and access, and (2) establishing a standard data model supporting an integrated view. While both approaches are complementary and address complex aspects of data integration and interoperability, accessing integrated data and reasoning over it remains a challenge in the offshore environment due to various service companies, operators, and platform leasing companies utilizing different systems. Our project aims to provide a better solution to this problem.

Our project’s challenge is to create a framework for the semantic interoperability of data used by a digital twin of a petroleum production plant. Data supporting production operation planning and control in petroleum plants is often spread across multiple systems from different service companies performing specific operation tasks. These systems exchange data using proprietary or partially standardized formats. Integrated operation centers receive this data, labeled with source and meaning, and analyze it to support short-term decisions. Petroleum engineers use this data to produce simulations for medium to long-term operations planning and evaluate economic viability.

A digital twin provides a virtual mirror of a production plant to support simulation, prediction, and data analytics on production and facility maintenance data (MINERVA; LEE; CRESPI, 2020). Achieving a real-time, integrated view of a petroleum plant necessitates a uniform data view to help operators monitor oil flow and facilities' behavior in real-time. However, converting the vast array of data into a unified platform or format is a highly complex and labor-intensive task. Semantic tools that identify and describe data using a common descriptive vocabulary, with explicit descriptions of each term's meaning and logical restrictions, offer valuable contributions to the data integration problem. Many researchers have explored ontologies for this purpose due to their potential to improve communication in digital twin environments (KHARLAMOV et al., 2019a).

Ontologies serve as an excellent approach to conceptual modeling. They are a set of concepts, relations, attributes, instances, and axioms in machine-readable language that define entities' essential, invariant properties in a specific domain and provide explicit knowledge to computational systems(ARP; SMITH; SPEAR, 2015). Ontologies can verify similarity based on the congruence of meanings, serving as valuable references for terminology when individuals from diverse backgrounds collaborate. Additionally, ontologies can facilitate data integration by mapping elements in databases to concepts, attributes, and relations in ontologies, acting as semantic bridges to connect and integrate different databases. Petroleum operating companies can access ontology-based semantic tools in real-time to retrieve meaning, provenance, and restrictions for entities modeled in data applications.

We aim to develop a semantic-level model that bridges the gap between data and the simulation system, facilitating the operations of a digital twin. By establishing O3PO, we clarify relevant entities' meaning and logical restrictions within a well-structured domain ontology for production plant physical assets and related properties, known as Offshore Petroleum Production Plant Ontology. Our model refines the Basic Formal Ontology (BFO) (ARP; SMITH; SPEAR, 2015) and aligns with the GeoCore (GARCIA et al., 2020) and IOF-Core (SMITH et al., 2019) middle-level ontologies. In our domain analysis, we take into account the primary standard formats developed in the petroleum industry, which encompass the entities relevant to our application's scope.

The specific goal of O3PO is to provide a uniform, formal vocabulary referring to entities that pertain to an offshore petroleum production plant. The amount of data inherent in modern offshore enterprises and the large number of companies working together during field production motivate this goal. The scope of the ontology comprises a set of assets that are part of the oil path between the reservoir and the platform, including wells and subsea equipment. The ontology does not include topside equipment at this moment. This work includes a comprehensive exam of the literature on petroleum production systems, the design of an ontology using modern criteria, the debate and validation with industry experts, and the implementation of the conceptual model utilizing data from Libra field wells.

The following chapters detail the development process, the resulting ontology definition, and the application. Chapter 2 gives a theoretical background on ontologies and ontology engineering methodology. Chapter 3 provides a literature review of previous work on ontologies in the oil and gas domain, both from academia and industry. Chapter 4 presents the development methodology. Chapter 5 discusses the problem domain and scope, the specified requirements, and the resulting conceptual model; a few points are discussed as well. Chapter 6 presents some applications of ontology based on use cases acquired from the industry. Finally, Chapter 7 presents the conclusions and future work.

2 THEORETICAL BACKGROUND

As this thesis offers a domain ontology in the context of a digital twin for petroleum production as a key contribution, this chapter aims to deliver comprehensive overviews of relevant topics, including ontologies, digital twins, the Basic Formal Ontology (BFO), and the NeOn ontology development framework. These overviews will provide essential background knowledge and context for understanding the significance and implementation of the proposed domain ontology.

2.1 Ontologies

A famous paper by Nicola Guarino and associates published in 2009 discusses the definition of an ontology (GUARINO; OBERLE; STAAB, 2009). The first consideration of the authors is the distinction between the philosophical field of Ontology and computational ontology. In this work, we refer to ontologies in the latter sense. In this paper, he considers the definition of Studer, Benjamins and Fensel (1998), which combines the definitions from Gruber (1993) and Borst (1997) to form the definition "*An ontology is a **formal, explicit** specification of a **shared conceptualization**.*" This definition considers many aspects of an ontology:

- **formal**: the ontology should be machine-readable; this aspect restricts through logic the meaning of a particular concept and permits reasoning by a computer given a particular set of axioms.

- **explicit**: the types of terms and relations should be explicit (STUDER; BENJAMINS; FENSEL, 1998);

- **shared**: this particular aspect considered in (BORST, 1997) acknowledges that a conceptualization must be the result of a shared consensus instead of an individual conceptualization.

- **conceptualization**: the set of concepts about a particular domain.

A different definition of ontology is seen in (ARP; SMITH; SPEAR, 2015):

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

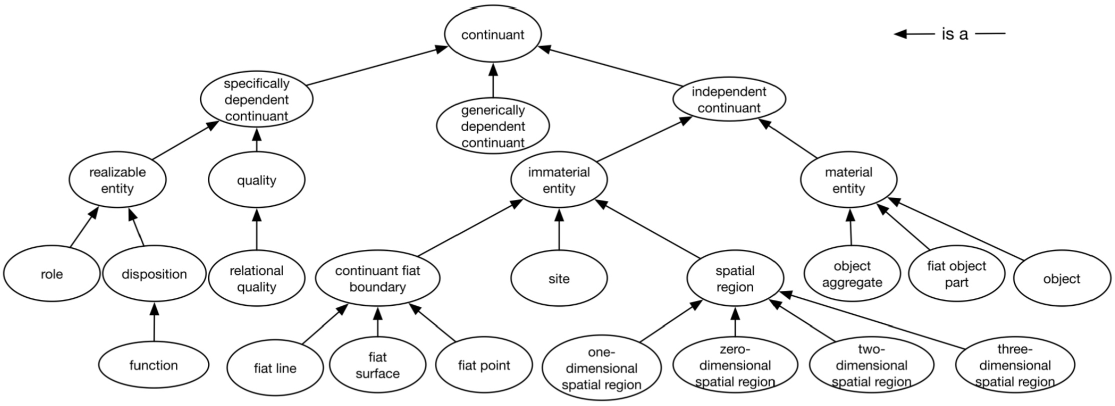


Figure 2.1 – Taxonomy of continuants in BFO. Taken from (OTTE; BEVERLEY; RUTTENBERG, 2022)

In this definition, "*universals*" define types of entities that exist in the world, i.e., color, car, person, football match; taxonomy refers to a hierarchy of types (e.g., universals) such as the one we see in Figure 2.1. Defined classes are entities with no corresponding universal demarcated by particular selection criteria, such as "human beings born in Brazil".

This definition doesn't consider terms referred to in (STUDER; BENJAMINS; FENSEL, 1998). This is because there is more than one view on how to represent the things in the world. The realist view considers that "*universals are mind-independent features of reality that exist only as instantiated in their respective instances*" (ARP; SMITH; SPEAR, 2015). In fact, this work assumes a realist view of the world due to a simpler approach and use of ontologies in related domains. However, due to the overall use of the latter definition and the important aspects of ontologies it shows, it is important to acknowledge this way of defining ontologies.

Now, that we have a general idea of what an ontology is, it is important to know what an ontology is not. We do that by differentiating ontologies from other types of knowledge artifacts. Figure 2.2 shows the variation in the degree of complexity when comparing taxonomies, thesaurus, conceptual models, and logical theories.

Ontologies, thesauri, glossaries, taxonomies, and conceptual models are all tools used in information science and knowledge management. In the context of semantic complexity, these tools - ontologies, thesauri, glossaries, taxonomies, and conceptual models - have different roles (Semantic Web Company, 2014). Ontologies describe concepts or types of things that exist. Thesaurus and glossaries describe terms; they are language resources. The information they represent is very different from ontologies, although they have relationships, of course. Terms in a glossary can label concepts in ontologies, for

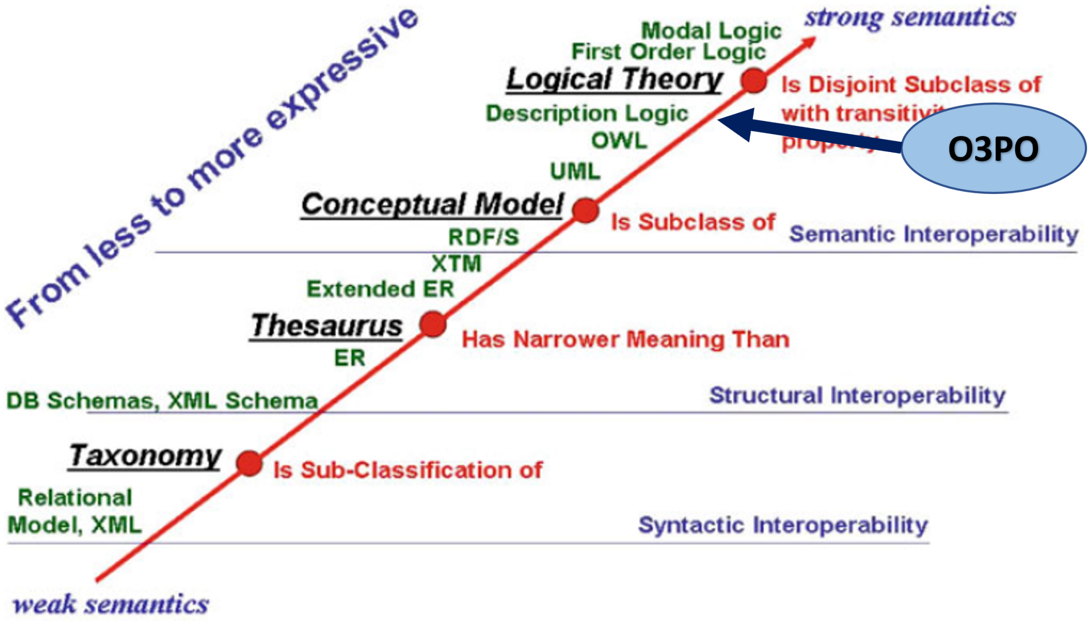


Figure 2.2 – Different degrees of semantic complexity in knowledge artifacts. Modified from (OBRST, 2010).

example.

Ontologies help to represent and organize knowledge in a structured and formal manner, which can be particularly useful in handling complex information and performing reasoning over data; they are on the same level as logical theories seen in Figure 2.2. Thesauri are controlled vocabularies of synonyms, related terms, and hierarchical relationships that describe concepts in a particular domain or language. Glossaries are lists of terms and natural-language definitions that explain the meaning and usage of specialized vocabulary in a particular field or domain. Taxonomies are hierarchical classifications of objects or concepts based on their characteristics, properties, or relationships. Usually, an ontology contains a taxonomy of classes and sub-classes as a backbone (REES, 2003). Finally, conceptual models are abstract representations of a system or process that capture its essential features and relationships.

But why are they important? Ontologies are essential for providing a shared understanding, enabling machine interpretation of information, facilitating knowledge management, and ensuring interoperability between different systems and applications. They help to organize and represent knowledge in a structured and systematic way, eliminating ambiguity and guiding the development of new knowledge. Overall, ontologies can be used in a wide range of applications where structured knowledge representation and sharing are required, enabling effective communication and collaboration between humans and machines, including:

- Data integration: Ontologies can integrate data from multiple sources by providing a common vocabulary and conceptual framework. An example of this would be in the healthcare industry, where ontologies can be used to integrate patient data from different hospitals, labs, and medical databases.

- Information retrieval: Ontologies can help improve search results accuracy and relevance by providing a structured representation of knowledge. This is useful particularly in systems such as an e-commerce platform, where an ontology can be used to improve product search results. By using a structured representation of knowledge about product categories, attributes, and relationships, the search engine can more accurately understand user queries and return relevant results.

- Knowledge management: Ontologies can organize and manage knowledge within an organization or community, facilitating knowledge sharing and reuse. An example of this would be in a large corporation, where an ontology can be employed to organize and manage internal knowledge across departments. For example, an ontology could capture information about company policies, employee roles and responsibilities, projects, and industry standards. This structured representation of knowledge facilitates easy sharing and reuse across the organization, allowing employees to quickly locate relevant information and collaborate more effectively.

- Decision support: Ontologies can provide decision support by encoding expert knowledge and providing automated reasoning capabilities. Integrating data with a machine-readable artifact such as an ontology enables full access to relevant data and supports decision-making.

- Natural language processing: Ontologies can be used to improve natural language processing by providing a semantic framework for understanding language. An example from an information retrieval context would be extracting tokens from natural language queries and mapping to types in an ontology to enrich the information retrieval process and deliver more accurate results.

- Machine learning: Ontologies can be used to improve machine learning algorithms by providing a structured representation of knowledge for training and inference. Annotating the attributes in the dataset for training results from a machine-learning algorithm could provide more meaning to the user.

- Standardization: Ontologies can provide a common standard for data exchange between different systems and applications. Considering a set of common functional requirements from a set of applications, a consensual model of types and relations of the domain could serve as a base for later data schemas generation that could serve different applications.

Considering the level of granularity in an ontology, there are three layers to ontological architectures (OBRST, 2010):

- Upper-level ontologies: sometimes called foundational or top-level ontologies, are domain-independent and high-level ontologies that serve as a root for lower-level ontologies (ARP; SMITH; SPEAR, 2015). Its purpose is to provide universal and fundamental concepts applicable across multiple domains. Such an ontology often represents general concepts such as object, event, quality, etc (PRESTES et al., 2013). Examples of such ontologies would be the Basic Formal Ontology (BFO) (OTTE; BEVERLEY; RUTTENBERG, 2022) and the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (BORGO et al., 2022).

- Middle-level ontologies: link the high-level, abstract concepts established in an upper ontology and the detailed, specific concepts featured in a domain ontology. Even though ontology mapping can be carried out at any level, the upper and middle-level ontologies strive to simplify the cross-domain mapping process (OBRST, 2010). In addition, core ontologies can be categorized as middle-level ontologies that sit between top-level and domain ontologies. They incorporate concepts outlined in top-level ontologies and introduce new concepts that can be utilized in specific tasks and domains. Examples of core ontologies can be seen in (PRESTES et al., 2013; GARCIA et al., 2020).

- Domain-level ontologies: define concepts in a particular domain (PRESTES et al., 2013), often use higher-level concepts from middle-level or top-level ontologies to anchor definitions and increase ontology reuse, and interoperability with other ontologies in similar domains (ARP; SMITH; SPEAR, 2015). Such type of ontologies are present in a broad range of domains (WOODS et al., ; SANTOS et al., 2022; ASHBURNER et al., 2000).

A second dimension for ontology classification relates to the potential use of the ontology by others (MENZEL, 2003):

- Application ontologies: these are domain-specific ontologies designed to capture the concepts and relationships within a particular domain or application. Application ontologies share and reuse information and knowledge within a particular community of practice and support specific tasks or applications. They are often developed by subject matter experts in a particular field and are intended to be used by practitioners within that field.

- Reference ontologies: these ontologies are designed as a reference point or standard within a particular field or community of practice. Reference ontologies are often general and are intended to provide a common vocabulary or framework for communication and collaboration within a domain. They are typically developed by a community of experts and designed to be a foundation for developing more specific application ontologies.

2.2 Digital Twins

Digital twins are virtual clones of real assets, processes, or systems capable of mirroring their behavior, simulating the system outcome based on previous data and possible actions, controlling the asset directly from the twin, and accurately predicting overall performance (GRIEVES; VICKERS, 2017). Integrating data from various sources, such as wells, reservoirs, flowlines, and equipment, is challenging when applying this technology to petroleum production systems.

The demand for AI and digital technologies in all fields of petroleum engineering has increased due to the requirement for lower-carbon output and greater efficiency. Such technologies, such as data analytics and cyber-physical systems, can help to make petroleum exploration and production more environmentally friendly and modern, according to (MATHIESON; MEEHAN; POTTS, 2019).

Considering academic research on its industrial applications, the topic of digital twins has been on the rise (TAO et al., 2019). One of the pillars of the so-called “Sector 4.0” is the idea of digital twins changing the industry.

Different definitions have been given to digital twins, but Tao and Zhang (2017)’s proposal adequately describes the notion. As shown in Figure 2.3, it comprises a five-dimensional framework, a physical entity, a virtual entity, services, data, and connections. PE denotes a physical entity; VE denotes virtual equipment; SS denotes services for physical and virtual entities (PE and VE); DD denotes digital twin (DT) data. CN denotes the link between PE, VE, Ss, and DD.

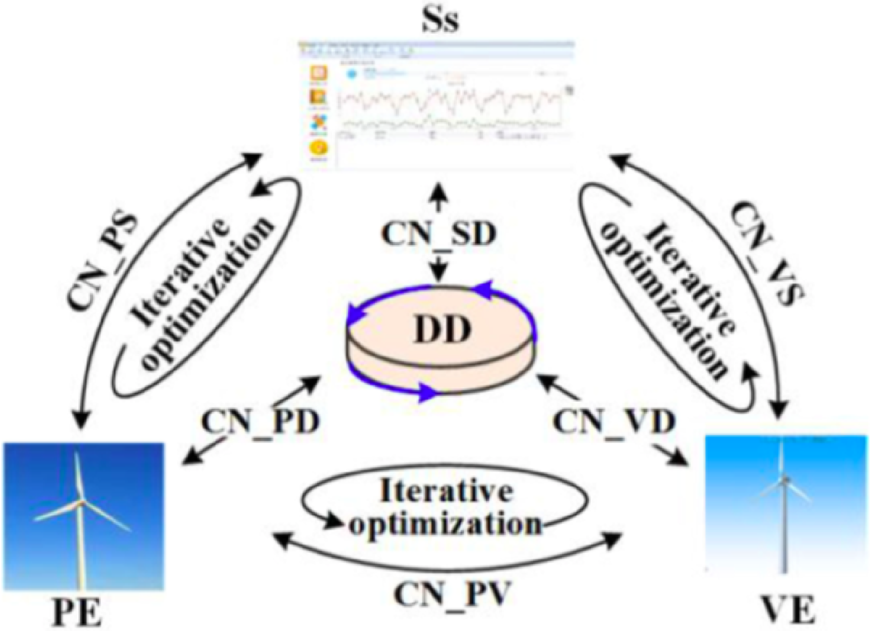


Figure 2.3 – Five-dimensional concept of a digital twin (TAO et al., 2018).

Applying a digital twin to an item, a system, or even an organization allows for the entity's simulation while allowing for the continuous flow of data from sensors in the actual system. The digital twin predicts behavior, accomplishes performance analysis, and gets faster and more accurate information, all of which are critical for decision-making.

The work of (TAO et al., 2019) presents a current evaluation of the state-of-the-art-related digital twins and their use in many industries is presented.

The creation of an intelligent control system for distributed industrial objects resulted from installing a digital twin in (PONOMAREV; KUDRYASHOV; POPELNUKHA, 2017).

Kritzinger et al. (2018) created a literature analysis that classified the applications of the digital twin idea based on the level of integration between the real and virtual entities. When it comes to the whole relationship between physical and virtual things, the authors found that there is insufficient research on digital twins. The three methods of integration proposed by the authors are shown in Figure 2.4: digital models, digital shadows, and digital twins, respectively.

The notion of digital twins was observed in a white paper in 2004 (GRIEVES, 2014). Aerospace was the first industry to benefit from the new technology, as evidenced by the first publication that coined the term "digital twin", "an integrated multi-physics,

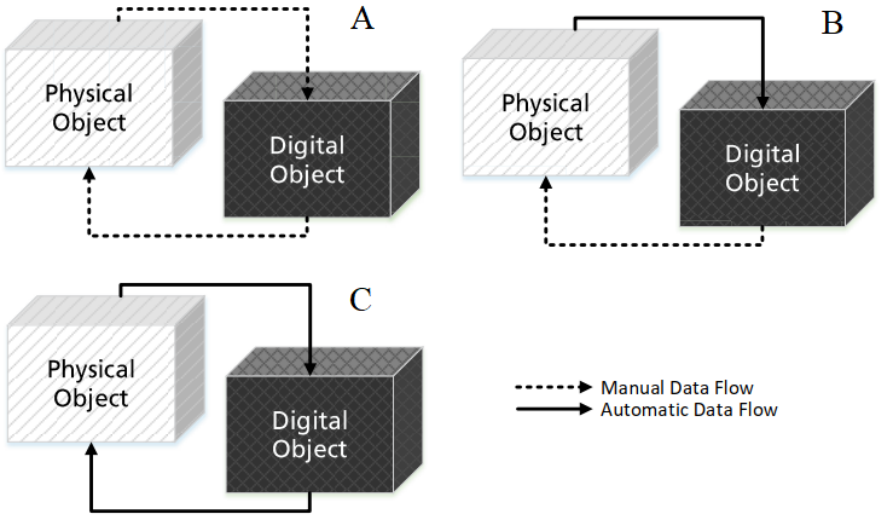


Figure 2.4 – (a) Digital model, (b) Digital shadow, and (c) Digital twin according to the level of integration between the physical and virtual entities. Modified from Kritzinger et al. (2018).

multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin", according to NASA (SHAFTO et al., 2012). According to NASA's definition, a digital twin is realistic and may consider one or more crucial and interconnected systems.

Following the aerospace community's original examination of the digital twin concept, the manufacturing industry began investigating the potential benefits of having a digital twin of production systems to boost overall productivity. The work of (NEGRI; FUMAGALLI; MACCHI, 2017) presented a review of research conducted between 2012 and 2016 that introduced the term "Industry 4.0" in 2011. Negri, Fumagalli and Macchi (2017) demonstrated the importance of semantic meta-data models in integrating simulations from several disciplines and predicting overall system behavior.

Why should businesses use digital twins, and how should they begin the transition? Voell et al. (2018) looks at the business side of digital twins and proposes an organic strategy for integrating them into businesses through experimenting, testing, piloting, and scaling.

Souza et al. (2019) presented the first design for digital twins based on the Industrial Internet of Things (IIOT) and the Open Platform Communication Unified Architecture (OPC UA) standard to combine data interchange from many devices in the physical twin. Receiving and transmitting data and commands in XML format was the responsibility of an internal server.

2.3 Basic Formal Ontology (BFO)

One of the most used top-level ontologies is the Basic Formal Ontology (BFO) (ARP; SMITH; SPEAR, 2015). The ontology shown in this work uses BFO as a top-level ontology. Additionally, all imported ontologies to produce O3PO use BFO as a top-level ontology. It commits to some ontology design principles (OTTE; BEVERLEY; RUTTENBERG, 2022):

- Ontological Realism: this principle states that the objective of an ontology is to represent reality, and not mental representations.
- Fallibilism: the ontology requires continuous updates when new scientific discoveries are met.
- Adequatism: an opposite tendency from reductionism, the representations of entities in a particular domain should not be reduced to some other entities in another domain deemed more important (OTTE; BEVERLEY; RUTTENBERG, 2022). All entities within a domain should be recognized and respected for their uniqueness, with room made in our understanding of reality for each type of entity across various levels of detail(ARP; SMITH; SPEAR, 2015).

One of the basic distinctions that BFO considers is between universals and instances. Universals are the kinds of entities that are repeated in reality across time and space (OTTE; BEVERLEY; RUTTENBERG, 2022). They are general concepts or categories that can apply to many specific instances, while instances are specific examples or occurrences of those concepts. Universals are often represented as classes or types of things, such as “animal”, “plant”, “car”, “person”, etc. Instances, or particulars, are entities in reality that cannot be repeated, such as “my computer” being an instance of the universal “computer”. They have unique identities, properties, and relationships that distinguish them from other instances of the same or different universals.

A second distinction present in BFO is between continuants and ocurrents. Figure 2.5 shows the taxonomy of continuants in BFO. A continuant is an entity or object that exists over time and persists through change, such as a person, a book, or a tree. Continuants have a relatively stable identity and maintain their properties and relationships over time, even though they may undergo physical or functional changes. On the other hand, an occurrent is an event, process, or activity that occurs in time and may involve one or more continuants. Occurrents are transient and have a duration, a location, and a set of

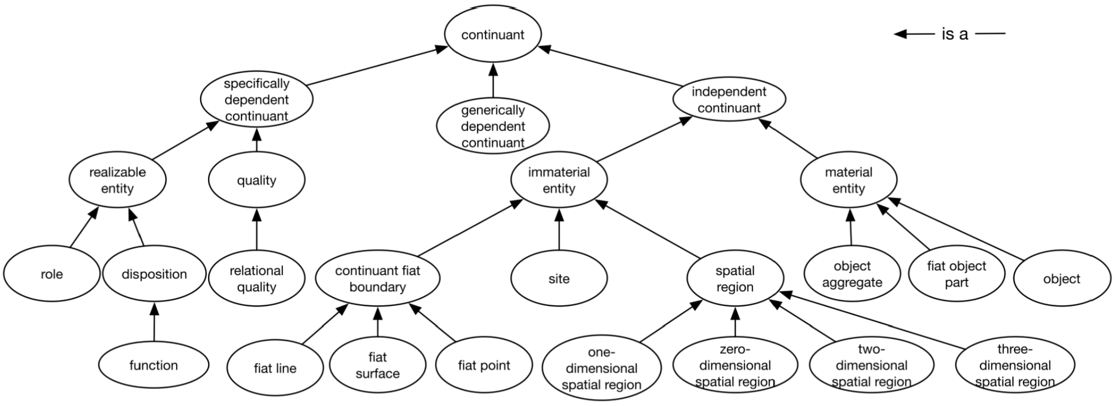


Figure 2.5 – BFO's *is_a* hierarchy. Taken from (SMITH et al., 2015).

causal or temporal relationships with other entities or events. Examples of occurrents include running, reading, or falling. The distinction between continuants and occurrents is important in ontology because it helps to differentiate between things that exist over time and things that happen in time, and to capture the dynamic and temporal aspects of the world.

Continuants are divided in independent and dependent continuants. An entity that can exist independently or as part of another entity is known as an independent continuant. Examples of independent continuants include a person, a ball, a pencil, a plant, a city, and the sky. It's worth noting that an independent continuant that is a part of another object, such as a heart, is still considered an independent continuant as it can be detached from the object and continue to exist on its own.

On the other hand, dependent continuants in Basic Formal Ontology (BFO) have a relationship of existential dependence with their bearers, meaning that they rely on the existence of some independent entity in order to exist. A dependent continuant that relies on one or more specific independent continuants for its existence is known as a specifically dependent continuant. These entities exhibit existential dependence, which means that in order for them to exist, another entity in which they are inherent must also exist, and that entity is typically more concrete. Examples of specifically dependent continuants include the color of an orange, the pain in one's left knee, the mass of a mouse, the smell of a perfume, the disposition of a fish to decay, one's role as a doctor, the function of the heart to pump blood, and the quality of a specific pixel array on a screen. A specifically dependent continuant can be a quality or a realizable entity. Qualities like roundness or redness are examples of specifically dependent continuants that require an independent continuant, such as a ball or a clown's nose, to exist. The mass of a mouse, for instance, could not exist without the mouse itself, and the color of an orange could not exist without the orange (ARP; SMITH; SPEAR, 2015). Realizable entities are specifically dependent continuants that can be manifested or actualized in corresponding events or processes and may exhibit periods of dormancy. Examples of realizable entities include the function of the separator to separate fluid that comes from the well, the role of being a student, and the disposition of saltwater to conduct electricity (ARP; SMITH, 2008).

Differently from a specifically dependent continuant that cannot exist unless an independent continuant exists, some dependent continuants appear capable of migration, such as when a digital file (e.g., this thesis) is copied from one computer to another. While a digital file depends on a bearer, there must be a physical storage device on which it is saved for the file to exist; it can still be transferred from one storage device to another. The same digital file can be saved to multiple storage devices and exist in multiple identical copies. We call this type of entity generically dependent continuants (ARP; SMITH; SPEAR, 2015).

Independent continuants can be either material entities or immaterial entities. Material entities have physical matter as a part of them and are localized in space. Examples of material entities include a person, a football team, a mountain, and a house. Immaterial entities, on the other hand, "*have no material entities as parts*" (ARP; SMITH; SPEAR, 2015). Examples of immaterial entities include the interior of your stomach that is a part of your body, the interior of a cave, or the annular space between the tubing and the inner walls of casing inside a well (SMITH, 2012).

Material entities can be objects, object aggregates, and fiat object parts (SMITH, 2012). Objects are material entities that are spatially extended in three dimensions, causally unified and maximally self-connected (ARP; SMITH; SPEAR, 2015). Examples would be a book, a computer, or a person. Object aggregates are material entities that are composed of a collection of objects, each of these objects is part of the object aggregate. Examples would be a band, a collection of coins, or the collection of students during a class. A fiat object part refers to a physical entity that is a constituent of a larger object but is not separated from the remainder of the object by any physical breaks or boundaries. The handle of a fork, the Northern hemisphere of our planet, and the "division of the brain into regions" (SMITH, 2012).

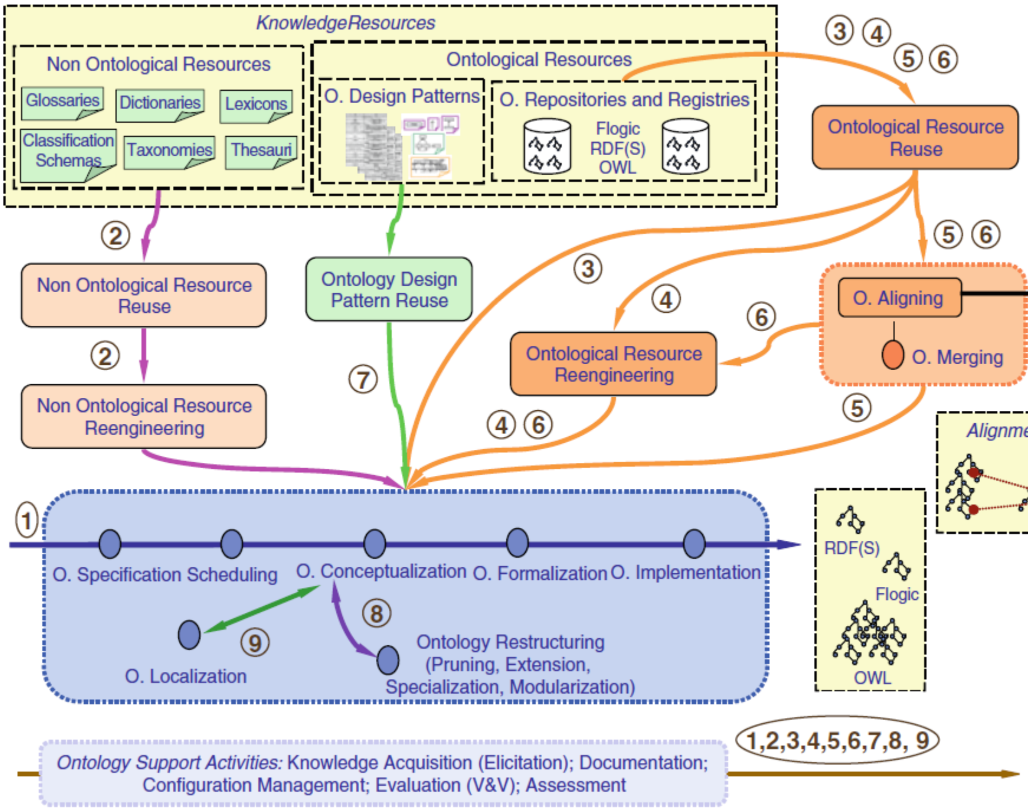


Figure 2.6 – Overall methodology process considering different scenarios in NeOn. Taken from (SUÁREZ-FIGUEROA et al., 2012)

2.4 NeOn ontology development framework

A recent review of ontology engineering methodologies (TUDORACHE, 2020) shows that one of the most comprehensive ontology engineering methodologies is a result of the NeOn Project (2006- 2010). Considering the organization of activities provided by their approach to ontology development, we followed the NeOn methodology (SUÁREZ-FIGUEROA, 2010) in the development of O3PO. Among the scenarios included in NeOn, this research combines scenarios where non-ontological and ontological resources will compound the resulting ontology.

This methodology considers nine scenarios of ontology development. Depending on the scenario that a possible user might have, the methodology prescribes a different sequence of activities. Figure 2.6 shows the overall methodology process considering different scenarios.

Below is the NeOn list of scenarios:

1. From specification to implementation,

2. Reusing and re-engineering non-ontological resources,
3. Reusing ontological resources,
4. Reusing and re-engineering ontological resources,
5. Reusing and merging ontological resources,
6. Reusing, merging, and re-engineering ontological resources,
7. Reusing ontology design patterns (ODPs),
8. Restructuring ontological resources, and
9. Localizing ontological resources.

Considering these nine scenarios included in NeOn, this research combines scenarios 2 and 3 and applies non-ontological and ontological resources to build the ontology. Like all the other scenarios covered by NeOn, this research follows the sequence of activities of the first scenario, from specification to implementation, with some alterations. The sequence of activities prescribed for the combination of scenarios 2 and 3 is:

1. Specification: this activity involves defining the requirements and goals of the system being developed. It includes identifying the scope, target users, and desired functionalities. This stage is critical for determining the purpose and direction of the project.

2. Scheduling: involves planning and organizing the tasks and resources required for the project. This includes allocating time for each activity, setting deadlines, and estimating the overall timeline. It ensures that the project progresses in a structured and timely manner.

3. Non-ontological resource reuse: this activity involves identifying and reusing existing non-ontological resources, such as databases, knowledge bases, and software components. By leveraging these resources, the development process can be accelerated and the overall quality of the system can be improved.

4. Non-ontological resource reengineering: in cases where existing non-ontological resources do not fully meet the project requirements, they may need to be reengineered. This activity involves adapting or modifying these resources to better fit the needs of the project.

5. Ontological resource reuse: similar to non-ontological resource reuse, this activity involves identifying and reusing existing ontologies or ontology modules. By reusing established ontologies, the development process can benefit from the collective knowledge and experience of the ontology engineering community.
6. Conceptualization: the process of creating a high-level, abstract representation of the knowledge domain. It includes identifying the main concepts, their properties, and their relationships. This activity lays the foundation for the ontology and helps ensure that it accurately represents the domain.
7. Formalization: formalization involves translating the conceptual model into a formal representation, such as an ontology language like OWL or RDF. This step is essential for enabling machine reasoning, interoperability, and automated processing of knowledge.
8. Implementation: the implementation activity covers the actual development of the knowledge-based system, including integrating the ontology with the application logic, building user interfaces, and ensuring proper functionality. This stage also involves testing, debugging, and refining the system to ensure that it meets the requirements specified in the first stage.

3 LITERATURE REVIEW

In this chapter, we present an analysis of the literature review. We start with a description of the problems related to data interoperability. Then we describe digital twins, focusing on how they can treat data interoperability problems.

Petroleum production consists of a sequence of tasks that begins with the drilling or definition of production and injection wells and the installation of the facilities that will support the operation. The whole operation will produce data supporting petroleum engineers' decisions about the well flow or water and gas injection and equipment maintenance. The operation and economy of a field are controlled by a large set of parameters captured from different sources by distinct professionals in engineering studies and should be integrated to produce predictive models.

Interoperability is the capability to offer to the user uniform access over data and information produced and managed by a distinct information system, allowing unified data analysis (WACHE et al., 2001) Interoperability enables disparate information systems to speak with one another, sharing data and information to create predictive models without requiring the integration of data models or applications. We can search and reach interoperability using different strategies for system and data integration:

(a) Applications can reach the same data repository, taking care of converting data format and extracting the meaning of the data.

(b) Applications can retrieve data from several data sources that share data in standard formats.

(c) One application can feed another application using a common format/standard.

(d) Applications can share standard servers, modules, and data objects.

(e) Full applications can play the role of data integrators, with the support of software vendors.

Any of these strategies requires the consideration and treatment of the semantics of the data that the applications exchange. The transfer and clarification of the data shared between applications are still the main bottleneck for interoperability.

We claim that formal ontologies can be used to make explicit the intended meaning of the engineering concepts embedded in the conceptual models and into the metadata of the databases that support the applications in production control and prediction. This ontology can then be used for documenting the business processes and implementing systems integration solutions based on clear conceptual models and entity definitions.

3.1 Contributions from industry

This section describes the main initiatives of petroleum associations and industries for providing standards to guide integration and interoperability tasks for petroleum production.

3.1.1 ISO 15926

ISO 15926, named "Industrial Automation Systems and Integration of Life-cycle Data for Process Plants, including Oil and Gas Production Facilities, is an International Standard to systematize the flow of information in a process plant life cycle. It provides a conceptual data model that, along with reference data, can support specific life-cycle activities. It was created to tackle the challenge of integrating different software systems used during the entire time a process plant is active. ISO 15926, along with ISO 10303-221, were the ultimate contributions provided by the EU ESPRIT project, ProcessBase (1991-1994)[1], that aimed at developing a data format for dealing with information along with all phases of a process plant (EBRAHIMIPOUR; YACOUT, 2015).

This first part of ISO15926 (International Organization for Standardization, 2004) gives an overview of the standards and their parts up to the year 2004. ISO15926 was created to provide a standardized framework for data integration in process plants so that operators, designers, and other interested parties can effectively communicate with each other using a standard data model.

Part 2 of ISO 15926 defines the data model, with rules and constraints on how to use the standard. This part delivers a neutral conceptual model in relation to external models to be used later on with reference data. It covers generic concepts such as thing, class, and individual at the top-level and other 201 entity types. The axioms provided don't really constrain the model except for disjointedness since external data models are expected to provide further constraints. This data model has a four-dimensional (4D) view of the world.

The ontology presented in ISO 15296-2 was the target of some criticism in academia due to some misconceptions. Barry Smith clarifies the errors and defects of the supposed ontology provided by ISO 15926. A series of examples shows common errors in ontology development and good practices for building ontologies (SMITH; CORPORATION, 2006).

ISO 15926-3 (International Organization of Standardization (ISO), 2009) and ISO 15926-4 (International Organization for Standardization (ISO), 2019) provide the reference data library. The former focuses on geometry and topology entities, while the latter delivers a set of definitions called Reference Data Library (RDL). Each definition is linked to Unified Resource Identifier (URI). Almost 20000 classes are available through this part as specialization and subtypes of the classes provided by ISO 15926-2 (EBRAHIMIPOUR; YACOUT, 2015). Together with Part 2, these two parts comprise the core of ISO 15926.

ISO 15926-7 delivers a methodology for building templates based on the conceptual model provided in ISO 15926-2. ISO 15926-8 describes guidelines for implementing the upper-level ontology from Part 2 and template methodology from Part 7 into the RDF and OWL languages.

ISO 15926-9 elaborates on the Façades, RDF quad stores for the Nodes and Templates specified in ISO 15926-7. These Façades can be used for data consolidation and transfer between different computational systems participating during the life cycle of a production plant. Part 11[2] specifies a methodology for the simplified industrial use of reference data, as specified in Part 4, and is relevant to the process industry supply chain's plant life cycle stages. ISO 15926-12 establishes an ontology for integrating industrial data over its entire life cycle. ISO 15926-13 is an ontology for asset planning in process plants, including oil and gas production facilities. It also includes an XML standard for data sharing for asset planning, developed from the ontology.

ISO 15926-14 planned to specify an OWL 2 Direct Semantics ontology based on the ISO 15926-2 Data model. Lifecycle modeling, in particular, requires a different representation in this standard than the ISO 15926-2 4D approach to lifecycle modeling. This ontology part was intended to offer efficient reasoning support for classes and properties, which the other parts of this standard did not prioritize. This part has the status of "Deleted" in its ISO webpage [3] at the present moment (February 2023).

3.1.2 CFIHOS

Capital Facilities Information Handover Specification (CFIHOS) standard aims to enhance how information is transferred between suppliers that own, operate, and build equipment for the process and energy industries. Its objective is to establish a single language for transferring information in various sectors, starting with a standard equipment name taxonomy and supporting specification (International Association of Oil & Gas Producers (IOGP), 2020a).

At this moment, the current version of the standard is 1.4.1, released in December of 2020[4]. It comprises the narrative documents, the reference data library, a data model, and supporting templates. This standard uses natural language definitions from ISO 15926-4 in its Reference Data Library.

3.1.3 OntoCAPE

OntoCAPE is a formal, heavyweight ontology comprising 62 OWL files related to the domain of Computer-Aided Process Engineering (CAPE)[5]. OntoCAPE captures process engineering domain consensus knowledge in a generic form that can be reused and shared by groups of people and across software systems.

OntoCAPE can be used to create novel software support for various engineering activities, such as the systematic management and retrieval of simulation models and design documents, electronic plant equipment procurement, mathematical modeling, and the integration of design data from distributed sources. (MARQUARDT et al., 2010).

3.1.4 IOF

The Industry Ontology Foundry (IOF)[6] is an organization that is working to co-create a collection of open reference ontologies to serve the demands of the manufacturing and engineering industries and promote data interoperability.

The ontologies developed by the Industrial Ontologies Foundry (IOF) has chosen the Basic Formal Ontology (BFO) as its top-level ontology, an endeavor to build a set of ontologies to assist digital manufacturing on the part of representatives from many areas of the advanced manufacturing industry. A first preliminary draft of the first-order logic (FOL) axioms and definitions of an IOF higher ontology that descends from BFO is presented in (SMITH et al., 2019). The first version of the IOF-Core ontology is presented in (DROBNJAKOVIC et al., 2022). Figure 3.1 shows the backbone taxonomy of the independent continuants in the IOF-Core ontology.

Figure 3.1 – Taxonomy of the independent continuants in the IOF-Core ontology. Modified from (DROBNJAKOVIC et al., 2022).

3.1.5 Open Subsurface Data Universe^{TM} (OSDU) Forum

The Open Group OSDU^{TM} Forum was launched in 2019. The Open Group is a worldwide consortium of over 800 companies that use technological standards to help businesses accomplish their goals. One of its initiatives, the Open Subsurface Data

Universe$^{\mathrm{TM}}$ (OSDU$^{\mathrm{TM}}$) Forum, is a collaborative initiative to provide a single Open Source data platform and ecosystem in a vendor-neutral environment. The objective is to exchange non-differentiating technologies to speed up industry innovation and transformation. At the date of this publication, OSDU's most recent update is the Mercury Release (R3)[7]. It provides a single data platform to be hosted in a cloud that takes data from conventional data silos and makes it all accessible.

3.1.6 PPDM

The PPDM Association is a global non-profit organization that promotes the best practices, standards and education, certification, and professional development opportunities to the petroleum sector. Its primary goal is to deliver solutions and services that provide a solid and comprehensive basis for data management in the oil sector.

A reference list, also known as a vocabulary or a list of values, is a regulated and well-defined vocabulary that may be utilized in data management creation and practice. PPDM maintains a set of reference lists to improve communication among various industry players, including operators, suppliers, regulators, and the service sector.

The benefits of a shared understanding are evident, as it improves stakeholder communication, eliminates ambiguity, and eliminates the need for descriptive information qualifying. Interoperability comes from this shared understanding of what it means, contributing, among other things, to data governance through effective communication, interoperability amongst stakeholders, and system standardization for those who share the same sort of information. Different standardizations are used to display this vocabulary. The What is a Well (Professional Petroleum Data Management Association (PPDM), 2014) and What is a Completion (Professional Petroleum Data Management Association (PPDM), 2018) papers collect basic ideas for novices concerning the well and completion concepts. Each of these papers has definitions (for example, a definition of a wellbore), important ideas (explaining the concept), and explanations including a variety of connections with other relevant components (e.g., hierarchy, composition).

3.1.7 Energistics

Energistics is an open collaboration that defines, develops, and maintains data standards for the worldwide oil and gas sector. Petrotechnical Open Software Corporation (POSC) was formed in 1990 by a consortium of international operators and was renamed Energistics in 2006. Energistics is committed to informing, educating, and assisting all stakeholders in the quest for interoperability, efficiency, and data integrity through the quick and successful adoption of standards. Its contribution focuses on agnostic data transmission patterns consisting of interest groups (SIGs - Special Interest Groups), which characterize the EnergyML family of standards.

For segments of the business, three EnergyML family standards have been defined:

- RESQML is a standard for exchanging reservoir model data, spanning from basic structural modeling and interpretation through reservoir characterization and monitoring
- PRODML is a standard for exchanging data needed to optimize oil and gas well output, emphasizing data from the reservoir-well border to the custody transfer point.
- WITSML: Covers data from drilling activities, such as drilling data in real-time mud reports and well construction and location information.

3.2 Contributions from academia

Ontologies in Oil and Gas have been a research target since 2001 when Chen (2000) used the Inferential Modeling Technique (IMT), a conceptual modeling tool, for knowledge analysis in the petroleum remediation domain. Later, Chen et al. (2001) proposed a method for ontology construction and its application in which the authors apply the models to the problem domain of petroleum waste management and address knowledge representation at the various levels of ontologies. Later works have used IMT for developing an ontology for inspection in a process plant (CHAN; PENG; CHEN, 2002; CHAN, 2004; CHAN, 2005).

Nimmagadda, Dreher and ... (2005) proposes to simplify heterogeneous data structures through an ontological data modeling approach, particularly to address the issues of data integration and information sharing. The petroleum ontology framework, which delivers the conceptualization of petroleum ontology and methodological architectural views, has been described (NIMMAGADDA; DREHER; RAJAB, 2007). Nimmagadda and Dreher (2008) suggests methods for constructing and using the ontology-based warehouse. Bilong, Fucai and Jing (2010) shows the method used to build Petro-Onto (petroleum exploration and production domain ontology), the top-level ontology was designed, and an approach was proposed to capture concepts and the relationships among concepts automatically from business models and data models. Du et al. (2010) introduces ontology into knowledge construction of the petroleum exploitation domain to solve problems on information sharing and application integration caused by the inconsistency of information terms between disciplines.

Contributions from the industry are also seen with the development of the ISO 15926-2 (International Organization of Standardization (ISO), 2003) standard that provides a data model to be used for information exchange in the oil and gas industry. This model was criticized by a prominent researcher in the field of ontologies, Barry Smith, in (SMITH; CORPORATION, 2006). Later, other parts of the same standards were released (International Organization for Standardization (ISO), 2019; International Organization of Standardization (ISO), 2013; International Organization of Standardization (ISO), 2011a; International Organization of Standardization (ISO), 2011b; International Organization of Standardization (ISO), 2019; International Organization of Standardization (ISO), 2015; International Organization of Standardization (ISO), 2018; International Organization of Standardization, 2018), addressing some of the faults and covering other areas in the domain.

Ontologies have many uses, Carbonera, Abel and Scherer (2015) proposes an ontology-based approach for visual interpretation tasks. The used approach applies ontological meta-properties of a foundational ontology and proposes a cognition-inspired representation structure for inferential knowledge. The study applied the approach for interpreting depositional processes in Petroleum Geology. Also in the geology field, Cicconeto et al. (2022) introduces the GeoReservoir ontology, the outcome of an ontological examination of the terminology used by geologists in studies of deep-marine depositional systems. Another example of using ontologies in similar fields is shown in (CAMERON et al., 2022), which presents the Zero-defect Manufacturing (ZDM) framework for the process industry's engineering design and operations supply chain. Using the Tennessee Eastman Process System, a working example of ZDM has been provided. Although ontology is not the main contribution topic, they have been used as a reference for creating the ZDM framework.

There are also applications of ontologies to different problems in the oil sector. Shankar et al. (2022) have proposed and constructed a knowledge-based digital twin prototype for the O&G upstream using a generalized IoT stack and schema-based ontologies. Expressing risks have also been dealt with in (SILVA et al., 2021), where the authors proposed a framework for the conceptualization of risks in oil and gas projects as documented by experts in the domain of Petroleum Reservoir. It provides domain specialists with a predefined set of concepts for their areas of expertise. Ontologies have also been used in the drilling domain, Skalle and Aamodt (2020) developed a knowledge model of the drilling process. This model is referred to as a drilling ontology. The paper describes how the authors used knowledge modeling and ontology engineering to construct it and how it has been utilized to predict downhole failures during drilling.

As previously mentioned, data collection is the most time-consuming aspect of the decision-making process for petroleum companies. The retrieval and utilization of information in petroleum firms are difficult due to the massive amount of historical and created data. Calvanese et al. (2013) demonstrates an ontology method for data search and access. The Optique project comprises a goal sequence of steps to access Big Data better and faster based on ontology-based data access. Kharlamov et al. (2019b) describes the experience of addressing this data challenge at Equinor. Exploration geologists at Equinor, in particular, must analyze data about specific regions to locate new oil or gas reserves that oil companies can exploit. The authors developed an Ontology-Based Data Access (OBDA) system to find relevant exploration data.

To facilitate the development of data semantics standards for the oil and gas sector, Yuan and Li (2023) offers a model of data semantics standardization and an Oil&Gas industry reference vocabulary standardization model (OIRVSM). This study provides significant suggestions for implementing knowledge graphs with standardized vocabulary. It also gives an operational algorithm for generating a standardized knowledge graph (AC-SKG) in the Oil-and-Gas business. This contribution can help to overcome the 'information island' issue in 'Smart Fields.' As stated in the paper's conclusion, "*the problem of semantic standardization in knowledge organization is ultimately the problem of vocabulary standardization used to designate concepts*."

3.3 Research gaps

The diversity of data required to support the exploration and production activity reflects in many proprietary software and specialized services that challenge information integration. The actors have answered this challenge by making industry standards in response to the digitization demands of Industry 4.0. The "What is a Well?" (Professional Petroleum Data Management Association (PPDM), 2014) and "What is a Completion?" (Professional Petroleum Data Management Association (PPDM), 2018) sections of the Professional Petroleum Data Management (PPDM) glossaries represent two of the processes involved in this procedure. There are also efforts to provide a semantic framework, like ISO 15926's ontology and reference library. A syntactic framework, like the Open Subsurface Data Universe Forum's (OSDU) (WHITLEY et al., 2020) integrated data platform, the Energistics Consortium's data standards, the WITSML, PRODML, and RESQML standards, and CFIHOS's equipment specifications (International Association of Oil & Gas Producers (IOGP), 2020b). In these earlier projects, data accessibility was improved in two ways: (1) by creating a standard data model that gives an integrated view and (2) by defining a typical software architecture for sending, storing, and getting data. Even though the two methods work well together, they both lack the implicit meaning of the entities to make it easier to align the data automatically.

Our work fills the gap to provide this support and proposes a semantic infrastructure for further modeling the occurrents of the petroleum production process. Although the modeling of occurrents itself is not in the scope of this thesis, we consider that our work is the first step toward this achievement.

More specific research gaps were identified as follows:

1. Broader works or standards that define a great set of terms from the domain, such as seen in (International Organization for Standardization (ISO), 2019; International Association of Oil & Gas Producers (IOGP), 2020b) lack complete semantic formalization by logically restricting the meaning of terms, which can cause some problems. For instance, let's consider the informal definition of a “christmas tree” in the documentation of the CFIHOS standard[8] is: “*a physical object that is an assembly of pipes and piping parts, with valves and associated control equipment that is connected to the top of a wellhead and is intended for control of fluid from a well*”. The informal definition is essentially correct. However, navigating through the subClassOf hierarchy in the standard, “christmas tree” is defined as a type of

“piping and pipeline equipment” [9] which in turn is defined as a type of “mechanical equipment class” [10]. The informal definition considers a christmas tree as a concrete entity (i.e., a physical object). In contrast, by the subClassOf relations in the standard, a christmas tree is an abstract entity (i.e., a class). There is a contradiction between the informal definition and the formalized definition.

2. Works related to upstream oil and gas such as (SKALLE; AAMODT, 2020; BILONG; FUCAI; JING, 2010) are not broad enough to provide the formalized meaning of entities (i.e., well or reservoir) that encompass the entire domain of oil and gas.

3. Higher-level ontologies such as (SMITH, 2019; GARCIA et al., 2020) are not about the domain of oil and gas.

The previous work mentioned above shows the gaps this research intends to fill. Another aspect is that most of the mentioned works that provide an ontology for the oil and gas domain as a product are not available to the public and/or haven’t used a top-level ontology or an accredited development methodology to anchor the meaning of the entities and later be reused by other ontologies. O3PO makes a significant contribution as a domain reference ontology specifically tailored for offshore petroleum production plants, utilizing BFO as its top-level ontology. This innovative approach bridges existing gaps in the field and fosters standardization and data integration across siloed systems within the oil and gas industry.

Reservoir simulations, wells, flowlines, christmas trees, platform equipment, and other data are used in developing petroleum fields. Although there are software solutions that use this data independently to construct models or monitor status utilizing databases, there isn’t a solution that brings all of the data together in a single application in a smart and timely manner.

The lack of conceptual frameworks to organize the knowledge surrounding petroleum production is one of the most significant barriers to achieving such integration. Such a contribution would allow for a more comprehensive perspective of petroleum production systems, with an accurate data correlation.

Unlike a top-level ontology that provides a comprehensive representation of the world (e.g., UFO, BFO, and DOLCE) or a core ontology that defines a minimal set of concepts and relationships that define the common knowledge of a domain (e.g., GeoCore, and IOF-Core), a domain ontology focuses on a particular area of knowledge or application. Domain ontologies can be reference or application ontologies. They provide a common vocabulary and shared understanding of the domain's essential entities, attributes, and relationships. It can serve as a source for vocabulary uniformization or a foundation for data schemas generation. This research will aim to deliver such a contribution, particularly to the offshore petroleum production domain.

Therefore, this research delivers a domain ontology for offshore petroleum production plants using well-accredited principles for ontology development, contributing to current industry and academic efforts to standardize and integrate data from siloed systems. Additionally, this research shows the long process of requirements specification for building a domain ontology and later applying the ontology in particular use cases that relate to the interoperability problem in the oil and gas industry.

4 DEVELOPMENT METHODOLOGY

Considering the nine scenarios included in NeOn, this ontology development combines scenarios 2 and 3, applying non-ontological and ontological resources to build the ontology. Like all the other scenarios covered by NeOn, this research follows the sequence of activities of the first scenario, from specification to implementation, with some alterations.

The ontology development started with the ontology requirements specification activity when we elicited a set of competency questions (CQs) from the technical staff of the petroleum company. These requirements further oriented the search for technical documentation. To capture the requirements, the authors accomplished six interviews with domain experts from the industry, mostly petroleum engineers responsible for the petroleum plant's daily operation, to elicit the goals of the ontology and identify the most relevant entities. Each interview lasted 1 to 2 hours and questioned the professionals about their daily tasks, information requirements, and overall activities. The analysis of the transcripts of the interviews enabled the identification of keywords and relevant properties in the domain.

The process of requirements specification was iterative. The interactions with petroleum company staff and analysis of available data, led the authors to identify new requirements, bringing new concepts for modeling along the way.

The non-ontological resources were selected based on their relevance to the domain. Also, according to the functional requirements identified, the search for non-ontological resources was focused on attending to the requirements such as the definitions for specific domain entities. The authors collected data from the following non-ontological resources:

Industry-wide glossaries such as:

- natural language definitions from the Reference Data Library present in ISO 15926-4 (International Organization for Standardization (ISO), 2019);
- the Society of Petroleum Engineers (SPE) Petrowiki;[11];
- the Professional Petroleum Data Management (PPDM) *What is a Well?* Standard (Professional Petroleum Data Management Association (PPDM), 2014);
- the Schlumberger Oilfield; Glossary[12];

- American Petroleum Institute (API) Glossary of Oilfield Production Terminology (American Petroleum Institute (API), 1988);
- and public-accredited sources of definitions such as Wikipedia, and standard dictionaries when more definitions were necessary.

Considering ontological resources, O3PO uses Basic Formal Ontology (BFO) 2020 (OTTE; BEVERLEY; RUTTENBERG, 2022) as a top-level ontology. This version of BFO became an international standard, ISO/IEC 21838-2. It is a small, and well-documented foundational ontology with a philosophical basis on realism that shows to be adequate for a material domain such as the facilities of a production plant. On top of that, BFO is widely adopted in different domains, making it suitable to work as a common umbrella for other related ontologies developed in the industry. Besides BFO, O3PO derives some concepts from GeoCore [10], the core ontology from Industrial Ontologies Foundry (IOF) (DROBNJAKOVIC et al., 2022) middle-level ontologies. The resource adapted some relations from Flow Systems Ontology (FSO) (KUKKONEN et al., 2022), but the ontology was not imported directly. Figure 4.1 shows the relations between the imported ontologies.

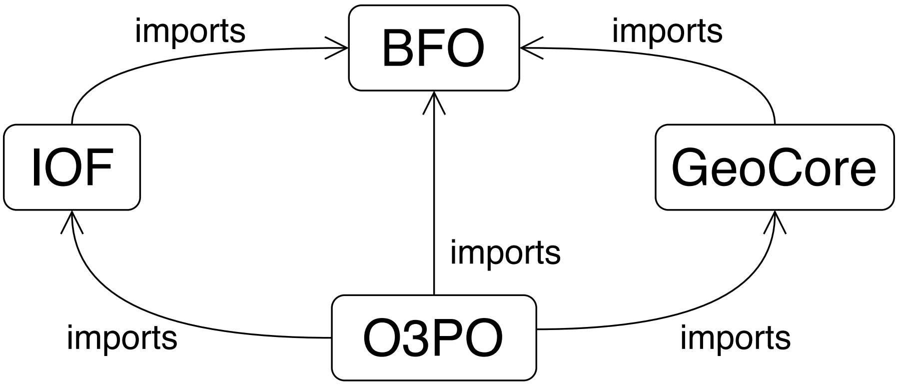


Figure 4.1 – Imported ontologies in O3PO. Taken from (SANTOS et al., 2022).

We gathered a set of natural language definitions for each term from the referred non-ontological resources. Based on those definitions, we elaborated semi-formal Aristotelian definitions using the genus-differentia form. In this form of definition, there are two primary elements: the genus serves as the head of the definition and represents the immediate superordinate type, while the differentiae are additional components that set apart the defined type from other types belonging to the same genus (SEPPÄLÄ; RUTTENBERG; SMITH, 2016). From those definitions, the entities were classified according to BFO, and the relations were considered object properties. Properties of each class such as dependence and part-whole relations were added according to the consensual knowledge about the entities in the domain.

The produced ontology could be implemented in any representational language. Nonetheless, we must constantly remember that every language has its expressiveness constraints compared with the wanted expressivity of an ontology. Regardless of our language, we will likely need to modify the model to arrive at compromise solutions that privilege some elements over others. In light of this, our ontology is implemented in OWL 2 - one of the most widely used ontology languages, a de facto standard in the life sciences industry, and a W3C recommendation. In a nutshell, ontologies are represented in OWL 2 by classes, individuals or instances, properties, and property restrictions. Also, it is possible to define property chains, which state that a sequence of relationships along a row of individuals implies a particular relationship between the individuals at the beginning and end of the row (for example, x hasMother y and y hasSister z could be defined to imply that x hasAunt z). Meanwhile, in developing the ontology, we have been checking the logical consistency with reasoners and the semantic consistency by the user and expert continuous verification in return meetings where the partial results are presented and discussed. Protégé (MUSEN, 2015) was the tool used for editing and checking logical consistency to build the model.

For validation, we used a dataset from an anticipated production system (SPA) in an offshore oil field from Brazil's Pre-salt. CSV files with time-series data composed the dataset used for validating O3PO. It covers two wells, a producer and an injector well. Engineers label the files with tags. A sequence of consecutive measurements of the same properties, such as pressure, temperature, and flow rates, by sensors spread in the plant, produced the labeled files.

Since O3PO doesn't deal with information artifacts such as those provided in the dataset, we used the Information Artifact Ontology (IAO) (SMITH; CEUSTERS, 2015). IAO used BFO as a top-level ontology for representing types of information content entities (ICEs) such as documents, databases, and digital images.

Thus, to demonstrate the usability of O3PO, which explicit the semantics of the material entities, we used two concepts from IAO, namely the class `iao:information content entity` (i.e., a generically dependent continuant that is about some entity) and the relation `iao:isAbout` (i.e., a relation that relates an information artifact to an entity). They were used to tie the information artifacts to the material entities, providing the means to query files that hold measurement data about instances of properties that inhere in the material artifacts of the plant.

Additionally, with the help of industry experts, we elaborated three use cases based on the motivating scenario. The use cases provided the means to prove the usability of the ontology in a real-world scenario. A set of DL queries was performed and compared with expected answers.

5 CONCEPTUAL MODELING

This chapter describes the problem domain and scope, the utilized methodology, the ontology development process and results in the creation of the domain ontology for the petroleum production domain with the goal of supporting a digital twin for monitoring and optimization. Later, we discuss a few misconceptions in the domain and the treatment of time in this work.

5.1 Problem domain and scope

In order to achieve a useful conceptual model for the production domain that can serve as a semantic core to the flow of information in a digital twin, we need a domain ontology that has a production view and not a design view. In other words, the ontology must describe the properties that refer to and change according to the production process of a platform. They must be the properties necessary and sufficient to identify the progress of the production process. To specify such properties, we considered a motivating scenario inspired by the interviews with domain specialists. Later, the ontology development methodology is presented.

5.1.1 Motivating scenario

Typically, an offshore production system has the following main components:

- The reservoir
- The wells
- The subsea lines and facilities
- A platform, with its primary processing facilities

These entities comprise the oil path and its immediate surroundings, from the porous medium to the storage and transfer tanks. These include units that separate the fluid produced in its oil, gas, and water phases. Also part of the system is the artificial lift and water or gas injection subsystems that aid in the displacement and recovery of the oil.

Finding relevant data during day-to-day workflows is a time-intensive activity for a petroleum engineer working off-site with offshore oil fields. An engineer in a petroleum company needs to access all production facility data to manage the field. This data is collected and labeled with tags that point to digital signals acquired by measurement instruments, along with timestamps, regarding particular properties of assets (e.g., downhole pressures, separation temperature, liquids flow rate, etc.). The most used tags to manage the facility's production are related to production and injection wells. Figure 5.1 shows an overview of the primary elements within an offshore petroleum production plant (considering only one production well), along with the data sources that cover the entire facility. There are usually three scenarios where the engineer must locate specific data: properties of the well that are, in fact, properties of parts of the well; properties of all the wells connected to a specific production facility or rig; and properties of all rigs located in a particular offshore field. At an exact time, the engineer intends to forecast the oil flow for the field's remaining lifetime and needs to access the flow rate, pressure, and temperature time series data of all wells and their components connected to all rigs in a particular field. To do that, he must locate where is each tag or file in the company's database that maps each measured property for all the time series of interest. These data-gathering activities take a lot of time (COSENTINO, 2001) and are prone to mistakes, mainly due to the lack of semantic context in those tags and files. Often the engineer retrieves the context only in the string that identifies the tag, formed by mnemonics of sensors, their relative positions, and well names. Although there is previous knowledge regarding the tag names and their meaning, someone can make mistakes or need to check the data to infer the signal effectively measured, drawing a complex scenario for the workers and a complex barrier for novices. This problem also impacts the data processing by machines, in which the lack of semantic context limits artificial intelligence applications.

5.2 Ontology requirements specification

As earlier mentioned, the NeOn methodology prescribes a set of activities to be followed for each considered scenario. For the ontology requirements specification activity, we started by eliciting a set of competency questions from the technical staff of the petroleum company that oriented the search for documentation.

A set of 60 competency questions were generated through the ontology requirements specification activity. Table 5.1 shows the set of competency questions.

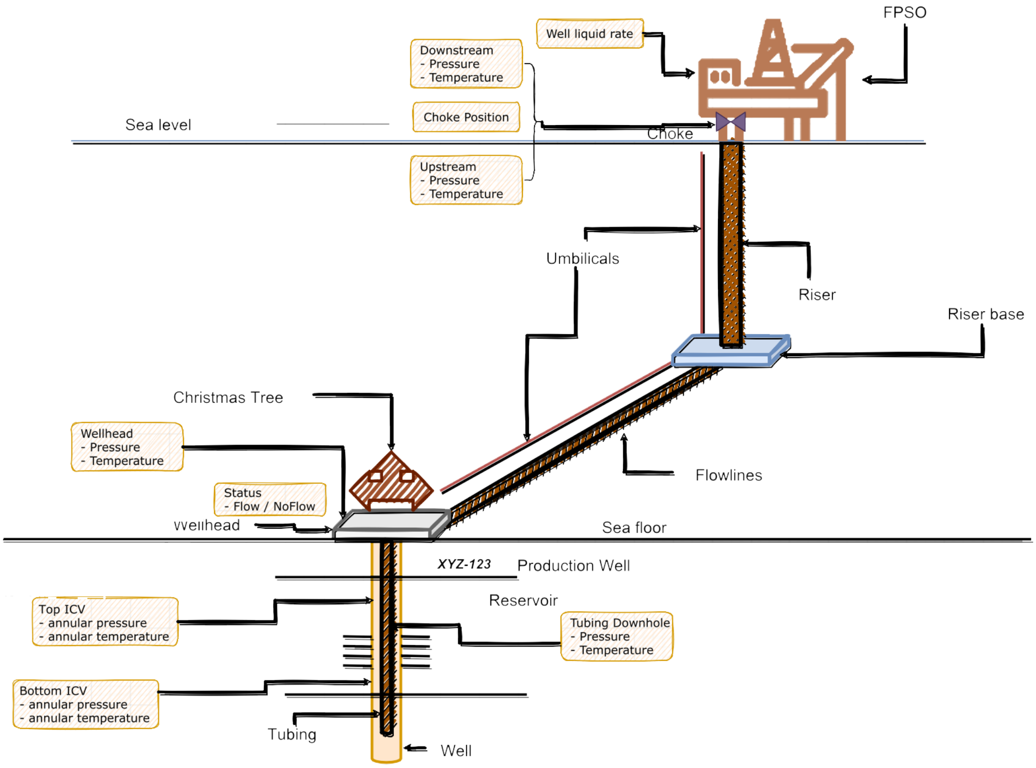


Figure 5.1 – Illustrative overview of the primary elements within an offshore petroleum production plant considering only one production well connected to the platform, along with the data sources that cover the entire facility. Produced by João César Netto.

Table 5.1 – List of acquired competency questions.

Nr.	Competency question
1	What is a field?
2	What is a basin?
3	What is a well?
4	What is oil?
5	What is gas?
6	What is petroleum?
8	What is pressure?
9	What is production?
10	What is injection?
11	What is a zone?
12	What is a reservoir?
13	What is a choke?
14	What is an ICV?
15	What is a wellhead?

Table 5.1 – List of acquired competency questions.

Nr.	Competency question
16	What is tubing?
17	What is a valve?
18	What is a christmas tree?
19	What is a manifold?
20	What is a wellbore?
21	What is a separator?
22	What is a tank?
23	What is a platform?
24	What is a plant?
25	What is a production well?
26	What is an injection well?
27	What are the properties of a well?
28	What are the properties of a field?
29	What is the name of the field?
30	Is the field onshore or offshore?
31	Which basin does the field belong to?
32	If offshore, what is the water depth?
33	What is the name of the well?
34	Where is the well located?
35	What is the well function?
36	What is a shape of a well?
37	What is the total depth of the well?
38	What is the oil production rate?
39	What is the total oil in place?
40	What is the density of the oil?
41	What is the associated gas flow rate?
42	What is the gas-oil ratio (GOR)?
43	What is the gas injection rate?
44	What is the non-associated gas production rate?
45	What is the condensate production rate?
46	What are the properties of a measurement?
47	What is the type of sensor used for the measurement?

Table 5.1 – List of acquired competency questions.

Nr.	Competency question
48	What is the timestamp of the measurement?
49	What is the depth of the measurement?
50	What is the type of measurement?
51	Which company is operating the measurement?
52	What is the tubing downhole pressure?
53	What is the wellhead pressure?
54	What is the unit of the measured pressure?
55	Which fluid is being injected?
56	What is the fluid injection rate?
57	From how many zones does the well produce?
58	What is pressure on the IC Valve located in the zone?
59	What is the temperature on the IC Valve located in the zone?
60	In how many zones in the well does the injection occur?

A preliminary set of 25 terms was selected for modeling. These terms fall within the ontology's scope and are considered the most pertinent terms based on the scenario described in the preceding section. However, the incremental and iterative nature of the followed methodology permits adjusting the specified requirements depending on the necessity or possible application of the ontology (SUÁREZ-FIGUEROA et al., 2012). Table 5.2 shows the initial modeling terms covered by the ontology.

christmas tree	duct	field	flow	flow rate
flowline	fluid	ICV	manifold	petroleum
pipe	pipeline	platform	pressure	pump
reservoir	riser	stream	temperature	tube
tubing	valve	well	wellhead	zone

Table 5.2 – Current terms included in the ontology.

After specifying requirements, a set of non-ontological and ontological resources was used to gather the most consensual knowledge in the domain and to reuse useful concepts from other ontologies as much as possible.

5.3 Conceptualization

This section shows the conceptualization of the ontology. First, it presents some combined natural-language definitions, then it shows the independent continuants, dependent continuants and relations covered in the ontology.

5.3.1 Entity definitions

For each term, a set of natural language definitions were gathered from the above sources, and a combination of these definitions using the genus-differentia form, following Aristotelian logic (PARRY; HACKER, 1991) was performed. To provide an example of a definition merging process, let's consider the term `christmas tree`. Table 5.3 shows a subset of the definitions gathered from the non-ontological resources chosen for ontology development.

Source	**Definition**
ISO 15926-4	A piping system that is an assembly of pipes and piping parts, with valves and associated control equipment that is connected to the top of a wellhead and is intended for control of fluid from a well.
SLB Glossary	The set of valves, spools, and fittings connected to the top of a well to direct and control the flow of formation fluids from the well.
API Glossary	An assembly of valves and fittings attached to the uppermost flange of the tubing head, used to control well production.
Wikipedia	An assembly of valves, casing spools, and fittings used to regulate the flow of pipes in an oil well, gas well, water injection well, water disposal well, gas injection well, condensate well and other types of wells.

Table 5.3 – Subset of natural language definitions gathered from non-ontological resources.

After generating a combined natural-language definition, a semi-formal definition is generated by replacing the terms in the definition with terms defined in the ontology. For this example, the semi-formal definition for `christmas tree` is "def. an **iof:assembly** that **o3po:has component** some **o3po:valves** that is **ro:located in** above a **o3po:well** and that is **o3po:connected to** the **o3po:well** to control the **o3po:flow** of fluids from/into the **o3po:well**."

After separating the genus and differentia for each definition, a combined natural-language definition is generated. Consider the first definition in Table 5.4, which shows the natural-language definitions for the resulting set of terms. Since we conceived the domain ontology for continuous evolution by the application of our described methodology, the terms included are subject to change.

Table 5.4 – Natural-language definitions of the entities.

Entity	**Definition**
christmas tree	an assembly of valves that is located above a well and that is connected to the well (i.e., to the wellhead) to control the flow of fluids from/into the well.
valve	a piece of equipment that controls (e.g., permits, obstructs, regulates, directs) a stream/flow of fluids (e.g., gases, liquids) or fluid-like amounts (e.g., slurries, fluidized solids).
wellhead	an assembly located to the top of the casing strings of a well, composed of spools, valves, fittings, tubing and casing heads, and the means for hanging and isolating the various tubular strings.
pump	an engineered system that moves (e.g., transfer, deliver, compress) fluids (e.g., liquids, gases) or fluid-like amounts (e.g., slurries) by increasing fluid pressure.
pipe	a piece of equipment that is enclosed and hollow and that acts as a passageway to carry different sorts of things (e.g., fluids, cables, other ducts).
pipeline	an assembly that is composed of tubes (or pipes) that are connected to one another in a single, maximal/end-to-end line, and that is can convey liquids, gases, or finely divided solids.
tubing	a pipeline that is installed in a well, located inside all casing strings, that extends from the wellhead to the production zones, to conduct fluids from the reservoir into the christmas tree, or from the christmas tree into the reservoir.
flowline	a pipe or pipeline that carries oil, gas, or water that connects the wellhead to a manifold or a production platform.

Table 5.4 – Natural-language definitions of the entities.

Entity	**Definition**
flow	continuous movement of some amount of fluid, fluid compound, or fluid-like substance (e.g., slurry) through a conduit in a single average direction that is parallel to the walls of the conduit.
riser	a pipe or pipeline that vertically carries materials (e.g., production fluids, drilling mud, drilling cuttings) or artifacts (e.g., cables, pipes) and that connects a wellhead or a manifold on the seafloor to a platform.
(portion of) fluid	a (portion of) substance that tends to flow in response to external pressure and has no fixed shape
pressure	a quality that corresponds to the intensity of the force applied perpendicular to the surface of a material entity per unit area over which that force is distributed
temperature	a quality that corresponds to the intensity of the thermal energy of a material entity.
flow rate	a quality of a flow that corresponds to the volume of fluid that flows per unit of time through a conduit.
well	an engineered system that has as part one or more wellbores for producing or injecting fluids or fluid-like materials (e.g. hydrocarbons, water) from or to a subsurface reservoir.
wellbore	an artificial hole that is drilled in the ground (i.e., surface or seabed) and that is intended to be used in an exchange of fluids between a subsurface reservoir and the surface (i.e., production or injection) and in the acquisition of information (e.g., measurement of temperature, rock properties).
zone	a reservoir or a section of a reservoir that is drilled into from which hydrocarbons are produced or fluids are injected.
reservoir	a rock containing one or more individual and separate natural accumulations of petroleum confined by impermeable rock.

Figure 5.2 presents the backbone taxonomy of independent continuants in O3PO.

We check the logical consistency with reasoners and the semantic consistency by the user and expert verification.

Some important definitions from IOF-Core (DROBNJAKOVIC et al., 2022) and GeoCore (GARCIA et al., 2020) should be considered to show the reasons for defining several terms in the ontology.

1. `iof:equipment`: "*material artifact which has an equipment role*"
    - `iof:equipment role`: "*role held by a material artifact when it is planned to be involved in or is involved in carrying out some part of a planned process and that is not consumed in that planned process*"
2. `iof:assembly`: "*material artifact that is composed of material components that are physically connected and that is capable of disassembly*"
3. `iof:engineered system`: "*system that is deliberately created to have a certain function*"; "*every instance of 'engineered system' is defined as exactly an instance of 'system' that is the 'bearer of' some 'function' which is 'prescribed by' a 'design specification*'"
4. `geocore:earth fluid`: "*Earth fluid is an 'earth material' that is fluid.*"
    - `geocore:earth material`: "*It is a natural amount of matter 'generated by' some 'geological process'.*"
5. `geocore:rock`: "*An amount of rock is a solid consolidated 'earth material' that is 'constituted by' an aggregate of particles made of mineral matter or material of biological origin.*"

The core ontology provided by IOF (DROBNJAKOVIC et al., 2022) offers a definition for `equipment`, which we have adopted. This definition emphasizes the significance of a `planned process` in providing identity to the term. Additionally, the IOF-Core ontology includes two other terms: `assembly` and `engineered system`. Although these terms appear closely related, they refer to distinct types of entities. An assembly is viewed as a material artifact that originates from an assembly process. In contrast, an engineered system possesses a specific function and adheres to a set of specifications.

`Assembly` is a helpful term for describing entities such as `pipelines`, `christmas trees`, and `manifolds` that can be disassembled. Meanwhile, the term `engineered`

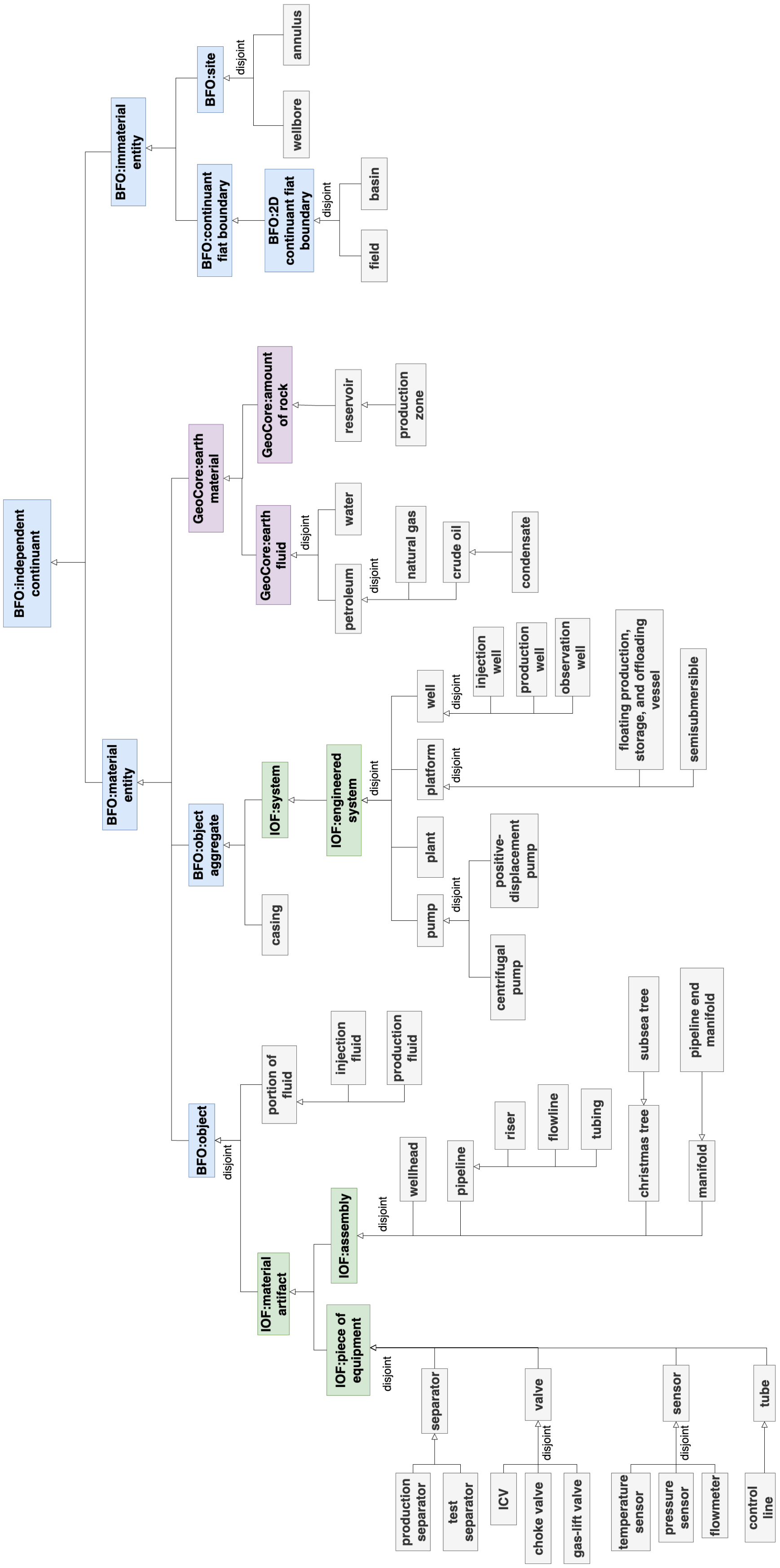


Figure 5.2 – Taxonomy of independent continuants present in O3PO.

`system` connects the concept of a system to possessing a function disposition. It also incorporates design specifications in its definition, which serves as a valuable superclass for terms like `well`, `pump`, `plant`, and `platform`.

To define the fluids extracted from reservoirs, we utilized the concept of `earth fluid` from the GeoCore ontology (GARCIA et al., 2020). With the superclass `earth material` this term relates the concepts to geological processes. This constraint proved beneficial as `petroleum`, `crude oil`, `natural gas`, and `water` were defined as `earth fluids`, while a `reservoir` was defined as a `rock` that assumes a `reservoir role`. This role relates to the containment of hydrocarbons within.

Another aspect of our ontology is that it provides categories for entities described in terms of their contingent component, such as how an asset is utilized. An example is producer and injector wells, wells engaged in the flow of oil from the reservoir to the platform or from the platform to the reservoir. There are also several roles for a pipeline, including tubing (pipeline put within a well), flowline (pipeline connecting the well to a manifold or a platform), and riser (i.e., pipeline connecting a well or a manifold to a platform). We identify these entities as having particular duties depending on their scenario. The tubing transfers the oil from the reservoir to the wellhead, whereas the riser is responsible for transporting the oil from the ocean floor to the platform. These are processes with particular characteristics; therefore, it makes sense to designate the relevant roles (for example, it is more challenging to make the oil rise than to move it sideways, so pipes are subject to different effects). Future work will elucidate the precise procedures by which these functions are carried out. Our ontology permits the tracking of an entity throughout whatever changes it may experience (e.g., distinguishing the notions of a well that is initially used to produce oil and later employed to inject fluids into the reservoir) by separating the essence of entities from their dependent characteristics.

5.3.2 Relations component_of and connected_to

Lastly, the ontology comprises two primary sorts of relationships that can exist between the assets under consideration. One is the `componentOf` relation, a binary, transitive mereological relation between instances of a material entity. We utilize it to describe the functional decomposition of plant assets. It specializes `bfo:partOf` relation constraining the domain and range to `bfo:material entities`.

The other relation is the `connectedTo` binary symmetric relation. It indicates

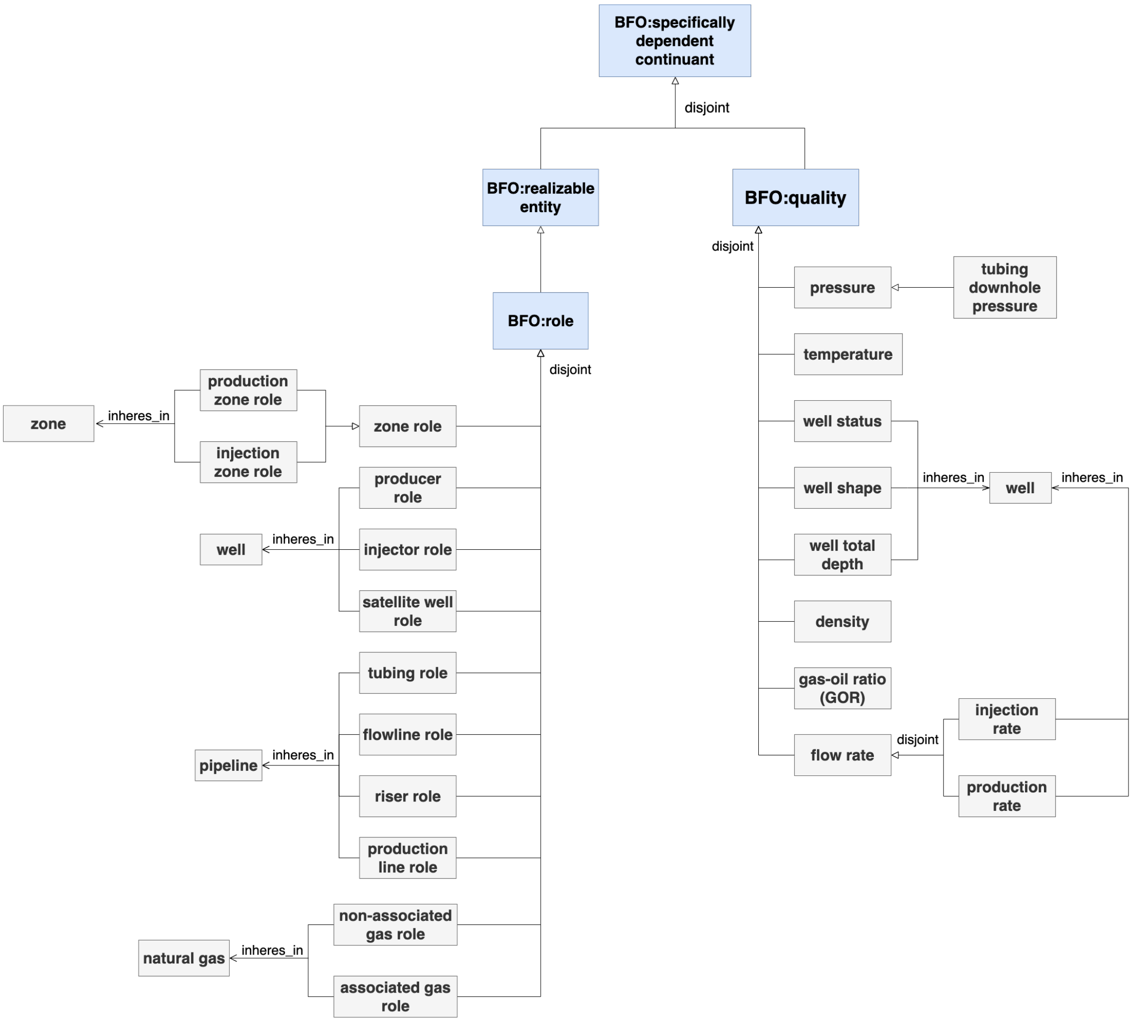


Figure 5.3 – Specific dependent continuants included in O3PO. Taken from (SANTOS et al., 2022).

the oil path through the various plant assets. In addition, if there are three items, X, Y, and Z, and X is `connectedTo` Y, which is a `componentOf` Z, then X is understood to be `connectedTo` Z. Similarly if Y is a `componentOf` X and Y is `connectedTo` Z, then X is likewise `connectedTo` Z. Both relations have as domain and range `bfo:material entities`. These relations are automatically derived by reasoning over the ontology.

The ontology adapts the Flow Systems Ontology (FSO) (KUKKONEN et al., 2022) to define various sorts of connectivity explicitly. As subclasses of the relation `O3PO:connectedTo`, we constructed equivalents of the `exchangesFluidWith` relation and its specializations, such as `feedsFluidTo` and `suppliesFluidTo`. These were not considered transitive relationships. Adding connection relations based on FSO to O3PO will provide reasoning about the fluid exchange between various plant systems.

Figure 5.4 illustrates the application of our ontology elements to describe the oil path in a generic production facility. The rectangles represent specific instances of the class that labels each rectangle. The diagram demonstrates a particular route of oil from a reservoir to a platform via a fluid supply chain and component relations. In addition, it depicts the partial disassembly of a well into some of its constituents and their inhering qualities. Different configurations may exist based on the parameters of the oil production environment.

In addition, some preliminary work has been done in modeling ocurrents. It is notable the dual aspect of the behavior of the plant. On one side, there are the material entities and the dependent continuants that inhere in them. On the other hand, the entities that extend in time, such as flows, activities, and events, are also inherent to the modeling of plant behavior.

Figure 5.5 depicts the relationships between ocurrents, material entities, and their attributes. We consider flow rate as an `iof:flow process characteristic`. We regard `production` and `injection` as subclasses of `flow`, which participants are `production well` and `injection well`, respectively. Also, `reservoir` and `earth fluid` are treated as participants of both subtypes of `flow`. It is essential to mention that the treatment of ocurrents in this field is in its infancy.

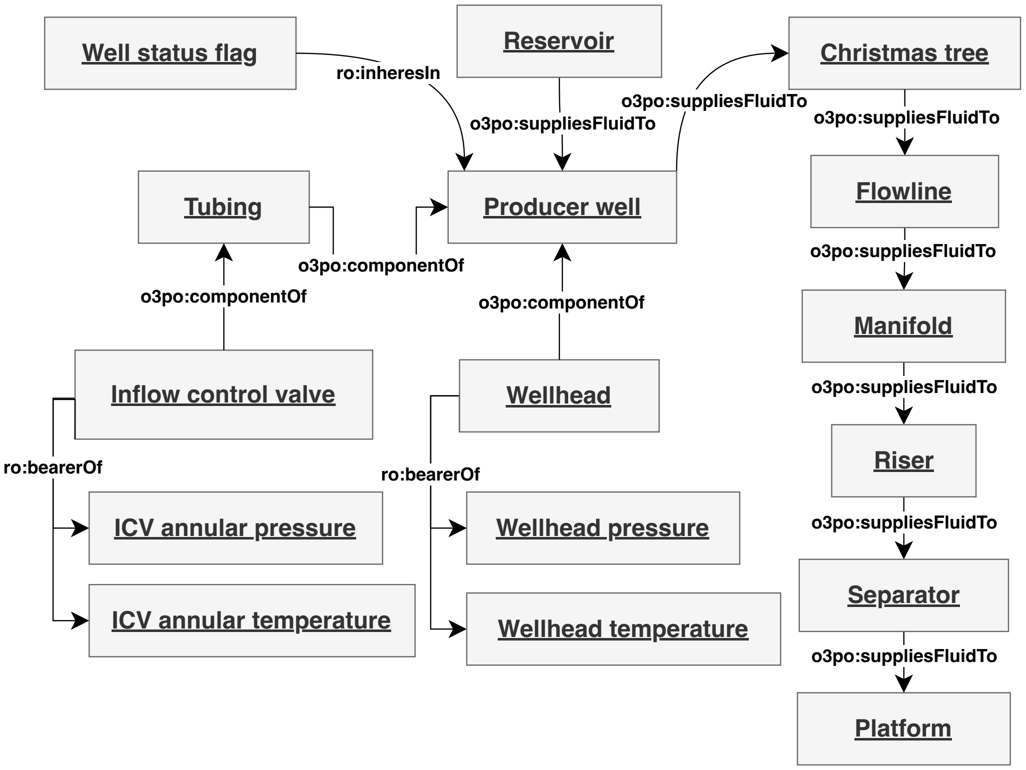


Figure 5.4 – Diagram showing possible relations between instances in a subsea production system. Taken from (SANTOS et al., 2022).

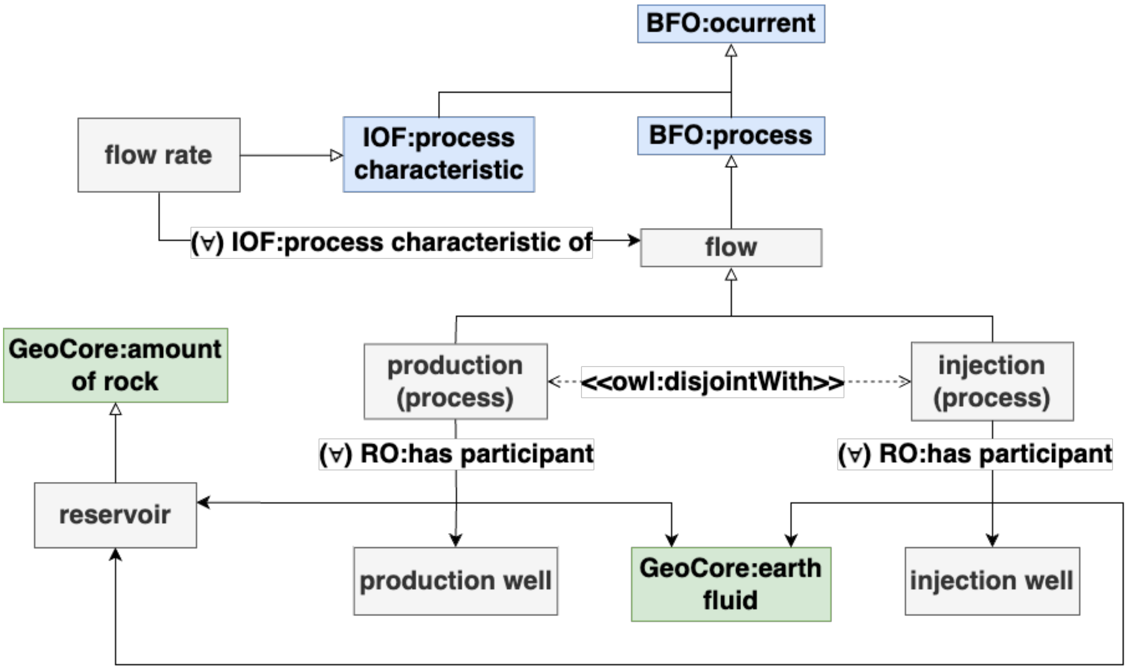


Figure 5.5 – A diagram showing ocurrents modeling in O3PO. Taken from (SANTOS et al., 2022).

5.4 Discussion

In this section, we raise some common misunderstandings on the terminology semantics of petroleum plant vocabulary and discuss how the domain's ontological analysis can help avoid them. Also, we discuss the treatment of time in our ontology.

5.4.1 Ontological analysis of common misunderstandings in the domain

One of the leading entities of a petroleum production plant, onshore or offshore, is the *well*. However, we should clarify what do we refer to when we say *well*. Does it represent a hole created by drilling in a particular location? Or a set of equipment such as a wellhead or the production string? One of the primary references used in the petroleum industry is the product of a workgroup from PPDM, *What is a Well? (Professional Petroleum Data Management Association (PPDM), 2014)* a booklet that defines the entities that pertain to a well and its context. From the definition adopted by the organization, the term refers to "a permitted or actual drilled hole in the ground designed to exchange (or facilitate the exchange of) fluids between a subsurface reservoir and the surface (or another reservoir) or to enable the detection and measurement of rock properties." This definition is ontologically inconsistent since it defines, for the same entity, the well as an informational entity, the permitted well (not yet existent), and an immaterial entity (the hole). The ontological analysis of the definitions proposed by "What is a Well" was already performed in (ABEL et al., 2017) pointing out these inconsistencies. The authors define a well as a "non-natural hole drilled in the ground in a particular location" and classify it as a BFO:Site, an immaterial entity. Accordingly, the equipment and other entities that occupy this site along the entire Well, from planning to abandonment, are called a Well Set. While both works have different perspectives, we consider the perspective of professionals in the petroleum production domain. So we assume that a well is the sum of a wellbore (the immaterial entity) and the equipment that are part of the Well to realize a particular function.

5.4.2 The treatment of time

The Offshore Petroleum Production Plants Ontology (O3PO) is a domain ontology that adopts BFO as its top-level ontology. While it is composed almost entirely of continuants, it does not necessarily adopt a perdurantist view of the world. To understand why, let's briefly discuss the difference between perdurantism and endurantism.

Perdurantism is a view in the philosophy of time and identity that claims that objects persist through time by having temporal parts, which means that they exist as different time-slices or stages at different points in time. In contrast, endurantism holds that objects persist through time as complete wholes and maintain their identity without having temporal parts.

BFO is an upper-level ontology designed to provide a coherent and consistent foundation for domain ontologies, such as O3PO. BFO distinguishes between continuants and occurrents (ARP; SMITH; SPEAR, 2015). Continuants persist through time while maintaining their identity and having no temporal parts. Occurrents, on the other hand, are entities that unfold through time in successive temporal parts or phases.

Since O3PO is based on BFO, it naturally incorporates the distinction between continuants and occurrents. Given that O3PO is composed almost entirely of continuants, it focuses on entities that persist through time without having temporal parts. This means that O3PO primarily adheres to an endurantist view of the world rather than a perdurantist one.

However, it is important to note that adopting BFO as a top-level ontology does not necessarily preclude O3PO from incorporating perdurantist elements or perspectives. BFO's occurrent category could provide a framework for representing processes that unfold through time with temporal parts. In such cases, O3PO could incorporate both endurantist and perdurantist elements, depending on the specific representation requirements of the domain. Future research will address the occurrents in the context of petroleum production.

6 VALIDATION

This chapter presents a succinct overview of how the ontology may be utilized, drawing on the competency questions that directed its creation as well as typical scenarios encountered in the industry's daily operations. Additionally, three applications of the ontology based on use cases inspired by the motivating scenario are presented.

6.1 Competency questions verification

In this section, we briefly present some possibilities for the use of the ontology based on the competency questions that guided the development of the ontology and some stereotypical scenarios of the daily routine in the industry.

A standard way of verifying if an ontology meets its requirements is by assessing whether it provides support to answer the competency questions on which it is based suitably. In what follows, we overview how to answer those questions using the ontology.

1. *What is the well connected to?* - Look for the instances that stand in a relation `connected_to` with the desired instance of `Well`.

2. *What are the components of the well?* - Look for the instances that stand in a relation `component_of` with the instance of `Well` of interest.

3. *What are the components of the production line?* - Analogous to the answer of CQ 2.

4. *What are the components of the tubing of a well?* - Analogous to the answer of CQ 2. Here, the *tubing of a well* is the instance of `Tubing` that is `component_of` the chosen instance of `Well`.

5. *How many production zones does a well have?* - Look for the instances of `Zone` with which the desired instance of `Well` stands in a relation `produces_from` and then count the results.

6. *How many zones does a well produce from?* - A different way to ask CQ 5.

7. *How many zones does a well inject in?* - Analogous to CQ 5, but looking for instances of `Zone` with a `injects_in` relation with the instance of `Well`.

8. *Does a well have a gas lift valve installed in it?* - Check whether there is an instance of `Gas Lift Valve` that stands in a `installed_in` relation with the instance of `Well` of interest.

9. *What is the type of the well?* - Look for which is the subtype of `Well` (e.g., `Producer`, `Injector`) that the particular well in question instantiates.

10. *What is the type of fluid that is being injected through the well?* - Look for the subtype of `Injector` (i.e., `Water Injector Well` or `Gas Injector Well`) that the particular well instantiates. Those types are defined as having all its instances related to some instance of `Water` or `Gas` via `is_injecting` relation.

11. *How many wells are connected to a platform?* - Look for the instances of `Well` that stand in a `connected_to` with the chosen instance of `Platform`.

12. *Which wells are connected to a manifold?* - Look for the instances of `Well` that are `connected_to` the desired `Manifold`.

13. *Which are the types of pressure associated with a well?* - Look for all the subtypes of `Pressure` that have some instance that `inheres_in` an entity that is `component_of` the chosen instance of `Well`.

14. *In which basin is the field located?* - Look for the instance of `Field` that has a relation `located_in` with the desired instance of `Basin`.

15. *To which field is the platform associated?* - Look for the instance of `Field` that has a relation `comprehends` with an instance of `Reservoir` that `RO:has_part` some instance of `Zone` to which an instance of `Well` `produces_from` or `injects_in` and such well is `connected_to` the desired instance of `Platform`.

16. *Questions 16 to 35* - Look for the value of the instance of the desired `BFO:Quality` that inheres in the instance of desired `IOF:Material Artifact`.

6.2 Ontology applicability based on connection, component and location relations

Besides using competency questions, we can evaluate ontology applicability through scenarios where we can apply the ontology to resolve a particular problem in a real-world setting. One of the benefits of O3PO is the capability to define the components and connected parts of the facilities in a plant. The properties of its components can infer properties from a whole. Related to this particular use for O3PO, we have two stereotypical scenarios in which a user would benefit from using the ontology.

#1 The vast number of assets in a production plant makes managing information during the operation life cycle difficult. For example, an engineer working in automation and control needs to map every piece of equipment in the plant. Due to equipment failure, changes in field development strategy, or maintenance schedules, there are frequent modifications in equipment functioning inside a production plant. It would be desirable if, after a maintenance procedure on a particular piece of equipment, changes in the components of the plant show up on the overall information system.

To deal with this scenario, one can take advantage of the compositional structure provided by the ontology through the `component_of` relation combined to the network of components linked by `connected_to` relations. For example, to identify all the equipment that compose a production plant, we can identify the instances of `Platform` that make up the plant. From that, given the transitive character of the relations of `component_of` and `connected_to`, we only have to look for all the instances of `Material Artifact` that have some `component_of` relation with each of the chosen instances of `Platform` or with any instance of `Material Artifact` that is `connected_to` any of the selected instances of `Platform`.

It is noteworthy that, with this arrangement, we only need to feed a knowledge base backed up by this ontology with local information to allow us to infer global information. For example, whoever installs a pump in a well only needs to establish a `component_of` relation between the corresponding instances of `Pump` and the `Well` in the knowledge base, without worrying about to which instance `Platform` the `Well` is currently `connected_to`. Analogously, whoever is responsible for the connection between a `Riser` and a `Platform` only needs to assert a `component_of` relation between the corresponding instances of `Riser` and `Platform` in the knowledge base, disregarding the instance of `Well` the `Riser` comes from or the pieces of equipment that compose it. Even so, anyone querying the knowledge base could find out to which instance of `Platform` the installed `Pump` is connected and of which production plant it is a part.

#2 Besides knowing all the parts of the plant, it is important to know the connection path of the components of the subsea production system, such as in a process flow diagram used by flow assurance engineers. Basically, the engineer wants to know where a particular valve is installed, considering the entire facility. Also, it is useful to be aware of the ways in which connected items can affect one another.

To handle the issue of locating an item, the `component_of` relation between instances of `bfo:material entity` is not enough to fully provide the location of a particular asset. So, we could improve the location of an item by adding contextual information regarding its position in relation to neighboring items in the chain of connections. To do that, one could identify the things that are `connected_to` the entity of interest.

Besides that, to some extent, the ontology also allows representing influences between equipment, such as when a change in an item modifies the status of another entity external to it. For example, if the choke valve of a Christmas tree on top of a well is closed, it will also close the well. However, it is not the closing of just any choke valve that will close a well, but only those suitably related to it. Making this differentiation would be useful, for example, for simulating the possible behavior of the diverse equipment in the plant. In this respect, our ontology can help by allowing the characterization of the conditions in which an artifact would affect another. In this particular case, the external valve whose closing affects the well could be represented as an instance of `choke valve` that is `component_of` an instance of `christmas Tree` that is `connected_to` an instance of `well`. This could be used to define a rule in a system associating modifications on the value of `choke position` property of the `choke valve` to modifications of the `well status` property of the `well`.

6.3 Application

As usual for a deep-water offshore field, many data sources are distributed in the production plant due to the number of sensors in the wells and equipment. The data acquired in each measurement performed by a sensor is usually stored in a CSV file next to its timestamp, forming a time series reflect the historical values for the measured property.

A difficulty that engineers have is tying such data to the corresponding physical assets in the plant. For example, if an engineer wants to gather all pressure time series from a given well, s/he has to inspect the label of each CSV file to check whether it corresponds to a property of interest of the desired well. It would be desirable to have the means to obtain all the time series from the well without looking for each data file spread on the company's data stores.

To handle this, we could use our ontology to annotate the files with reference to the type of property that was measured, the instance of the entity that has the property, and the unit of measurement. Moreover, given the representation of relations between artifacts, it would be possible to perform more elaborated semantic searches on the data. For example, it would allow searching files referring to all the pressures associated with all the wells connected to a given manifold - information that would probably not be present in the label of the data file.

To further validate the ontology, we considered three specific use cases derived from the motivating scenario described earlier which industry experts validated. The use cases consider production data from an Early Production System (EPS) in the context of the extended well test described in (ROVINA et al., 2019). The time-series data comes from an offshore petroleum production plant in the Mero field, from Brazilian Pre-Salt, operated by Libra Consortium.

In this scenario, there is a production well and an injection well, both connected to a Floating Production Storage Offloading (FPSO) rig, a specific type of platform used in ultradeep water fields. Figure 6.1 shows a scheme adapted from (ROVINA et al., 2019) that relates entities of the scenario and their properties to classes in the O3PO ontology.

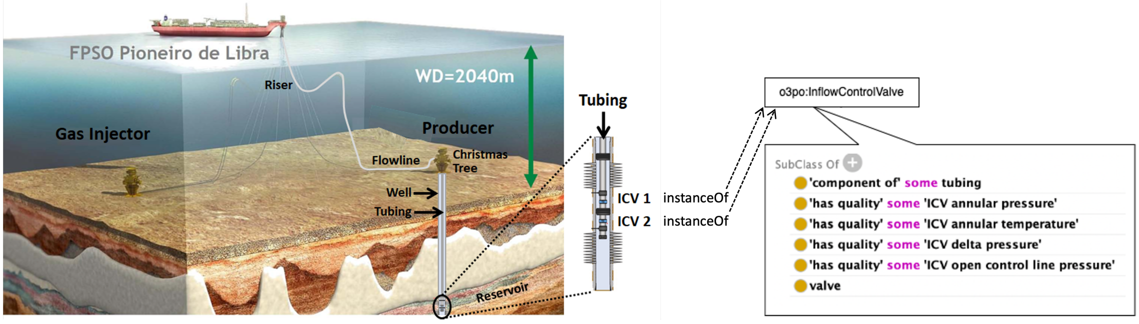


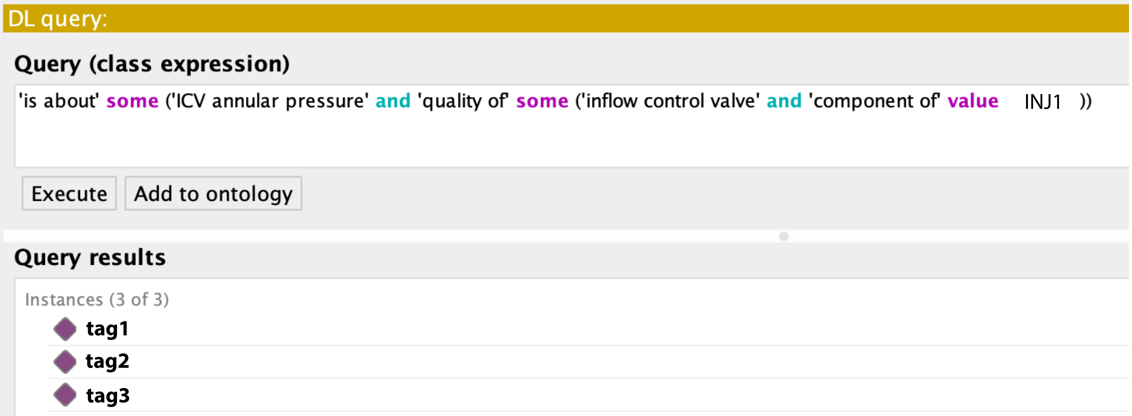
Figure 6.1 – A simplistic scheme displaying the main entities in an offshore petroleum production system. Produced by Régis Romeu.

The instantiation process of O3PO involves the addition of instances to the OWL file, representing real-world entities and relationships in the offshore petroleum production domain. These instances serve for reasoning and as the basis for querying information using DL queries, allowing users to retrieve and analyze relevant data within the Protege environment. These queries were performed on the instances added to the OWL file, which were used to illustrate the ontology's functionality in the three use cases discussed in the thesis.

We used two concepts from the Information Artifact Ontology (IAO) for these use cases: the relation `isAbout` and the universal `symbol`. We created a class under `symbol` called `tag` which refers to the names of the files where time series data is stored. In this scenario, each `tag` `isAbout` a particular instance of a `quality`. An example would be an instance of `tag` with the name **Wellhead-pressure-well-1** which `isAbout` and instance of the `quality` `wellhead pressure` that `inheresIn` **Well1**, an instance of the class `well`.

6.3.1 Use case 1: Investigation of production loss

An engineer wants to investigate the tightness of the ICV valves in a particular injection well. To do this, he must compare the annular pressures in each ICV within the given well. Figure 6.2 shows a DL-query that provides the files' tags corresponding to the annular pressure time series for all ICVs in the well. Notice that the user doesn't need to know how many ICVs there are or the names of the files particularly. With the tags, a visualization application can display the information according to the user's needs.



'is about' some ('ICV annular pressure' and 'quality of' some ('inflow control valve' and 'component of' value INJ1))

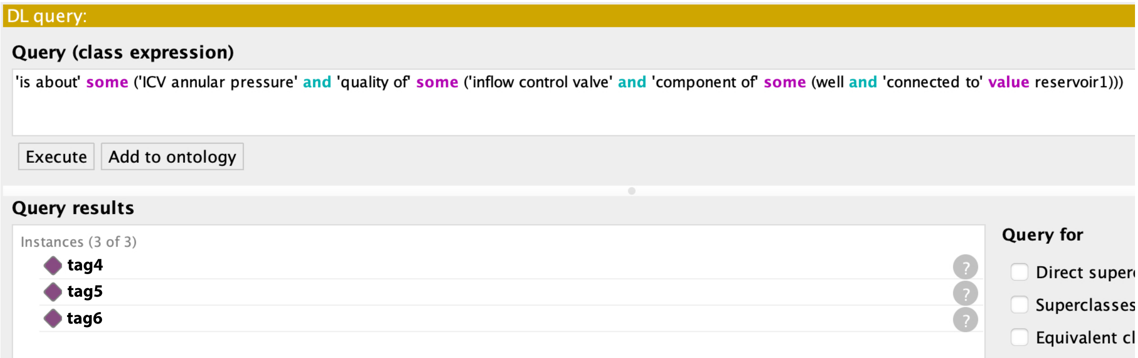
Instances (3 of 3)
tag1
tag2
tag3

Figure 6.2 – DL-query corresponding to use case 1.

6.3.2 Use case 2: Reservoir connectivity

A reservoir engineer wants to know the influence of an injection well on the pressure in a production well to evaluate the reservoir connectivity between the wells.

Figure 6.3 shows a DL-query that looks for the tags corresponding to ICV annular pressures in wells connected to a particular reservoir. The ontology can provide the appropriate tags for this particular necessity.



```
'is about' some ('ICV annular pressure' and 'quality of' some ('inflow control valve' and 'component of' some (well and 'connected to' value reservoir1)))
```

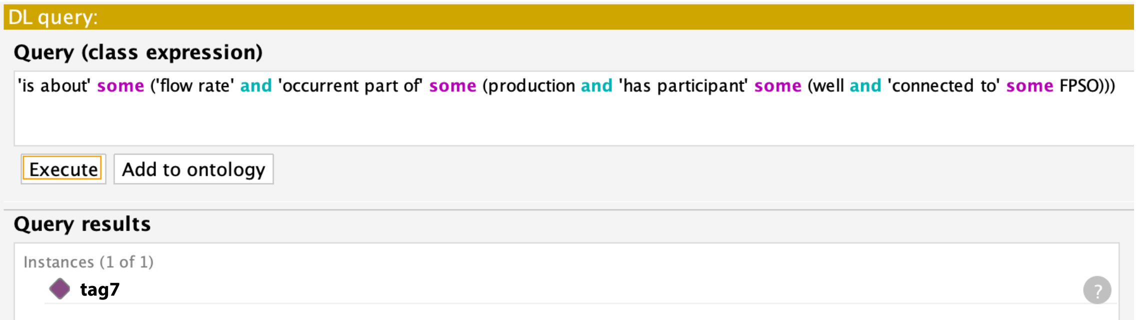
Instances (3 of 3)
- tag4
- tag5
- tag6

Figure 6.3 – DL-query corresponding to use case 2.

6.3.3 Use case 3: Platform production

For business purposes, a field manager needs to know how much oil is being produced by a particular platform. To do that, he needs to calculate using the files corresponding to flow rates for all production wells connected to the platform.

Figure 6.4 shows the corresponding DL-query providing the tag about the flow rate of the only production well instantiated in the ontology. Without the user needing to know how many wells connect to a specific platform, this DL query can provide tags about any of them.



```
'is about' some ('flow rate' and 'occurrent part of' some (production and 'has participant' some (well and 'connected to' some FPSO)))
```

Instances (1 of 1)
- tag7

Figure 6.4 – DL-query corresponding to use case 3.

7 CONCLUSIONS

This research contributes to the broader goal of the PeTwin project, which is to create a semantic-level model that connects data with the simulation system, making it easier to operate a digital twin of an offshore petroleum production plant. In light of this, a domain ontology has been developed to provide a standardized vocabulary for engineers and information technology professionals working with offshore petroleum production plants. The ontology addresses the challenge of integrating data from multiple sources and is based on industry-oriented requirements. It reuses a top-level ontology, BFO, two core ontologies, GeoCore, and a core ontology developed by the Industry Ontology Foundry (IOF). The ontology is grounded in various resources, resulting in a well-founded ontology that can be useful in various applications in the petroleum production domain.

This research has focused on defining the entities that are present in an offshore petroleum production plant. Particularly, it focuses on the set of entities that are present in the oil path between the reservoir and the platform. The main scope of the ontology is on the pertaining continuants in the plant, even though an initial treatment to ocurrent has been shown. Also, connection and component relations were modeled to give a semantic treatment to the relations between equipment in the plant. When considering a digital twin of an oil plant, which is supposed to be a virtual replica of a physical plant, the formal definition of the entities, universals, defined classes, and relations are essential for an effective semantic layer to operational streaming data.

Through an extensive requirements specification activity, a set of competency questions and terms were gathered. The development of the ontology followed the specified requirements. The O3PO ontology delivers natural-language, semi-formal, and formal definitions of terms commonly used in offshore petroleum production plants. From this, different parties involved in petroleum enterprises can refer to the same terms when communicating amongst themselves. The formal definitions will restrain the meaning of the entities, and this will provide a framework for a system to map different databases to the same entities and their inhering properties.

Furthermore, the amount of data in a usual offshore production plant is substantial. Without a semantic level to this data, the possibilities of reasoning over data and efficient data querying would be challenging. The ontology present in this paper can serve as a basis for the semantic layer between data and equipment properties pertaining to a plant. During ontology development and with data instantiation, semantic reasoning was accomplished and proved the usefulness of the semantic model.

Besides validating the ontology by showing the capability of representing the competency question and the expected answers with the modeled entities, some use cases from a real-world scenario in an offshore petroleum field in Brazil were considered for ontology application and validation.

Moreover, this ontology can be reused by other ontologies in related domains, as most of the model's terms are generic. We hope that this research will contribute to future initiatives in knowledge engineering in the oil and gas industry and possibly in other industrial domains as well.

7.1 Limitations

O3PO provides many benefits to information management in offshore petroleum production plants. However, there are some limitations to the capabilities of the ontology.

- Even though O3PO provides the means to tie values of properties (e.g., pressure and temperature) to material entities, the ontology cannot handle missing data, erroneous data, and other problems related to the observation of properties and not just to the properties themselves.

- The ontology, as of yet, does not handle measurements and information content entities in full, which are fundamental to production surveillance and safety maintenance.

- Currently, the treatment of connections between the components of a plant could be more robust (e.g., not dealing with the effect of closed valves, obstructions, and other issues on the fluid path).

- There is no explicit treatment to harmonize the common configuration view of the petroleum standards provided by ISO 15926, CFIHOS, and OSDU with the production view of O3PO. This may affect the understanding of the users related to the expected role of O3PO in the data interoperability scenario of petroleum plants. We intend to solve this limitation in further steps of our project.

7.2 Future work

Current and future progress in ontology development can be seen in a public repository in GitHub (https://github.com/BDI-UFRGS/O3POntology). Future work will consider the limitations described in the previous section to enhance the ontological model further and contribute to a richer and more useful semantic layer in a digital twin of offshore petroleum production plants.

Besides dealing with the listed limitations, from the research performed so far, it's clear that there is a necessity to treat the dynamic functioning of the plant. By effectively mapping the ocurrents in petroleum production and giving the means for the dispositions of the entities in a plant to be realized, new causality relations can be defined. Such a model of the dynamic behavior of a plant would be useful later for behavior prediction and production optimization, given that the data from a real plant would be available.

There is potential for extending the underlying DL to support a Natural Language interface. This extension would facilitate more intuitive user interactions with the ontology by enabling queries in natural language. Although incorporating such an interface presents certain challenges, including the complexity of natural language processing and the need for accurate semantic interpretation, the benefits of improved usability and accessibility make it a promising avenue for future research and development.

Furthermore, incorporating topside equipment into the model would not only enhance the ontology but also yield a more all-encompassing semantic artifact within the realm of offshore petroleum production. To effectively model topside facilities, the set of ontologies provided in OntoCAPE (MARQUARDT et al., 2010), along with its approach to connections and systems, can prove invaluable for this purpose.

REFERENCES

ABEL, M. et al. "What Is a Well" in an ontology-based standard. In: **Petroleum Network Education Conferences (PNEC)**. [S.l.: s.n.], 2017.

American Petroleum Institute (API). **Glossary of Oilfield Production Terminology (GOT) (Definitions and Abbreviations)**. First edition. [S.l.]: American Petroleum Institute (API), 1988.

ARP, R.; SMITH, B. Function, Role, and Disposition in Basic Formal Ontology. **Nature Precedings**, 6 2008. ISSN 1756-0357.

ARP, R.; SMITH, B.; SPEAR, A. D. **Building Ontologies with Basic Formal Ontology**. [S.l.]: MIT Press, 2015.

ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, 2000. ISSN 1546-1718. Available from Internet: <https://doi.org/10.1038/75556>.

BILONG, W.; FUCAI, B.; JING, Y. Build Petroleum Exploration and Production Domain Ontology from Reference Models. **2010 Second International Workshop on Education Technology and Computer Science**, v. 1, p. 502–505, 2010. ISSN VO - 1. Available from Internet: <files/4292/Bilongetal.-2010-BuildPetroleumExplorationandProductionDomain.pdffiles/4593/bilong2010.pdf.pdf>.

BORGO, S. et al. DOLCE: A descriptive ontology for linguistic and cognitive engineering. **Applied Ontology**, IOS Press, v. 17, p. 45–69, 2022. ISSN 1875-8533.

BORST, W. N. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. Thesis (PhD) — University of Twente, Netherlands, 9 1997.

CALVANESE, D. et al. Optique: OBDA solution for big data. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2013. ISBN 9783642412417. ISSN 03029743.

CAMERON, D. B. et al. A semantic systems engineering framework for zero-defect engineering and operations in the continuous process industries. **Frontiers in Manufacturing Technology**, Frontiers Media SA, v. 2, 9 2022.

CARBONERA, J. L.; ABEL, M.; SCHERER, C. M. S. Visual interpretation of events in petroleum exploration: An approach supported by well-founded ontologies. **Expert Systems with Applications**, v. 42, n. 5, p. 2749–2763, 2015. ISSN 09574174 (ISSN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84918837822&doi=10.1016%2fj.eswa.2014.11.021&partnerID=40&md5=b743a070c5a9e915630314f6c1a7024afiles/4597/carbonera2015.pdf.pdf>.

CHAN, C. W. The Knowledge Modeling System and its application. In: **Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513)**. [S.l.: s.n.], 2004. v. 3, p. 1353–1356.

CHAN, C. W. An expert decision support system for monitoring and diagnosis of petroleum production and separation processes. **Expert Systems with Applications**, v. 29, n. 1, p. 131–143, 2005. ISSN 09574174 (ISSN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-16244389339&doi=10.1016%2fj.eswa.2005.01.009&partnerID=40&md5=0fbf71a212baa58d983da42d0ddc6a0dfiles/4309/chan2005.pdf.pdf>.

CHAN, C. W.; PENG, Y.; CHEN, L. L. Knowledge acquisition and ontology modelling for construction of a control and monitoring expert system. **International Journal of Systems Science**, v. 33, n. 6, 2002. ISSN 00207721.

CHEN, L.-L. et al. Ontology design and its application in the petroleum Remediation Domain. **6th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2000**, Springer Verlag, ["University of Regina, Faculty of Engineering, Saskatchewan, S4S 0A2, Canada", "University of Regina, Department of Computer Science, Saskatchewan, S4S 0A2, Canada"], v. 2112, p. 16–23, 2001. ISSN 03029743 (ISSN); 3540425977 (ISBN); 9783540454083 (ISBN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84943249559&doi=10.1007%2f3-540-45408-x_3&partnerID=40&md5=174ae79248adb11b8f14799318116b1c>.

CICCONETO, F. et al. GeoReservoir: An ontology for deep-marine depositional system geometry description. **Computers and Geosciences**, Elsevier Ltd, v. 159, 2 2022. ISSN 00983004.

COSENTINO, L. **Integrated reservoir studies**. [S.l.]: Editions Technip, 2001.

DROBNJAKOVIC, M. et al. The Industrial Ontologies Foundry (IOF) Core Ontology. 2022.

DU, R. et al. Study on ontology-based knowledge construction of petroleum exploitation domain. **2010 International Conference on Artificial Intelligence and Computational Intelligence, AICI 2010**, v. 2, p. 42–46, 2010. ISSN 9780769542256 (ISBN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-78651457463&doi=10.1109%2fAICI.2010.132&partnerID=40&md5=1729502819542806392634238c52126ffiles/4605/du2010.pdf.pdf>.

EBRAHIMIPOUR, V.; YACOUT, S. ISO 15926. In: **Ontology Modeling in Physical Asset Integrity Management**. [S.l.]: Springer International Publishing, 2015. p. 1–16.

GARCIA, L. F. et al. The GeoCore ontology: A core ontology for general use in Geology. **Computers & Geosciences**, Pergamon, v. 135, p. 104387, 2 2020. ISSN 0098-3004.

GRIEVES, M. **Digital Twin: Manufacturing Excellence through Virtual Factory Replication**. [S.l.], 2014.

GRIEVES, M.; VICKERS, J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: **Transdisciplinary perspectives on complex systems**. [S.l.]: Springer, 2017. p. 85–113.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, Elsevier, v. 5, n. 2, p. 199–220, 1993.

GUARINO, N.; OBERLE, D.; STAAB, S. What Is Ontology? In: **Handbook on Ontologies, International Handbooks on Information Systems**. [S.l.: s.n.], 2009. ISBN 978-3-540-70999-2.

International Association of Oil & Gas Producers (IOGP). **Capital Facilities Information Handover Specification (CFIHOS) - Specification Document**. [S.l.], 2020. Available from Internet: <https://www.jip36-cfihos.org/cfihos-standards/>.

International Association of Oil & Gas Producers (IOGP). **CFIHOS - Scope and Procedure**. [S.l.], 2020.

International Organization for Standardization. **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 1: Overview and fundamental principles (ISO 15926-1:2004)**. 2004. Available from Internet: <https://www.iso.org/standard/29556.html>.

International Organization for Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 4: Initial reference data (ISO/TS Standard No. 15926-4)**. Geneva: International Organization of Standardization (ISO), 2019. Available from Internet: <https://www.iso.org/standard/73830.html>.

International Organization of Standardization. **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 13: Integrated asset planning life-cycle (ISO 15926-13:2018)**. 2018. 1–96 p. Available from Internet: <https://www.iso.org/standard/70694.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 2: Data model (ISO 15926-2:2003)**. 2003. 1–225 p. Available from Internet: <https://www.iso.org/standard/29557.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 3: Reference data for geometry and topology (ISO/TS 15926-3:2009)**. 2009. 1–29 p. Available from Internet: <https://www.iso.org/standard/42875.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 7: Implementation methods for the integration of distributed systems: Template methodology (ISO/TS 15926-7:2011)**. 2011. 1–116 p. Available from Internet: <https://www.iso.org/standard/52455.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 8: Implementation methods for the integration of distributed systems: Web Ontology Language (OWL) implementation (ISO/TS 15926-8:2011)**. 2011. 1–50 p. Available from Internet: <https://www.iso.org/standard/52456.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 6: Methodology for the development and validation of reference data (ISO/TS 15926-6:2013)**. 2013. 1–59 p. Available from Internet: <https://www.iso.org/standard/56819.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 11: Methodology for simplified industrial usage of reference data (ISO/TS 15926-11:2015)**. 2015. 1–67 p. Available from Internet: <https://www.iso.org/standard/57859.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 12: Life-cycle integration ontology represented in Web Ontology Language (OWL) (ISO/TS 15926-12:2018**). 2018. 1–67 p. Available from Internet: <https://www.iso.org/standard/70695.html>.

International Organization of Standardization (ISO). **Industrial automation systems and integration — Integration of life cycle data for process plants including oil and gas production facilities — Part 10: Conformance testing (ISO 15926-10:2019)**. 2019. 1–41 p. Available from Internet: <https://www.iso.org/standard/68200.html>.

KHARLAMOV, E. et al. Finding Data Should be Easier than Finding Oil. In: **Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018**. [S.l.: s.n.], 2019.

KHARLAMOV, E. et al. Finding Data Should be Easier than Finding Oil. **2018 IEEE International Conference on Big Data, Big Data 2018**, p. 1747–1756, 2019. ISSN 9781538650356 (ISBN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062592946&doi=10.1109%2fBigData.2018.8622035&partnerID=40&md5=7d2417d1486e19055fe3043fcb2e5701files/4629/kharlamov2018.pdf.pdf>.

KRITZINGER, W. et al. Digital Twin in manufacturing: A categorical literature review and classification. **IFAC-PapersOnLine**, Elsevier B.V., v. 51, n. 11, p. 1016–1022, 6 2018. ISSN 24058963.

KUKKONEN, V. et al. An ontology to support flow system descriptions from design to operation of buildings. **Automation in Construction**, Elsevier B.V., v. 134, 2 2022. ISSN 09265805.

MARQUARDT, W. et al. **OntoCAPE - A Re-Usable Ontology for Chemical Process Engineering**. [S.l.]: Springer-Verlag Berlin Heidelberg, 2010. ISBN 978-3-642-04654-4.

MATHIESON, D.; MEEHAN, D. N.; POTTS, J. The End of Petroleum Engineering as We Know It. In: . [S.l.]: Society of Petroleum Engineers (SPE), 2019.

MENZEL, C. Reference Ontologies-Application Ontologies: Either/Or or Both/And?. In: **KI Workshop on Reference Ontologies and Application Ontologies**. [S.l.: s.n.], 2003. v. 16, p. 1–10.

MINERVA, R.; LEE, G. M.; CRESPI, N. Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models. **Proceedings of the IEEE**, v. 108, n. 10, 2020. ISSN 15582256.

MUSEN, M. A. The protégé project: a look back and a look }forward. **AI Matters**, v. 1, n. 4, p. 4–12, 2015. Available from Internet: <https://doi.org/10.1145/2757001.2757003>.

NEGRI, E.; FUMAGALLI, L.; MACCHI, M. A Review of the Roles of Digital Twin in CPS-based Production Systems. **Procedia Manufacturing**, Elsevier B.V., v. 11, p. 939–948, 2017. ISSN 23519789.

NIMMAGADDA, S. L.; DREHER, H.; ... Data warehouse structuring methodologies for efficient mining of Western Australian petroleum data sources. **INDIN'05. 2005 3rd IEEE ...**, 2005. Available from Internet: <https://ieeexplore.ieee.org/abstract/document/1560446/files/4653/data-warehouse-structuring-methodologies-for-effilcient-mining-o.pdf.pdf>.

NIMMAGADDA, S. L.; DREHER, H. Ontology based data warehouse modelling - A methodology for managing petroleum field ecosystems. In: **2008 2nd IEEE International Conference on Digital Ecosystems and Technologies, IEEE-DEST 2008**. [S.l.: s.n.], 2008. p. 221–228. ISBN 1424414903.

NIMMAGADDA, S. L.; DREHER, H.; RAJAB, M. R. Ontology-based warehouse time-depth data modelling framework for improved seismic interpretation in onshore producing basins. **International Petroleum Technology Conference 2007, IPTC 2007**, ["Wafra Joint Operations, Kuwait Gulf Oil Company, Kuwait", "Curtin University of Technology, Perth, WA, Australia", "Kuwait Gulf Oil Company, Kuwait"] Dubai, v. 1, p. 477–486, 2007. ISSN 9781615673360 (ISBN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-70450042739&partnerID=40&md5=887118391310bbac7c1e759a820af443>.

OBRST, L. Ontological Architectures. In: HEALY, M. P. R.; ACHILLES, K. (Ed.). **Theory and Applications of Ontology: Computer Applications**. Dordrecht: Springer Netherlands, 2010. p. 27–66. ISBN 978-90-481-8847-5. Available from Internet: <https://doi.org/10.1007/978-90-481-8847-5_2>.

OTTE, J. N.; BEVERLEY, J.; RUTTENBERG, A. BFO: Basic Formal Ontology. **Applied Ontology**, IOS Press, v. 17, p. 17–43, 2022. ISSN 1875-8533.

PARRY, W. T.; HACKER, E. A. **Aristotelian logic**. [S.l.]: Suny Press, 1991.

PONOMAREV, K.; KUDRYASHOV, N.; POPELNUKHA, N. Main Principals and Issues of Digital Twin Development for Complex Technological Processes. In: . [S.l.: s.n.], 2017. p. 523–528.

PRESTES, E. et al. Towards a core ontology for robotics and automation. **Ubiquitous Robotics**, v. 61, n. 11, p. 1193–1204, 2013. ISSN 0921-8890. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S0921889013000596>.

Professional Petroleum Data Management Association (PPDM). **What is a Well?** Calgary, CA: PPDM Association, 2014. Available from Internet: <https://ppdm.org/ppdm/PPDM/Standards/What_is_a_Well/PPDM/What_is_a_Well.aspx>.

Professional Petroleum Data Management Association (PPDM). **What is A Completion?** 2018. Available from Internet: <https://ppdm.org/PPDM/PPDM/Member/What_is_a_Completion.aspx>.

REES, R. V. Clarity in the usage of the terms ontology, taxonomy and classification. **CIB REPORT**, CIB, v. 284, n. 432, p. 1–8, 2003.

ROVINA, P. S. et al. Extend well test EWT libra project overview and technological highlights. In: **Proceedings of the Annual Offshore Technology Conference**. [S.l.]: Offshore Technology Conference, 2019. v. 2019-May. ISBN 9781613996416. ISSN 01603663.

SANTOS, N. O. et al. Towards an Ontology of Offshore Petroleum Production Equipment. In: SANFILIPPO, E. et al. (Ed.). **FOMI 2022 - Proceedings of the 12th International Workshop on Formal Ontologies Meet Industry**. Tarbes, France: CEUR-WS, 2022.

Semantic Web Company. **From Taxonomies over Ontologies to Knowledge Graphs**. 2014. Available from Internet: <https://semantic-web.com/from-taxonomies-over-ontologies-to-knowledge-graphs/>.

SEPPÄLÄ, S.; RUTTENBERG, A.; SMITH, B. The Functions of Definitions in Ontologies. In: FERRARIO, R.; KUHN, W. (Ed.). **Formal Ontology in Information Systems. Proceedings of the Ninth International Conference (FOIS 2016)**. [S.l.]: IOS Pres, 2016. p. 37–50.

SHAFTO, M. et al. Modeling, simulation, information technology & processing roadmap. **National Aeronautics and Space Administration**, 2012.

SHANKAR, R. et al. Knowledge-based Digital Twin for Oil and Gas 4.0 Upstream Process: A System Prototype. In: **Proceedings of the 2022 IEEE International Conference on Internet of Things and Intelligence Systems, IoTaIS 2022**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2022. p. 344–350. ISBN 9798350396454.

SILVA, P. F. et al. How do specialists express risks: an applied ontology for the oil & gas domain. In: **ONTOBRAS**. [S.l.: s.n.], 2021. p. 114–125.

SKALLE, P.; AAMODT, A. Petrol 18 946: Downhole failures revealed through ontology engineering. **Journal of Petroleum Science and Engineering**, v. 191, p. 107188, 2020. ISSN 0920-4105. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S0920410520302758files/4641/skalle2020.pdf.pdf>.

SMITH, B. On classifying material entities in Basic Formal Ontology. In: **Interdisciplinary Ontology. Proceedings of the Third Interdisciplinary Ontology Meeting**. [S.l.: s.n.], 2012.

SMITH, B. **Ontology as Product-Service System Lessons Learned from GO, BFO and DOLCE**. [S.l.], 2019. Available from Internet: <http://obofoundry.org.>

SMITH, B. et al. **Basic Formal Ontology 2.0 Specification and User's Guide**. [S.l.], 2015. 1–104 p. Available from Internet: <https://github.com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference.pdf>.

SMITH, B. et al. A First-Order Logic Formalization of the Industrial Ontology Foundry Signature Using Basic Formal Ontology. In: **Proceedings of the Joint Ontology Workshops (JOWO), Graz**. [S.l.: s.n.], 2019.

SMITH, B.; CEUSTERS, W. Aboutness: Towards foundations for the information artifact ontology. In: **CEUR Workshop Proceedings**. [S.l.: s.n.], 2015. v. 1515. ISSN 16130073.

SMITH, B.; CORPORATION, O. W. T. M. Against idiosyncrasy in ontology development. **4th International Conference on Formal Ontology in Information Systems, FOIS 2006**, v. 150, p. 15–26, 2006. ISSN 09226389 (ISSN); 1586036858 (ISBN); 9781586036850 (ISBN). Available from Internet: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-78249290752&partnerID=40&md5=1a6451e76570e25dd3210c51b87c2fccfiles/4342/SmithandCorporation-2006-Againstidiosyncrasyinontologydevelopment.pdf>.

SOUZA, V. et al. A Digital Twin Architecture Based on the Industrial Internet of Things Technologies. In: **2019 IEEE International Conference on Consumer Electronics, ICCE 2019**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2019. ISBN 9781538679104.

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.

SUÁREZ-FIGUEROA, M. C. NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse. Facultad de Informática (UPM), 6 2010.

SUÁREZ-FIGUEROA, M. C. et al. Ontology engineering in a networked world. **Ontology Engineering in a Networked World**, Springer Berlin Heidelberg, p. 1–435, 1 2012.

TAO, F. et al. Digital Twin in Industry: State-of-the-Art. **IEEE Transactions on Industrial Informatics**, IEEE Computer Society, v. 15, n. 4, p. 2405–2415, 6 2019. ISSN 15513203.

TAO, F.; ZHANG, M. Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing. **IEEE Access**, Institute of Electrical and Electronics Engineers Inc., v. 5, p. 20418–20427, 6 2017. ISSN 21693536.

TAO, F. et al. Digital twin driven prognostics and health management for complex equipment. **CIRP Annals**, Elsevier USA, v. 67, n. 1, p. 169–172, 6 2018. ISSN 17260604.

TUDORACHE, T. Ontology engineering: Current state, challenges, and future directions. **Semantic Web**, IOS Press, v. 11, n. 1, p. 125–138, 2020. ISSN 22104968.

VOELL, C. et al. How Digital Twins Enable the Next Level of PLM–A Guide for the Concept and the Implementation in the Internet of Everything Era. In: **IFIP International Conference on Product Lifecycle Management**. [S.l.: s.n.], 2018. p. 238–249.

WACHE, H. et al. Ontology-Based Integration of Information-A Survey of Existing Approaches. In: **OIS@ IJCAI**. [S.l.: s.n.], 2001.

WHITLEY, S. et al. **Understanding the OSDU Opportunity - A Schlumberger Guide to the OSDU Data Platform**. [S.l.]: Schlumberger, 2020.

WOODS, C. et al. **An ontology for maintenance activities and its application to data quality**. [S.l.].

YUAN, J.; LI, H. Research on the standardization model of data semantics in the knowledge graph construction of Oil&Gas industry. **Computer Standards & Interfaces**, Elsevier BV, v. 84, p. 103705, 3 2023. ISSN 09205489.

APPENDIX A — RESUMO EXPANDIDO

A integração de dados de múltiplas fontes é um problema persistente em diversas indústrias comerciais, incluindo o setor de petróleo. Com o avanço da tecnologia no setor petrolífero, houve um aumento no número de prestadores de serviços e, consequentemente, um aumento no número de silos de informações. Esses silos de informações dificultam o acesso e a integração de dados, levando a uma perda de tempo significativa na busca por informações em diferentes bancos de dados.

Para enfrentar esse desafio, os autores apresentam a Offshore Petroleum Production Plant Ontology (O3PO), uma ontologia de domínio projetada para representar entidades em plantas de produção de petróleo offshore. A ontologia fornece um vocabulário de referência claro e bem definido para ajudar profissionais das áreas de engenharia e tecnologia da informação a rotular e ligar o monitoramento de plantas de produção, medições de simulação e instalações.

A O3PO é baseada em uma compilação abrangente de requisitos específicos da indústria, como casos de uso e questões de competência. A ontologia se baseia em ontologias de nível médio, tais como GeoCore e a ontologia central da indústria de fundição (IOF), enquanto adere à ontologia de nível superior do BFO. Além disso, os autores utilizaram recursos adicionais, como glossários específicos da indústria e ontologias relevantes, para construir uma ontologia de domínio abrangente e robusta.

Essa ontologia oferece conceitos universais, classes definidas e relações que podem ser aplicadas em diversos domínios, além de ser validada por meio da análise de dados de um campo petrolífero offshore no Brasil. Os autores demonstraram com sucesso a aplicabilidade prática da O3PO nesse cenário real.

Para confirmar a validade da ontologia, analisamos três casos de uso específicos, originados do cenário motivacional descrito, que foram validados por especialistas no assunto. Esses casos de uso levam em conta dados de produção de um Sistema de Produção Antecipada (SPA) no contexto do teste de poço estendido. Os dados de séries temporais são provenientes de uma unidade de produção de petróleo offshore no campo Mero, localizado no Pré-Sal brasileiro, operado pelo Consórcio Libra.

Em resumo, a O3PO é uma solução inovadora para o desafio da integração de dados em plantas de produção de petróleo offshore, que pode ser aplicada em diferentes domínios e indústrias para facilitar a integração e o acesso aos dados, economizando tempo e recursos valiosos.

Os avanços atuais e futuros no desenvolvimento da ontologia podem ser acessados em um repositório público no GitHub (https://github.com/BDI-UFRGS/O3POntology). Trabalhos futuros abordarão as limitações mencionadas para aprimorar ainda mais o modelo ontológico e enriquecer a camada semântica em um gêmeo digital de instalações de produção de petróleo offshore.

Fica claro que há uma necessidade de lidar com o funcionamento dinâmico das instalações. Ao mapear efetivamente os eventos na produção de petróleo e fornecer meios para representar as disposições das entidades em uma instalação, novas relações de causalidade podem ser estabelecidas. Um modelo que represente o comportamento dinâmico de uma planta seria útil para prever e otimizar a produção, desde que os dados de uma instalação real estejam disponíveis.