INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Application of Ontologies to Contracts for Difference Key Information Documents

Renato Valentim Figueira Franco

Mestrado em Gestão de Sistemas de Informação

Orientadores(as):

Doutora Ana Maria Carvalho de Almeida, Professora Associada, Iscte – Instituto Universitário de Lisboa.

Doutor Francisco José Rosales Santana Guimarães, Professor Associado Convidado, Iscte – Instituto Universitário de Lisboa.

October, 2023

Acknowledgements

This project would not have been possible without the guidance and orientation from Professor Ana Maria de Almeida from Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura, Lisboa, Portugal and CISUC - Centro de Informática e Sistemas da Universidade de Coimbra, Coimbra, Portugal, and Professor Francisco Guimarães from Instituto Universitário de Lisboa (ISCTE-IUL).

Resumo

Num cenário de crescimento paulatino do setor financeiro onde a partilha de conteúdos é fundamental, emergiu a necessidade de uma representação formal dos conteúdos chave dos documentos de informação fundamental destinados ao esclarecimento dos investidores não profissionais de instrumentos financeiros complexos. A presente investigação visa a criação de um sistema representacional de informação aplicado ao mercado de valores mobiliários, em especial o desenvolvimento de uma ontologia aplicada à análise de documentos de informação fundamental sobre produtos de investimento de retalho e de produtos de investimento com base em seguros. A ontologia permitirá contribuir para um entendimento comum deste domínio da realidade de tal modo que seja possível facilitar a sua comunicação entre humanos e sistemas, e entre sistemas, e procurará contribuir para o incremento do conhecimento científico no domínio de estudos ontológicos e a construção de uma ontologia especificamente aplicada ao presente domínio.

Aproveitando as capacidades da *Web Ontology Language*, a ontologia representa formalmente as classes, relações e semântica associados aos principais conteúdos nos documentos de informação fundamental de contratos por diferença, incluindo fatores de risco, dinâmica de mercado, quadro normativo e informações relacionadas com os investidores. Através de implementação de SPARQL e sistemas de inferência integrados no editor Protégé, afigura-se demonstrada a aplicabilidade prática da ontologia através da identificação de tendências e conformidade regulatória. Neste âmbito, a dissertação explora o processo de construção da ontologia, em particular as fases de representação de conhecimento, especificação, fase de implementação e a fase de avaliação.

Palavras-chave: Ontologia, Mercado de Capitais, Documentos de Informação Fundamental, CFD.

Abstract

In a rapidly evolving landscape of the financial sector, the need for a comprehensive and standardized data representation has become increasingly pertinent. This dissertation aims to produce an ontology designed to address the particularities associated with the complex financial instruments such as Contract for Differences (CFD), thereby providing a framework for knowledge representation in the financial domain. Leveraging the capabilities of the Web Ontology Language, the ontology formally represents the attributes, connections and semantics associated with the key information documents of CFDs, including risk factors, market dynamics, legal frameworks and investor-related information.

Through the implementation of SPARQL queries and reasonings systems built-in the Protégé editor, the ontology's practical applicability in facilitating trend identification and regulatory compliance is demonstrated. The dissertation explores the ontology's structure, namely the knowledge representation phase, the specification phase, the implementation phase and evaluation phase. The scope of this study is restricted to the analysis of capital markets ontologies in order to capture its structure, semantics and knowledge sharing between people and systems.

Keywords: Ontology, Financial Markets, CFD, PRIIPs, Key Information Documents.

Acronyms

AI	Artificial Intelligence
BdP	Banco de Portugal
BRO	Bank Regulation Ontology
CFD	Contract for Difference
CMVM	Comissão do Mercado de Valores Mobiliários
ESMA	European Securities and Markets Authority
EU	European Union
FIBO	Financial Industry Business Ontology
FRC	Financial Regulation Ontology
FSB	Financial Stability Board
KID	Key information Document
LEI	Legal Identity Identifier
LKIF	Legal Knowledge Interchange Format
NLP	Natural Language Processing
OTC	Over-the-counter
OWL	Ontology Web Language
PRIIPs	Packaged retail and insurance-based investment products
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
SRI	Synthetic Risk Indicator
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
XML	Extensible Markup Language

Table of contents

List of tables

List of figures

CHAPTER 1

Introduction

1. Scope delimitation

This investigation aims to create a representational information system applied to the securities market, particularly the development of an ontology applied to the analysis of the key information documents of Contracts for Difference.

The acronym "PRIIPs" stems from the English term Packaged Retail and Insurance-based Investment Products, as mentioned in Regulation (EU) no. 1286/2014 of the European Parliament and of the Council of 26 November, which approves the legal framework applicable to this matter.

Pursuant to the article 4(1) of the Regulation (EU) no. 1286/2014 of the European Parliament and of the Council of 26 November 2014, a PRIIP can be defined as an investment where, regardless of the legal form of the investment, the amount repayable to the retail investor is subject to fluctuations because of exposure to reference values or the performance of one or more assets which are not directly purchased by the retail investor. An investment in a PRIIP is considered a complex investment because the return on investment is based on the referenced value or price of its underlying asset. The underlying assets may vary from securities (namely, stocks or bonds) to commodities, such as metals or other goods. The performance of the underlying asset will determine the repayable amount to the investor (Perchet, Romain et. al., 2023).

Included within the scope of the PRIIPs definition exists many types of investment products, such as structured deposits, structured investment products, derivative instruments, among others. Considering the sheer number of types of investment products contained in the PRIIPs broad legal definition, a scope limitation for this dissertation is required. As such, this dissertation is limited to Contract for Difference (CFD), which is a derivative financial instrument and one of the most traded by retail investors. According to the most recent statistics published by the Portuguese Securities Markets Commission (CMVM), CFDs were the most traded instrument in the derivatives market as of June 2023 (2.9% of the total). By definition, a CFD is a financial derivative product that pays to the investor the difference in settlement price between the opening and closing of a certain transaction. It's considered a derivative financial product because the amount or price payable to the investor will depend on the performance or value of the underlying asset of the CFD (for example, stocks, bonds, commodities, among others).

A CFD, similarly to other derivative investment products, is a legal contract signed between an investor and a CFD issuer that stipulates that one of the contractual parties will pay the other contractual party the difference in the value of a financial product between the opening and closer duration period of the position. This means that, unlike typical financial instruments (e.g., stocks or bonds), the CFD is negotiated in over-the-counter (OTC) markets between the investors and the issuer. An OTC market is a venue where supply meets demand in a decentralized manner, where the buyer trades directly with the seller without the need for an intermediary entity.

In Europe, the CFD issuer must be registered with the national supervisory authority in order to issue such investment products and, pursuant to the EU legal framework, must draft a document for the investors, called Key Information Document. The national supervisory authority is responsible for regulating and supervising the marketing process cycle of these investment products.

The study of ontologies was identified and recognized as an important component for the Semantic Web (Berners-Lee, 2001), and there are numerous initiatives in the literature for the construction of ontologies applicable to various knowledge domains. Although the term "ontology" has its roots in the field of philosophical knowledge and assumes different conceptual configurations, when applied to the context of information systems, it refers to an artifact consisting of a specific vocabulary used to describe a particular reality, its objects, and the relationships established between them (Guarino, 1998). In practice, ontologies are conceptual models that clarify a specific semantic vocabulary in such a way as to eliminate inherent ambiguities, facilitating their communication and usage. In Javed et. al. refers that the NLP (Natural Language Processing) processes may be optimized and experience an increase of efficiency using ontologies (Javed, 2022). In fact, according to the literature, ontology-based NLP can be used in compliance management of software engineering processes to analyze standardized documents.

However, in the domain of financial markets, particularly in capital markets and especially in the field of complex investment products, the interest in developing ontologies has lacked enthusiasm. Despite this fact, it's critical to emphasize the growing interest of the stakeholders in the usage of natural language processing (NLP) models to analyze these types of documents (FSB, 2020). In fact, new regulatory and supervisory technologies are being developed and enhanced in order to improve the detection of fraud capabilities of the supervisors and other issues, such as regulatory reporting information, risk management and data collection. In this regard, in the beginning of 2019, the European Securities and Markets Authority (ESMA) began exploring the application of NLP to the analysis of more than 20,000 Key Information Documents, from 500 issuers, in 21 European Union (EU) languages (Armstrong and Harris, 2019).

The importance of building ontologies that facilitate the communication of certain concepts becomes evident when seeking to build information systems that process information more efficiently, accurately, and quickly. In addition, ontologies facilitate and promote interoperability between information systems, allowing for a common understanding of a particular domain of reality that is intended to be described in such a way that it can be communicated by humans who come into contact with it, as well as by information systems that use it (Pinto and Martins, 2004).

The knowledge base to represent in the KID ontology will be the key information documents used as a vehicle to convey information about PRIIPs to potential investors.

The KID is a pre-contractual information document, clearly distinct from promotional materials, which provides key information for non-professional investors to fully understand and compare the main characteristics, risks and returns, and costs of the investment product in which they intend to invest. The KID has a legally defined format and content, structured into sections. In addition to the initial sections entitled "Purpose" and "Product," and a final section called "Other Relevant Information," reserved for any additional relevant information, there are more sections legally provided for.

Despite these current developments, we’ve observed that there is scarcity of ontologies related to capital markets, specifically focusing on the key information documents. As a result of these recent developments and shift in perspective regarding the usage of supervisory and regulatory technologies, this dissertation seeks to create an ontology for the KIDs.

This study is divided into two main phases. The first phase involved examining the text of the KIDs in order to extract its structure and semantics as the main relevant information. Specifically, it was necessary to thoroughly examine all the sections of the document to enable the representation of the entire knowledge within this specific domain, i.e., the domain of the key information documents required by the legislation to be drafted by a manufacturer of PRIIPs.

The second phase involves the creation of the ontology in OWL language, using the Protégé editor, based on the knowledge acquired in the previous phase. The design of the KID ontology comprises four main activities: (i) specification, (ii) knowledge acquisition, (iii) implementation and (iv) evaluation.

2. Motivation

The motivation behind the choice of the dissertation topic stems from the demand to address challenges and enhance information processing within the securities market sector, specially concerning the analysis of KID for CFD in the context of the PRIIPs European regulation.

The scope limitations applied to CFDs stem from the fact that these investment products are widely traded in the securities market among retail investors. The statistics that are frequently published by the Portuguese Securities and Markets Commission highlights the substantial trading volume of CFDs, making them an interesting subject of study.

Furthermore, the dissertation draws motivation from the unexplored application of ontologies in the domain of financial markets, especially in the context of complex investment products like CFDs and the intricate relations from the European regulation on the subject. While ontologies have found extensive application in multiple knowledge domains, their utility in the financial sector, particularly for enhancing natural language processing models, has been relatively unexplored. The recent emergence of regulatory technologies, namely the application of NLP models to the analysis of legal documents, has created a compelling need for ontological representations of financial documents.

The motivation is accentuated by the evolving landscape of regulatory and supervisory technologies, exemplified by the ESMA exploration of NLP models for analyzing the KIDs. In this context, developing an ontology for KIDs is seen as valuable contribution to the growing interest in enhancing the detection of fraud, improving regulatory reporting and optimizing the risk management processes within the financial sector.

In summary, the dissertations motivation arises from the convergence of the factors mentioned above, including the complexity of PRIIPs and specifically the CFDs, the need for advanced analysis in the securities market, the underutilized potential of ontologies in the financial sector, in particular in the context of European regulation, and the evolving landscape of regulatory technologies by the regulators. The goal is to create an ontology that enhances information processing, facilitates regulatory compliance and enhances understanding and communication within the context of CFDs documents.

3. Potential contributions

The potential contributions that may arise from the KID ontology may impact both academia and industry in the financial sector and regulatory compliance.

The development of specialized ontologies, such as the ontology for KIDs within the context of CFDs, contributes to the field of ontology development, as we intended to demonstrate how ontological constructs can be applied to legal documents that govern complex investment products, enriching this discipline with practical application in the financial sector.

Furthermore, by creating an ontology that correctly represents the regulatory requirements, content and structure of KIDs, this dissertation directly addresses the practical needs of the financial sector, which can serve as preliminary tool for ensuring compliance with PRIIPs regulation concerning the KIDs.

Considering that ontologies may be used to enhance NLP models, this ontology provides a semantic foundation for such models to extract, interpret and compare information from KIDs, thereby contributing for a step towards regulatory compliance and CFD document analysis, including in artificial intelligence domain.

The KID ontology's structured representation of KIDs documentation facilitates knowledge sharing and interoperability between information systems. This can help reduce issues in the communication between stakeholders in the financial market, including regulators, financial institutions, investors and contributing to greater transparency and efficiency in the KID documents analysis, including in artificial intelligence domain.

An ontology development process with characteristics such as scalability and reusability in mind ensures that it can serve as an initial framework for representing other types of PRIIPs related documents or even broaden the horizon to other financial markets regulations, such as the legislation that regulates the content of a prospectus, thereby expanding the potential of reuse to practical applications of the ontology to other regulatory contexts.

Moreover, the KID ontology development process is aligned with the principles of Semantic Web, promoting the structured and semantically fertile representation of financial knowledge, especially in the context of the European financial regulations of complex products. Likewise, it has a point of intersection with the emerging field of regulatory technologies, which leverages technology to address regulatory issues caused, in part, by substantial volumes of data gathered from the market.

In the end, this investigation aims to contribute to multifaceted objectives, encompassing advancements in the ontology development process, regulatory compliance, NLP models, knowledge sharing, promotion of reusability and alignment with emerging international trends. These contributions can be justified by the demanding need for innovative solutions to navigate the challenges that emerge from regulatory compliance in the financial sector while harnessing the potential of semantic technologies and interdisciplinary collaboration.

4. Main dissertation goals and research questions

The present dissertation is centered around two main goals:

i. The identification of relevant concepts of the key information documents of CFDs; and

ii. Defining the classes, hierarchies, properties, and instances of the KID ontology to represent the CFD concepts.

The accurate identification of the terms within the KID ensures that the ontology aligns precisely with the relevant regulatory framework. Furthermore, the identification of these terms is akin to laying the semantic foundation for the KID ontology. These concepts constitute the lexicon used to describe and represent the core regulatory terms and information found in KIDs. A well-defined semantic foundation mitigates ambiguity and facilitates communication and comprehension. Moreover, the process of identification of the relevant terms empowers the ontology to faithfully represent the details of the applicable regulatory framework in order to ensure an accurate capture of the knowledge contained within the KID.

The goal of defining the classes, hierarchies, properties and instances of the KID ontology hold critical significance in establishing a comprehensive, structured representation of the domain. The classes and hierarchies highlight a structured representation, promoting clarity and organizational coherence which is critical for an effective knowledge representation of the domain.

These two articulated goals are integral facets of crafting the KID ontology. The first goal ensures that the ontology is firmly based in the regulatory context while the second goal shapes the KID ontology. Together, these goals foster the creation of a resilient and versatile ontology that captures the complexities of the regulatory framework.

The contemporary financial landscape requires methods for knowledge representation and information sharing to effectively navigate the difficulties inherent in financial regulations and investment products. To address these challenges, the use of ontologies and Semantic Web technologies has gained considerable attention, providing a promising way for knowledge sharing and exchange. By recognizing the importance of such technologies, this research aims to explore the advantages of employing OWL for the purpose of sharing and utilizing knowledge within an ontology, specifically focusing on its applicability to complex investment products.

In this context, this research explores the capabilities that the OWL offers in terms of knowledge representation, reasoning and information sharing. This study aims to explore the ways in which this language can contribute to an understanding of complex investment products, particularly the content of essential information documents.

In light of the above, the main questions that this research aims to address are the following:

i. RQ1 - What is the advantage of using OWL to share and use knowledge as an ontology?

ii. RQ2 - Can ontologies capture the structure and semantics of classes of complex investment products such as PRIIPs?

5. Document structure

The dissertation document is organized into distinct chapters, each contributing to ensure a comprehensive and coherent presentation of the research topic. As such, the dissertation is divided into five main chapters. Chapter 1, the introduction, sets the stage for the research, defining the scope, motivation, potential contributions, main goals and the research questions that this research aims to address. The intention behind this specific structure is to guide the reader across the key elements of the study, setting the main framework for the subsequent chapters.

Chapter 2, regarding the literature review, delves into the foundational knowledge of ontology, exploring its definition, main components and classification. It also briefly introduces the concept of Semantic Web and emphasizes the benefits of using ontologies. The chapter further examines the ontologies development life cycle, exploring prominent methodologies such as Enterprise methodology, TOVE and Methontology.

In addition, it explores related work within the financial market's domain, highlighting the relevance of knowledge extraction from Key Information Documents using Natural Language Processing.

Chapter 3 focusses on the KID sections and structure, proving information about the essential elements withing these documents.

Chapter 4 expands on the ontology development process for KID, discussing the stages of specification, knowledge acquisition implementation and evaluation. The implementation phase is detailed, covering the application of Web Ontology Language and the creation of the KID ontology classes, object properties, data properties and instances. Challenges encountered during the implementation process are also examined, contributing to the understanding of the development process.

Also, within this chapter, the evaluation process was conducted though reasoning and validation using SPARQL queries, demonstrating the ontology's practical application and utility. The chapter culminates with a note about the analysis of similar ontologies.

Finally, chapter 5, concerning the conclusions, summarizes the main research findings, provides answers to the research questions stated and future research.

CHAPTER 2

Literature Review

1. Ontology definitions and its main components

The present chapter is dedicated to the ontologies literature review diving into the foundational aspects of ontology development process, examining the different methodologies employed by the diverse authors and the theoretical basis that contribute to the definition of ontology.

Traditionally, the term ontology has been identified as a category within the field of philosophical studies, dedicated particularly to the study of ontological reality, mainly focused on the Being and its essential characteristics. More recently, this term has been applied by many authors to the context of information systems but with a different meaning. In this specific field of knowledge, ontologies are envisioned as an information representation tool capable of collecting, mapping, and disseminating knowledge of specific fields of study. Fundamentally, ontologies are abstract models that allow the architecture of a lexicon of technical expressions used in a particular scientific domain, enabling a language free of ambiguity which facilitates the transfer and usage of knowledge through all stakeholders.

In the literature there are several attempts to define the term "ontology", with some notable notions being that ontology consists of an explicit and formal specification of a shared conceptualization (Gruber, 1993). The specification refers to concepts, attributes, relations, and axioms that are explicitly predefined. According to this definition, formalization refers to its interpretability by the information systems, the conceptualization refers to the abstract model of a particular phenomenon in the real world that is being mapped, and the transferability refers to consensus in the community (Borst, 1997).

In Guarino and Giarreta (1995), the authors adopt a divergent approach from Gruber's. For them, ontologies are both a partial and explicit description of specific concepts. According to these authors, ontology fulfils two essential purposes: (i) conceptualization should be a true syllogism independent of the different subjects who are going to utilize it. This means that the ontology should be based on a conceptual construction independent of subjective dimensions, as it is intended to be used as a basis for sharing knowledge of a specific domain; and (ii) ontology should consist of a set of premises through which strict restrictions are developed according to inferences designed to be shared by users who agree with its conceptual construction.

In Fikes and Farquhar (1999), ontology is the study of a specific domain that defines a lexicon of entities, classes, properties, predicates, functions, and a set of relationships that necessarily exist between such concepts. Such definition is considered one of the most comprehensive in scientific literature as it clearly identifies all the attributes that are inherent to every ontology.

The research fields related to ontologies are expanding in computer science and encompass multiple areas of knowledge, many of which are related to artificial intelligence systems. In fact, there are plenty of benefits that can be obtained through the ontological formalization of a specific domain of knowledge, and ontologies are being used in many and diverse scientific domains, including natural language processing, knowledge representation, knowledge management, among others.

Despite the wide range of ontological notions that appear in academic literature, the formalization of a knowledge domain through ontology will always result in a language that represents the existing knowledge about one specific domain.

The scientific literature identified a set of essential elements in the development process of an ontology, namely (Gruber, 1993):

i. Classes: organize concepts associated with a particular domain, constructed based on a taxonomy;

ii. Properties: represents the type of interaction established between classes in a particular domain;

iii. Instances: examples or use cases of classes used to represent specific objects;

iv. Competency questions: questions designed to be answered by the ontology. They help define the scope and characteristics of the ontology, specify the tasks and problems to be addressed.

In addition to the main elements above, it is also important to address the concept of axiom. An axiom, in the context of an ontology, refers to a statement that represents a logical assertion concerning a certain link or properties stated within the conceptual framework of a domain. These axioms fulfill a relevant role defining the structural and semantic aspects of the ontology. In practical terms, an axiom can be a class statement, a property assertion and/or a data property specification of the ontology (W3C Recommendation, 2012).

Ontologies have thrived in various areas of expertise in scientific literature, namely within the legal domain. However, legal ontologies have certain features that differ from the ontologies in other areas (Uschold and Gruninger, 1996). As legal rulings must be justified by reason and supported by solid evidence, legal ontologies are more inclined to cover epistemological concepts, such as norms, legal and/or natural person, duties, rights, legal documents, among others technical terms (Corcho et.al., 2005).

In this phase, it is relevant to emphasize some characteristics that can be observed in legal ontologies. The Law, which intends to regulate human actions, relies on documents to support the reasoning behind any legally binding decisions. That's why documents are the main infrastructure behind all legislative processes. Documents have three main dimensions: (i) the physical dimension (which is the document); (ii) the representational dimension (the form in which the language is represented); and (iii) the cognitive dimension (which is the intended content by its author). Also, the concepts used in legal science, similarly to other sciences, carry a specific meaning that cannot be confused with other meanings that might be associated with that concept. As such, depending on the specific legal domain of analysis, concepts such as contracts, liability, property, markets, financial products or documents must be understood from the point of view of the legal domain where they are inscribed and with the definition provided by the relevant legislation.

Ontologies have frequently been employed within the context of information systems research as a supplementary tool alongside other framework This is due to ontologies possessing an enhanced capacity for articulating formal knowledge, as they allow users the opportunity to employ logical formalisms for representing knowledge within a designated scientific domain.

Among the potential practical applications of ontologies in the realm of information systems, the following can be listed:

i. Facilitation of information extraction from libraries or scientific literature sourced from several web-based outlets;

ii. Automatic integration of a set of standardized vocabularies or data dictionaries pertaining to a specific domain, thereby contributing to the establishment of a unified standard vocabulary;

iii. Assistance in natural language translation, thereby aiding in the resolution of language ambiguity related problems;

iv. Integration of databases, software or business models;

v. Support in the development of systems capable of handling complex and heterogeneous information, suitable for human semantic distinctions. Such systems are typically tailored for specialized applications.

2. Semantic Web

The connection between the semantic web and ontologies is broadly cited in scientific literature. The semantic web aims to establish infrastructure for enhancing the efficiency of information systems (Berners-Lee, Hendler and Lassila, 2001). It aims to ensure that certain tasks, namely the search for information, attain the utmost precision and are devoid of ambiguities.

Ontologies constitute conceptual models designed to consolidate and formalize the technical lexicon utilized within semantic applications. They form the foundation upon which unambiguous communication is built. From this perspective, ontologies can be considered as one of the pillars of the semantic web. They provide a framework of concepts and insights into the intricate relationships that underpin the semantic web's structure and functionality, offering a set of notions and a hint concerning the close connections that are established between them. Ontologies play a fundamental role in enhancing interoperability, data integration and knowledge sharing in multiple scientific domains. They allow the machines and systems to better comprehend and reason about data, thereby assisting more meaningful interactions within the context of the semantic web.

At the heart of ontologies rests the characteristic of interoperability. This characteristic entails the ability of multiple diverse information systems to collaboratively operate, facilitating the efficient and effective exchange of information among individuals, organizations and systems. Consequently, ontologies fulfill the important function of increasing the potential for information exchange between information systems that use them. Interoperability between systems emerges from the need for data to adopt a uniform and integrated structure, allowing its widespread sharing (W3C, 2022). To ensure that information is represented formally and explicitly there are specific languages for its representation. Within scientific literature, two widely recognized languages for the formal representation of ontologies are the languages Resource Description Framework (RDF) and Ontology Web Language (OWL), both created and standardized by W3C. These languages enable the sharing and integration of ontological knowledge across different information systems, promoting greater interoperability and semantic consistency in multiple domains.

Specifically, the RDF model provides its users with a simplified vocabulary that is typically valuable in handling metadata but presents some gaps in its characteristics that hinder its ability to meet de comprehensive representation demands posed by ontologies (Mika, 2007).

Ontologies normally require more expressive languages, such as OWL, to capture complex connections, define classes and establish hierarchies and specify formal semantics.

On the other hand, the OWL language allows the description of the concepts and their respective connections in a more comprehensive way than the RDF model, allowing the construction of interoperable ontologies. OWL extends RDF and provides a more comprehensive framework for ontology modelling, while RDF serves as a valuable foundation for representing data, ontologies often require additional capabilities that OWL offer in order to achieve a robust and formalized knowledge representation.

Despite the differences between the two models, both RDF and OWL allow the formalization and representation of the knowledge that is being studied and can ensure an unambiguous and polysemy-free interpretation of concepts.

By using ontologies in web applications and enabling information systems to process them, an assurance is established that future information systems will be more efficient and faster.

In the context of developing the sematic web, it becomes imperative that all input information supplied to information systems possesses a certain degree of intelligibility. This level of intelligibility is achieved by defining clear rules applied to metadata and the creation of rules governing the transformation of said metadata into other forms. These steps are required in order to ensure the semantic consistency, interoperability and the ability of systems do derive meaningful insights from data, thereby enhancing the overall efficiency and effectiveness of information systems in the ever-evolving digital landscape.

3. Ontology classification

According to the scientific literature, ontologies exhibit multiple potential classifications, which the main are the following:

i. High-level ontologies: these ontologies describe general concepts that are independent of specific problems or domains;

ii. Domain ontologies: they describe the vocabulary of a specific domain by specifying concepts introduced in high-level ontologies;

iii. Activity ontologies: these ontologies describe a vocabularies related to a certain generic activity by specializing existing from high-level ontologies. Examples include the analysis of medical software or diagnostics; and

iv. Application ontologies: these ontologies describe concepts present in both in domain and activity ontologies, serving as specifications of both ontologies.

The current typological categorization of ontologies relies on the distinctive property of concepts as the main classification criterion. It’s important to note that alternative classification methods can be found in the literature, some of which relate ontologies to their function or base the classification on the degree of vocabulary formalism.

Notwithstanding the numerous classification criteria proposed in the literature, this investigation adopts the classification proposed by Guarino (1998). The author introduces an additional distinction between “unrefined” and “refined” ontologies. An ontology is considered “unrefined” when it contains a minimal number of axioms and aims to be shared by users who adhere to a specific worldview. Conversely, a “refined” ontology includes a certain number of axioms written in a highly expressive language and is intended for sharing among users who have already reached a generalized consensus regarding the underlying conceptualization.

Guarino’s classification criteria provides a valuable framework for understanding ontologies varying levels of sophistication and intended usage, offering insight into the diverse roles that ontologies play in knowledge representation and semantic applications.

4. The benefits of using ontologies

Using ontologies offers many advantages, namely regarding knowledge representation, interoperability and machine learning or artificial intelligence applications. By providing a structured and formal way to represent knowledge, through the definition of concepts, their properties and respective connections, makes it easier for machines and humans to understand and reason about complex domains. Using a common vocabulary and shared semantics, ontologies bridge the gap between different data formats and standards allowing the interoperability between systems and the respective exchange information.

Ontologies also enable machines to reason, infer and make decisions based on formalized knowledge, which constitutes valuable tools for machine learning and artificial intelligence applications. For instance, in natural language processing, ontologies can provide unambiguous language, which can be particularly useful in the analysis performed in the financial domain.

The ontological representation of a certain domain can bring several benefits, among which the following stand out:

i. The possibility for any stakeholders to reuse ontologies and databases knowledge, even with adaptations and extensions. The impact attributed to the development of information systems based on formal knowledge was substantial, as the construction of knowledge bases requires expensive and slow tasks of a given information systems project;

ii. The availability of a wide range of ontologies, ready to be reused and shared among various stakeholders. Currently, the most extensive ontologies, some of them with more than 2,000 axioms, include satellite image metadata, database integration genome, product catalogs, robotics, semiconductors, terminology medical, the Institute of Electrical and Electronics Engineers standard for interconnections between tools, and among others;

iii. Allows the representation of knowledge that facilitates its reuse and can enable more efficient communication between agents;

iv. Online access to ontology servers which would serve various communities and that can function as tools to maintain the integrity of the knowledge shared between them, ensuring a uniform vocabulary.

5. Ontology development life cycle

The scientific literature provides multiple references to methodological processes aimed at the creation of ontologies, some with convergent development stages that are universally accepted by the scientific community as essential to the development of any ontology, namely the specification phase, the knowledge acquisition phase, the conceptualization phase and the implementation (Pinto and Martins, 2004).

Concomitantly with these main phases of the development process of ontologies, there are parallel activities that are performed, and which are known as support activities, namely the documentation and integration phases with existing ontologies.

Resuming to the main ontology development phases, the specification phase is intended to identify and define the purpose and scope of the ontology. This phase should include a prior analysis with a view to deciding whether it is possible, necessary, or appropriate to resort to reuse of pre-existing ontologies. It will be at this stage that questions such as “why purpose will the ontology be constructed?” and/or “what will be your purpose and that of your users?”.

The implementation phase is very important for the development of the ontology, being the moment where the authors define the classes, properties and the instances. The implementation phase will transform the ontology into something computable.

In this context, it is important to heed the tips of the Protégé system that contains a list of stages logically designed to clarify this phase, namely:

i. As a first point, it is important to list the main domain concepts. At this stage, aspects such as similarity, reiteration and establishment of relationships between the concepts do not assume particular relevance;

ii. Next comes the class definition task. It is important to determine the concepts that contain within themselves a set or universe of objects that present similar characteristics to each other;

iii. Next, it is important to define a hierarchy for the classes, if applicable. It's a process that occurs concomitantly with the previous one by creating subclasses and providing some clarity and consistency to it. Clarity relates directly to the number of subclasses defined for a certain class - the greater the number of subclasses, the less clarity of the hierarchy. When clarity decreases, it is important to consider the need to use intermediate classes. To this end, an assessment must be conducted whether the class and its subclass have intermediate classes. The scientific literature identifies three distinct approaches to building hierarchies (Uschold and Gruninger, 1996), namely: (i) the top-down approach, defining the most general classes and then the most specific; (ii) the bottom-up approach, which starts by defining the most specific classes and then the more generic ones; and (iii) the middle-out approach, which begins by defining the central concepts and subsequently generalized or appropriately specialized;

iv. Subsequently, it is important to define the attributes of each identified class. This phase interacts with the two previous ones, as they are the new attributes that will contribute to the definition of the class;

v. Regarding the creation phase, it will correspond to the more specific concepts of the ontology;

vi. The naming of all elements that make up the ontology must be as easily understandable as possible, being widely considered as good practice to use different names for classes, properties and instances. Also considered good practice is the least possible use of word contractions, so that the ontology can be readable for all users who consult them.

Finally, the evaluation phase is where tests will be carried out to verify whether the ontology meets the requirements specified in the initial phases, in particular in the specification.

Over the years, the scientific literature proposed several methodological development processes, which include the Enterprise Ontology processes (Uschold and King, 1995), the Toronto Virtual Enterprise (TOVE) process (Gruninger and Fox, 1995) and the METHONTOLOGY process (Fernández, 1997).

Although there are several attempts to create a shared methodology for the development of ontologies, nobody manages to obtain total consensus in academia and practice has shown that many research groups choose to create their own development method depending on the type of ontology to be created or in an adaptation of the mentioned methodological processes.

5.1.Enterprise Methodology or Uschold and King method

The Enterprise methodology, also known as the Uschold and King method in tribute to its creators, provides a set of techniques, methods and guidelines for each stage. According to the present method, the creation of an ontology of a given domain requires the following stages (Pinto and Martins, 2004):

- i. Identification of the scope. This first parameter requires the identification of the reason that justifies the construction of the ontology, the intention of its use and the universe of potential users;
- ii. Building the ontology requires:
    - a. Knowledge acquisition phase:
        - a) Identification of the main concepts and relationships between them;
        - b) Producing unambiguous definitions for these concepts and relationships.
    - b. Knowledge formalization phase;
    - c. Reuse phase of pre-existing ontology knowledge, if applicable;
- iii. Assessment;
- iv. Documentation of the ontology in order to eliminate obstacles to sharing the knowledge.

In the Enterprise methodology, it is considered good practice to start by defining the concepts that have a greater range of connections with other concepts, as these will be those that present greater complexity in ensuring a correct and precise definition. In case of ambiguity or polysemy between concepts, an attempt should be made to identify all their possible meanings (for example, using dictionaries and other technical support documents).

Subsequently, it must be determined which concepts must be represented in the ontology and select a term to represent each concept, always avoiding ambiguous terms.

5.2. TOVE Methodology or Grüninger and Fox method

To create an ontology using the TOVE methodology, the following steps needs to be adopted (Pinto and Martins, 2004):

i. Creation of scenarios that describe the motivation underlying the ontology proposal to identify its possible applications;
ii. Formulation of questions for which the ontology should be able to provide an appropriate response taking into account the scenarios previously developed;
iii. Formalization and representation of stated terminology;
iv. Specification of axioms and definitions for the terms previously formalized;
v. Assessment of adequacy and completeness. In this phase, an assessment of the adequacy of the ontology should be prepared considering the set of questions initially formulated.

In this methodology, secondary activities like maintenance and documentation of the ontology, are not expressly defined as being part of the ontology development process. However, the adoption of these activities is considered a good practice in the development of ontology.

5.3. Methontology

Developed by the Artificial Intelligence Laboratory of the Polytechnic University of Madrid, Spain, in 1997, its largely influenced by software development methodologies. This methodological process, characterized by the structured form of its processes, outlines eight key activities for ontology development, namely:

i. Requirements specification: The result of this stage is the preparation of a document, written in natural language, containing specific information regarding the main objective for the ontology construction;
ii. Knowledge acquisition: This stage translates into the gathering of knowledge from various relevant sources, namely interviews with experts, consultation of technical books or consultation of pre-existing ontologies on the same or similar topic;
iii. Conceptualization of the knowledge domain: This stage requires the organization of the chosen knowledge domain into a conceptual model, based on the vocabulary obtained through the previous phases;
iv. Formalization of a conceptual model: This phase consists of formalizing the previously mentioned model through a formal language;
v. Formal implementation of the model in order to become computable;

vi. Maintenance: Consists of an auxiliary activity that seeks to make changes or corrections when necessary;

vii. Documentation: This phase, also ancillary, is relevant for purposes of sharing and respective reuse of the ontology by different stakeholders. It is characterized by the writing of documents associated with the ontology and its entire process of elaboration;

viii. Assessment: In this stage the ontology is assessed and validated.

The literature that addresses this methodological process also includes activities of project management, particularly the activities of planning and control, prior to the main activities of the present ontology development methodology (Fernández, 1997).

6. Related work on the financial markets

6.1. Financial Industry Business Ontology

In the field of capital markets, there are not many ontologies mentioned by scientific research. However, the Financial Industry Business Ontology (or FIBO) has been extensively cited among the authors who have researched this domain. The FIBO ontology provides relationships between financial constructs, provide high-level descriptions, and help its users to describe the financial business, namely regarding legal entities, market data, contracts, and the contractual obligations the arise from them and for many different financial instruments (e.g., Contracts for Difference, Swaps, Options, Futures, Forwards, and many others) (Petrova, 2017).

One way to represent an ontology in FIBO is from a formal OWL description made with the Protégé ontology editor, which was an editor developed by a research team from Stanford University. The FIBO ontology can be used by anyone interested in working in the financial sector. As stated above, the FIBO ontology provides a large set of financial business-related notions, definitions, and relations between them with which organizations can use as a complement to their own models of the field.

FIBO can be more accurately described as an intricate web of ontologies rather than just a single master ontology. This web is divided into subcategories and some of them contain sets of shared ontologies that link to other subcategories. The ontologies that make up the FIBO "web" are based on the top-level ontology including groups called "sections". In turn, these sections contain a description of various types of fundamental constructs. Such formal models allow for separate description, application and extension of concept groups contained within separate modules.

6.2. Bank Regulation Ontology

Within the financial markets field of study, we find the Bank Regulation Ontology (BRO), the Financial Regulation Ontology (FRC) and the Legal Knowledge Interchange Format (LKIF). The BRO[1] is a structured and comprehensive knowledge representation framework designed to capture and model the intricate landscape of regulatory guidelines, regulations and standards in the banking industry sector. It serves as a valuable resource for regulatory authorities, financial institutions, researchers, and policy makers to enhance their understating and compliance with the banking regulations. This ontology is particularly important in the context of a highly regulated industry where compliance with various regulatory frameworks is essential for maintaining financial stability and safeguarding the interests of the stakeholders.

At its foundations, the BRO categorizes and defines key concepts and connections related to the banking sector regulations. It encompasses high-level categories such as capital adequacy, risk management, consumer protection and reporting requirements. Each individual category is further refined to include specific regulations and guidelines issued by regulators at national and international levels. For example, it may include Basel III standards, Dodd-Frank Act legal provisions and from the Financial Stability Board.

The ontology also captures temporal aspects by tracking the evolution of the regulations over time. In fact, it can express revisions and effective dates of regulatory documents in order to ensure the most up to date information concerning that specific topic. The BRO also incorporates semantic connections between regulations, such as dependencies, conflicts and hierarchical relations, enabling users to navigate the complex web of regulatory requirements and assess their impact on financial institutions.

Overall, the BRO serves as a valuable tool for regulatory compliance and risk management within the banking sector. The ontology aims to provide transparency, consistency and efficiency regarding the compliance of regulatory requirements.

6.3. Financial Regulation Ontology

[1] https://bankontology.com/

Regarding the FRO[2], these are specialized knowledge structures designed to systematically capture, represent and organize the area of financial regulations. These ontologies are essential in creating complex legal frameworks within the financial sector in a machine-readable and semantically rich format. The difference between the FRO and BRO is that the FRO focuses on modelling legal frameworks beyond the banking sector and encompasses the sector of the financial markets. Contrarywise, the BRO focuses on modelling and representing the legal framework applicable to the banking sector.

The FRO encompasses a wide range of regulatory domains, including banking, insurance and capital markets. Each ontology is composed to represent a specific law or regulations issued by national and/or international regulators. Similarly, to the BRO, the FRO is constantly being updated in order to assure the most accurate and recent knowledge about the regulatory frameworks mapped and constantly refined to keep pace with evolving regulatory environments and emerging compliance challenges that arise within the finance sector.

6.4. Legal Knowledge Interchange Format

Another example of relevant related work is the Legal Knowledge Interchange Format (LKIF), which is a specialized framework designed to assist the structured representation and exchange of legal knowledge and information withing the field of law and jurisprudence. LKIF is a standardized ontology that leverages semantic technologies to encode legal concepts, rules, regulations and legal documents in a machine-readable format. Its main objective is to enhance accessibility, interoperability and understanding of legal information, making it a valuable resource for legal professionals, scholars and policymakers.

The LKIF employs ontological concepts that define legal entities, namely laws, legal proceedings, judges, lawyers and other legal related concepts (Gordon, 2010). Moreover, LKIF enables the integration of legal knowledge with other domains, such as natural language processing, which helps the integration with legal applications, reasoning systems, contract analysis tools and legal information capture systems. LKIF adherence to semantic web standards and principles ensures that legal information can be seamlessly integrated with other knowledge domains, fostering interdisciplinary collaborations and enhancing the capabilities of legal technology solutions.

In summary, the LKIF is a standardized ontology for encoding and sharing legal knowledge

[2] https://finregont.com/fro/html_widoco/index-en.html

in a structured and machine-readable format, allowing the comprehension of legal concepts which makes it an important resource for legal research and legal practice.

However, these ontologies do not address the specific domain of the KID ontology, since its primary focus is to provide a structured representation and standardized format for KID documents as mandated by European regulations.

6.5. KID analysis using Artificial Intelligence and Natural Language Processing

The field of research of Artificial Intelligence (AI) is comprised of many subfields, one of which is Natural Language Processing (NLP). This subfield of AI aims to train and enable computers to comprehend, interpret, and generate human language, enabling to obtain efficiency gains in communication between humans and the machines by providing the means to the latter to read human language.

By using NLP models, it is possible to analyze text based on a predefined set of rules and techniques (Liddy, 2001). This capability is being leveraged by the financial sector in order to acquire efficiency improvements in the services provided to the clients, such as chatbots for improving the customer experience, but also by helping the supervisory activity of the financial markets regulators by enabling the analysis of large volumes of documentation (Maple et.al., 2023).

The regulatory activity of the financial market supervisors requires the analysis of several legal documentation provided by the financial market operators to the investors, such as KIDs, prospectuses, financial statements, policies and procedures, among other information. In order to address this issue, “the European Securities and Markets Authority (ESMA) began exploring the use of NLP to analyze the information of more than 20,000 KIDs, from more than 500 issuers, in 21 European languages” (FSB, 2020).

More recently, ESMA published a comprehensive report detailing the results of its endeavor to apply NLP methods to a dataset of 3,220 documents with more than 593,000 pages of text. The overall results were positive and ESMA concluded that the algorithms behind NLP solutions opens new possibilities for helping the analysis of large volumes of information and lengthy documents (ESMA, 2022).

The scientific research shed light on the application of ontology-based NLP in compliance management of software engineering processes to analyze standardized documents (Javed, 2022).

In this sense, ontologies play a determining role in the formalization of knowledge as pillar above which is going to be built the NLP systems that optimize compliance management software tools. By formalizing knowledge and providing a structured framework for understanding the intricate domain-specific details within the KID, ontologies can contribute significantly to the effectiveness of NLP-based compliance management solutions.

CHAPTER 3

Key Information Document Analysis

1. Document structure

This chapter dives into the examination of key information documents analysis, exploring their structure, content and significance within the European capital markets. By analysing the elements of these documents, such as the required template and the information they convey, this chapter aims to contribute to the understanding of these document. Through this path, this chapter sets the stage to comprehend the object of this ontology.

The repercussions of the 2008 financial crisis paved the way to concerted efforts to develop KIDs (Key Information Documents) for PRIIPs, with the overarching objective of improving consumer literacy regarding financial instruments, specifically their risk, reward profiles and costs and charges associated with those financial instruments (Kling, 2020).

Insufficient understanding was identified as a significant factor contributing to the unanticipated losses experienced by certain investors during the financial crisis. Against this backdrop, the primary impetus behind PRIIP regulation has been to foster transparency and facilitate the comparison of diverse products for retail investors through the provision of a comprehensible pre-contractual document.

The pursuit of consistent transparency rules at the European Union level aimed to mitigate discrepancies and bolster investor protection, considering the variations that had previously existed in disclosure requirements across sectors and Member-States. The lack of harmonization between the applicable legal framework of the countries impeded or even precluded the comparison of distinct financial instruments, resulting in an inequitable playing field among different investment products and distribution channels. The PRIIP regulation sough to resolve this issue by mandating improved disclosures in a uniform manner for all products, irrespective of their form or structure, with the ultimate objective of rebuilding retail investors’ confidence in financial markets.

The Regulation no. 1286/2014 of the European Parliament and of the Council of 26 November 2014, set the framework applicable to PRIIPs. This legal document underscored the indispensability of disclosure requirements for investment products, enabling retail investors to comprehend the risks entailed when making investment decisions. Traditionally, both manufacturers and distributors drafted a prospectus for each product marketed to investors.

Because these documents were too complex to be understood by the average investor, they heavily relied on the advice and explanations provided by the marketing entities.

In order to mitigate this issue, the European Commission introduced a preliminary draft of regulatory technical standards for the KID. The KID is a pre-contractual information document provided to retail investors prior to making any investment decision. To ensure that the retail investors make an informed decision regarding a specific investment product, the KID must be drafted with accuracy, correctness, and clarity.

The information conveyed through the KID must be consistent with the information contained in other documents, namely prospectus, particularly regarding the terms and conditions of the PRIIP.

The legal framework that regulates the drafting of the KID requires this document to be concise and succinct, consisting of maximum of three A4-sized pages, and should ensure the comparability between different PRIIPs. The entire document should be drafted with characters that are easily legible to any readers.

According to the PRIIP regulation, the KID aims to:

i. Provide general information regarding the investment product;
ii. Identify the degree of the risk for each PRIIP in the form of a risk class by using a synthetic risk indicator (SRI);
iii. Identify the performance scenarios; and
iv. Identify all the costs and charges related to the PRIIP.

In terms of its structure, the PRIIPs Regulation stipulates that the KID should adhere to a standardized format, emphasizing brevity. This is crucial to avoid overwhelming the reader with excessive information overload. In this context, the KID should be drafted by the manufacturer in a concise way and without any unnecessary details that do not contribute to making well-informed investment decisions.

To fulfil the information needs of retail investors and enable easy comparison of various PRIIPs, it is crucial that the document is prepared in a standardized format. This ensures consistent ordering of items and adherence to uniform headings across all KIDs.

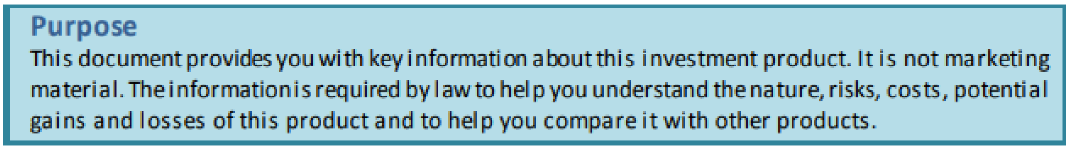
In terms of quality of information requirements, the KID is expected to present precise, equitable, lucid, and non-misleading information that aligns with other binding documents and terms and conditions applicable to the specific PRIIP. Furthermore, the PRIIPs Regulation emphasizes that the KID should be an independent and unbiased document, clearly separated from any marketing materials. It should also refrain from including cross-references to promotional content.

2. Document sections

The KID is a crucial document required under the PRIIPs regulation in the European Union market. It is designed to provide retail investors with clear, concise and standardized information about a specific complex financial product. The KID is divided into several sections, each with a distinct purpose and displaying specific information content.

The KID can be divided between 10 sections or areas, each one designed to include specific information:

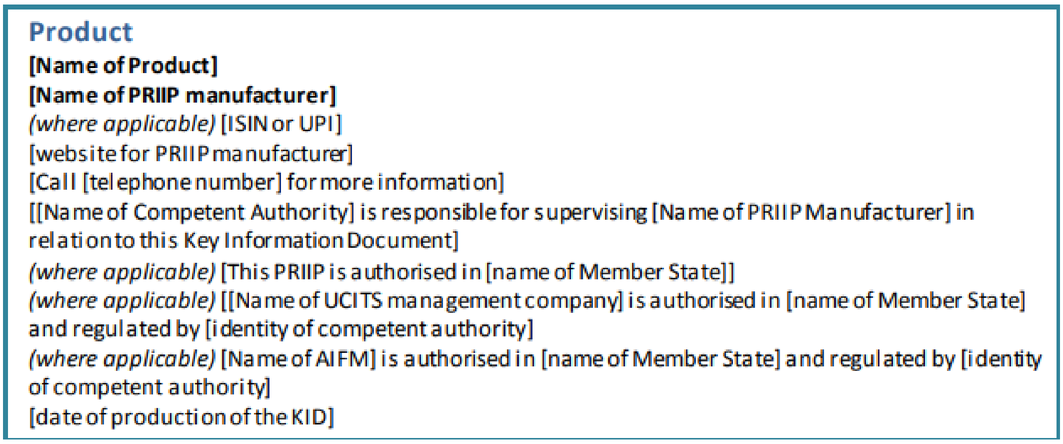
i. Section 1 contains an information alert to the investor (Figure 1):



Purpose
This document provides you with key information about this investment product. It is not marketing material. The information is required by law to help you understand the nature, risks, costs, potential gains and losses of this product and to help you compare it with other products.

Figure 1 - Section 1 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

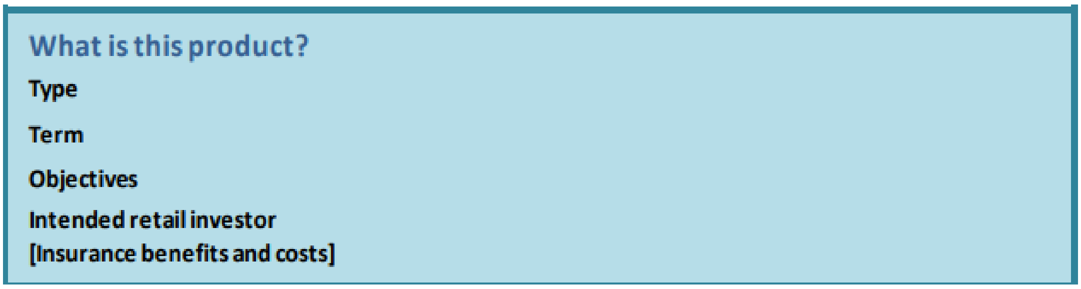
ii. Section 2 offers general information about the investment product, namely the contact details of the manufacturer or the national competent supervisory authority. This section includes the name of the product and where can be found it's legal authorizations and by which regulatory body (Figure 2):



Product
[Name of Product]
[Name of PRIIP manufacturer]
(where applicable) [ISIN or UPI]
[website for PRIIP manufacturer]
[Call [telephone number] for more information]
[[Name of Competent Authority] is responsible for supervising [Name of PRIIP Manufacturer] in relation to this Key Information Document]
(where applicable) [This PRIIP is authorised in [name of Member State]]
(where applicable) [[Name of UCITS management company] is authorised in [name of Member State] and regulated by [identity of competent authority]
(where applicable) [Name of AIFM] is authorised in [name of Member State] and regulated by [identity of competent authority]
[date of production of the KID]

Figure 2 - Section 2 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

iii. Section 3 contains information about the PRIIP, namely the type of product, supervisory authority, production date, among other details. In here the investor may also find information about other financial entities that are involved in the marketing cycle of the product, namely the depositary (if applicable) and where the specific documentation can be obtained (Figure 3):



What is this product?

Type

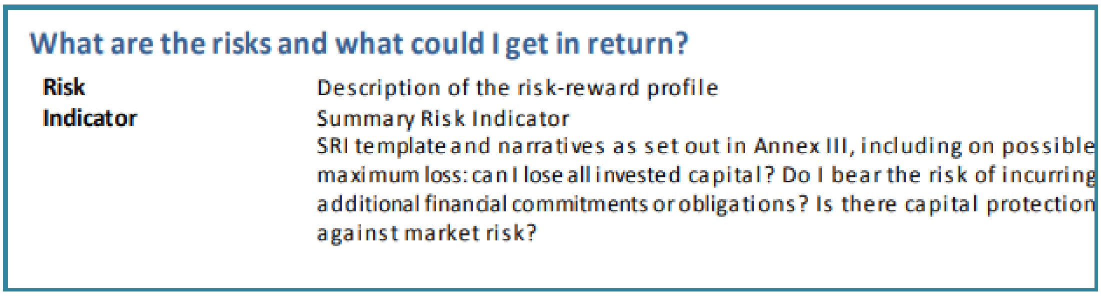
Term

Objectives

Intended retail investor
[Insurance benefits and costs]

Figure 3 - Section 3 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

iv. Section 4 is crucial section that focuses on the risk-reward profile of the product, highlighting potential risks investors may face and the expected returns. Contains information concerning the risks involving these products. This section details the degree of risk associated with product, which is represented using the summary risk indicator that ranges from 1 to 7, with 1 being the lowest risk level and 7 being the highest risk level. The SRI takes into account the volatility of the financial instrument (market risk) as well as the credit rating of the issuer (credit risk) (Figure 4):

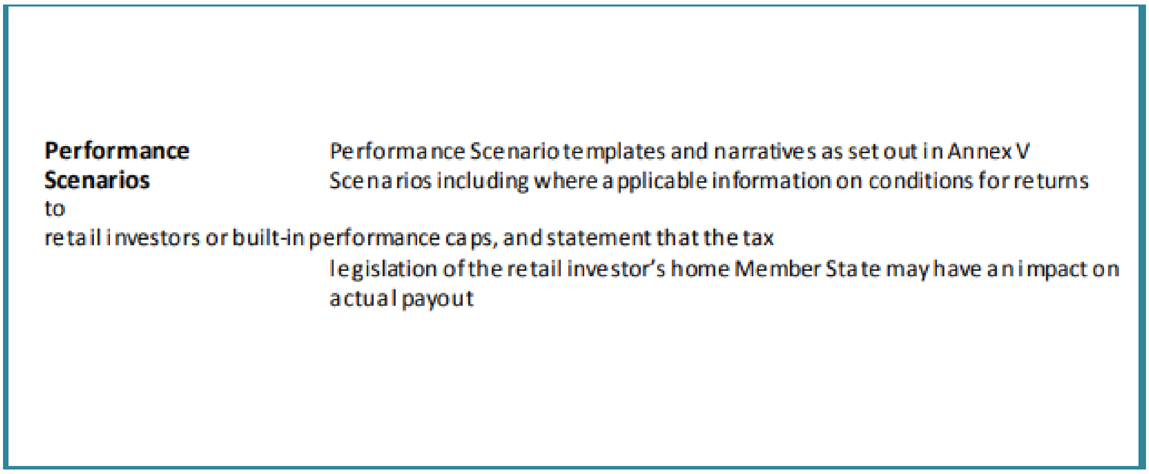


What are the risks and what could I get in return?

Risk Indicator	Description of the risk-reward profile Summary Risk Indicator SRI template and narratives as set out in Annex III, including on possible maximum loss: can I lose all invested capital? Do I bear the risk of incurring additional financial commitments or obligations? Is there capital protection against market risk?

Figure 4 - Section 4 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

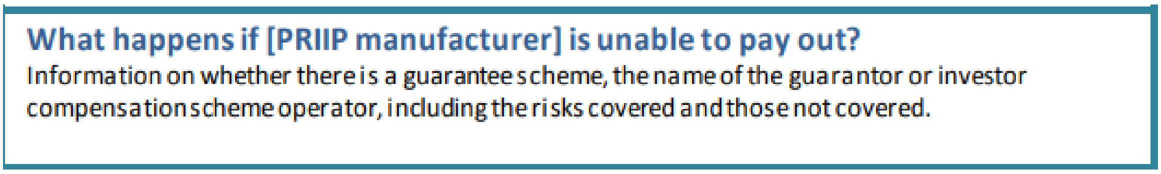
v. Section 5 contains information about the performance scenarios. These scenarios explain what the investor might obtain in return after costs across a range of performance scenarios, based on historical returns. The scenarios that are represented are "stress", which illustrates the return on the investment in extreme market conditions, the "unfavourable", the "moderate", "favourable", which represents the worst, average and best performance of the product during a certain period of time (Figure 5):



Performance Scenarios
Performance Scenario templates and narratives as set out in Annex V
Scenarios including where applicable information on conditions for returns
to
retail investors or built-in performance caps, and statement that the tax
legislation of the retail investor's home Member State may have an impact on
actual payout

Figure 5 - Section 5 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

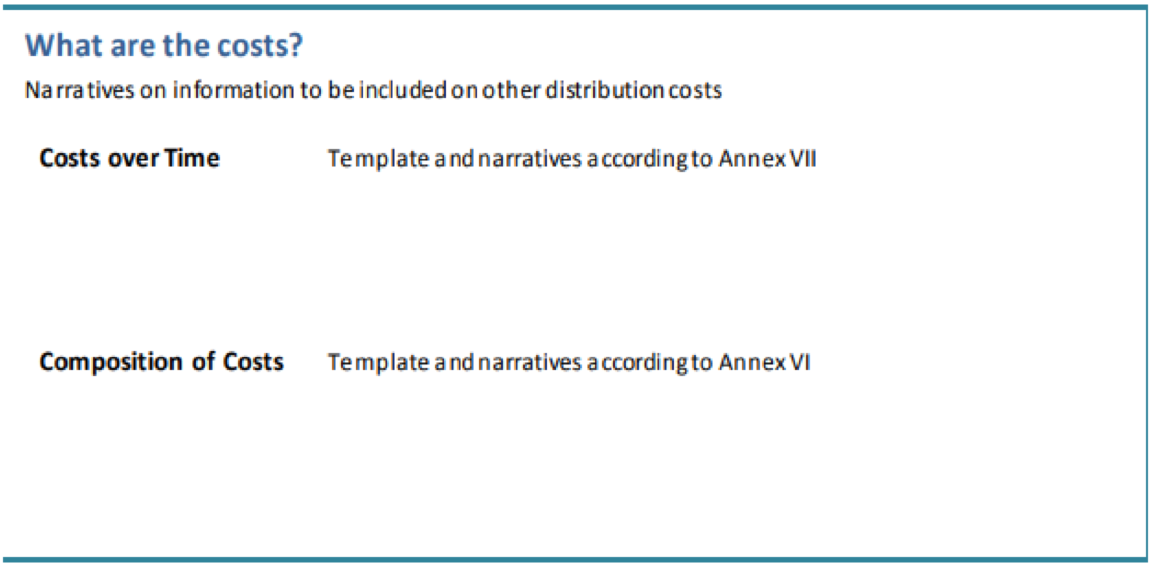
vi. Section 6 contains information regarding the investor compensation scheme accessible to the investor when the institution is deemed to become unavailable to repay the investment. It typically details the issuer's financial health, credit rating and any compensation scheme in place and outlines potential scenarios and consequences for investors if the issuer defaults (Figure 6):



What happens if [PRIIP manufacturer] is unable to pay out?
Information on whether there is a guarantee scheme, the name of the guarantor or investor compensationscheme operator, including the risks covered and those not covered.

Figure 6 - Section 6 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

vii. Section 7 contains information about the costs and charges to investors and their prospective impact on the return of the investment. The section contains information that allows the investor to compare the overall costs between different investment products. This includes upfront charges like entry fees, ongoing fees such as management fees and administration fees and any other expenses related to holding the investment (Figure 7):



What are the costs?

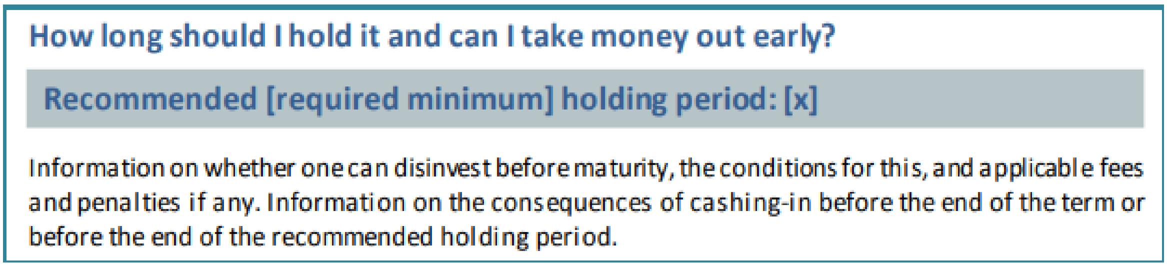
Narratives on information to be included on other distribution costs

Costs over Time Template and narratives according to Annex VII

Composition of Costs Template and narratives according to Annex VI

Figure 7 - Section 7 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

viii. Section 8 contains information about the recommended holding period and potential consequences applied if the investor chooses to withdraw before that recommended period. It also addresses liquidity concerns by explaining whether early withdrawals or redemptions are possible (Figure 8):



How long should I hold it and can I take money out early?

Recommended [required minimum] holding period: [x]

Information on whether one can disinvest before maturity, the conditions for this, and applicable fees and penalties if any. Information on the consequences of cashing-in before the end of the term or before the end of the recommended holding period.

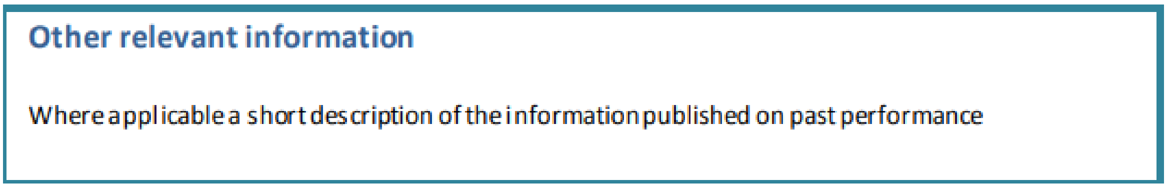
Figure 8 - Section 8 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

ix. Section 9 contains information regarding the complaint channels available to the investor, namely on how and whom to contact in order to address some issue that may arise. Investors can find contact details for relevant complaint channels, including the product manufacturer, financial ombudsman services or supervisory bodies. It explains the steps to follow when initiating a complaint (Figure 9):

How can I complain?

Figure 9 - Section 9 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

x. Section 10 contains other relevant information, namely on the methodologies used to calculate the costs and charges, the performance scenarios and the risk calculations. It may include miscellaneous details that investors should consider. Information found in this section varies but encompass tax considerations, legal provisions or any other relevant factors that may impact the investor’s decision-making process (Figure 10):



Other relevant information

Where applicable a short description of the information published on past performance

Figure 10 - Section 10 of KID Template: Commission Delegated Regulation (EU) no. 2021/2268

Together, these sections create a comprehensive, standardized document that empowers retail investors to make informed decisions about investing in PRIIPs while ensuring transparency and investor protection in the financial markets.

CHAPTER 4

Ontology development process

1. Ontology development for Key Information Documents

This chapter explores the steps involved in building the KID ontology, a semantic framework designed to capture and represent this specific domain. From the specification and knowledge acquisition phases to the implementation and evaluation stages, the chapter navigates through the landscape of the ontology development process followed to build the KID ontology. Here, we explore the tools used and unravel the formal representation of the domain classes, object properties, data properties, instances and share some of the issues encountered during the process.

The KID ontology can be categorized as a domain ontology, which describe the vocabulary of a specific domain by specifying concepts introduced in high-level ontologies. By defining the main concepts, links and properties of the KID domain, the KID ontology establishes a structured knowledge representation that enables the understanding of this specific scientific area.

The KID ontology was developed using Protégé editor that serves as an indispensable platform that accommodates a wide spectrum of formats, including RDF and OWL, enabling the robust creation and management of complex ontologies. The use of OWL, an XML-based framework, facilitated the description and formalization of the concepts inherent to the KID ontology, ensuring a standardized and systematic representation of the information associated with PRIIPs. Particularly, the Protégé editor has emerged as the industry standard for ontology development and maintenance, primarily due to its user-friendly interface, comprehensive features, and extensive support for ontology formats. Although the software is available in both web and desktop systems, the KID ontology was developed using the desktop version. This decision was based on the specific requirements of the ontology development process that considered factors such as data security, enhanced performance, offline accessibility, and documentation process. The Protégé editor's extensive functionalities and toolkit served as a cornerstone for the creation and implementation of the KID ontology, underscoring the relevance of employing this software in the field of ontology development process.

The development of this ontology was based on the Methontology methodology, an established approach that guides the systemic construction of ontologies. This development methodology is characterized by a set of steps that includes the specification, conceptualization, implementations and evaluation phases. By choosing this specific methodology, it is possible to leverage a well-structured development process that facilitates the representation of the KID domain and its validation and verification.

In light of the above, to develop the KID ontology the following main tasks were employed: (i) specification; (ii) knowledge acquisition; (iii) implementation and (v) evaluation. The Protégé editor enabled for the creation of a hierarchy of concepts (classes), which can be further categorized and enables the definition of connections between the classes.

1.1.Web Ontology Language

The formal representation and modelling of the knowledge extracted from the KID was conducted by using Web Ontology Language (OWL). OWL is a World Wide Web Consortium (W3C) standard language, which ensures interoperability and compatibility across different software and applications. This standardization characteristic is essential for the ontology development process, as it ensures that the ontology can be utilized and integrated into multiple systems in a consistent manner.

The OWL also enables the use of semantic querying using SPARQL, which can be used to retrieve information for the created ontology. The integration of SPARQL in OWL further strengthens the ontology’s capacity to serve as a robust and reliable knowledge repository.

1.2. Specification

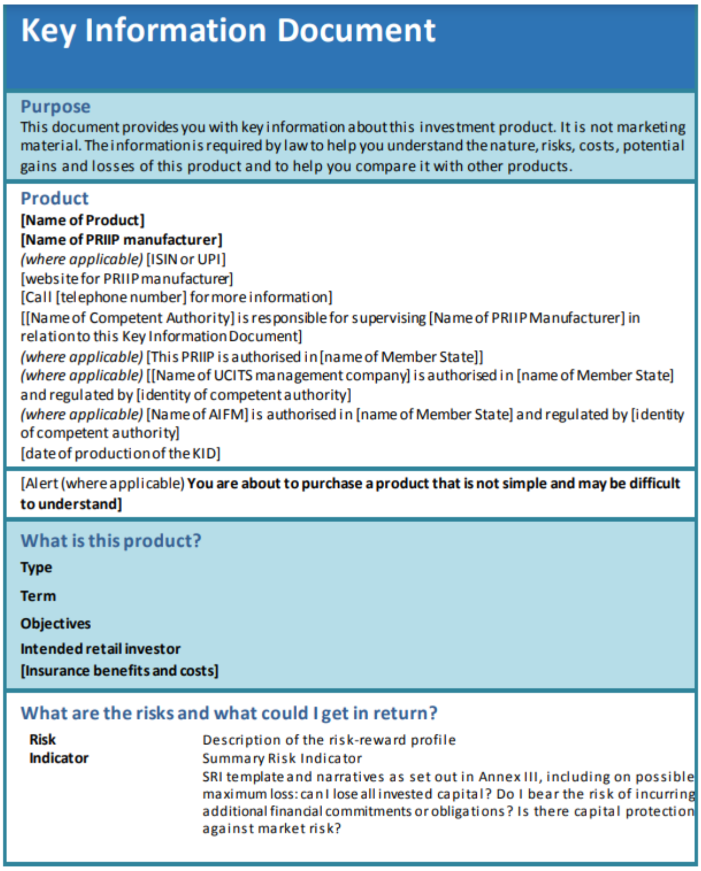
In the ontology development process, articulating the purpose serves as the foundation for delineating the key objectives and aspirations that the ontology aims to accomplish. The purpose, scope and the degree of detail are essential parameters that need to be defined for any ontology development process. The purpose of an ontology serves as the guiding principle behind its creation. It defines the specific objectives and goals that the ontology seeks to achieve. In the context of the KID ontology, the purpose is to develop a semantic ontology that supports the KID analysis within the securities market enabling automated processes for regulatory compliance purposes. It is also intended to enhance the understanding of the KID content, its structure and CFD-related information to investors by creating an instantiated knowledge database that describes the concepts and the properties of the KID developed by any manufacturer of PRIIPs using the template provided by the regulation.

In the context of the KID ontology, establishing a precise scope is imperative to ensure that the ontology's focus remains well-defined and aligned with its intended purpose. The scope of the KID ontology is limited to certain types of PRIIPs, particularly the CFDs. By confining the scope to specific PRIIPs, notably centring in CFDs, the ontology can concentrate its efforts on the nuances that are most relevant to the targeted domain. This scoping ensures that the ontology remains focused and manageable and prevents the addition of unrelated concepts, maintaining its relevance to the defined objectives. Given that the underlying financial instruments of the CFD do not alter the configuration of the KID, it is not necessary to limit the scope of the investigation to a certain type of an underlying assets.

Regarding the KID ontology degree of detail, a balanced approach is adopted, focusing on maintaining a high-level concept granularity, for instance the concepts of "Financial_Instruments, "Competent_Courts" and/or "Performance scenarios", in order to cover the wide spectrum of CFD collectively abstracting from specific considerations, since the type of CFD doesn't affect the configuration of the KID, therefore maintaining its versatility and applicability to different contexts.

1.3. Knowledge acquisition

This stage is characterized by the data extraction from various sources. The main source of knowledge used to produce the KID ontology structure originated from the KID template structure that has been published in the Annex I of the Commission Delegated Regulation (EU) no. 2021/2268 of 6 September 2021, as shown in Figure 11:



Key Information Document

Purpose

This document provides you with key information about this investment product. It is not marketing material. The information is required by law to help you understand the nature, risks, costs, potential gains and losses of this product and to help you compare it with other products.

Product

[Name of Product]
[Name of PRIIP manufacturer]
(where applicable) [ISIN or UPI]
[website for PRIIP manufacturer]
[Call [telephone number] for more information]
[[Name of Competent Authority] is responsible for supervising [Name of PRIIP Manufacturer] in relation to this Key Information Document]
(where applicable) [This PRIIP is authorised in [name of Member State]]
(where applicable) [[Name of UCITS management company] is authorised in [name of Member State] and regulated by [identity of competent authority]
(where applicable) [Name of AIFM] is authorised in [name of Member State] and regulated by [identity of competent authority]
[date of production of the KID]

[Alert (where applicable) **You are about to purchase a product that is not simple and may be difficult to understand]**

What is this product?

Type

Term

Objectives

Intended retail investor

[Insurance benefits and costs]

What are the risks and what could I get in return?

Risk Indicator

Description of the risk-reward profile
Summary Risk Indicator
SRI template and narratives as set out in Annex III, including on possible maximum loss: can I lose all invested capital? Do I bear the risk of incurring additional financial commitments or obligations? Is there capital protection against market risk?

Figure 11 - KID Template: Commission Delegated Regulation (EU) no. 2021/2268

The mentioned regulatory framework serves as the core source for the comprehensive and standardization of the essential components and data elements within the KID documents. The adherence to the template structure enables the ontology development process to incorporate key data points, conceptual hierarchies and links into the KID ontology.

In this context, the relevant information was extracted from a sample of thirteen KIDs of CFDs issued by Portuguese financial entities, actively marketed in Portugal, and regulated by the CMVM. In delineating the sample for the KID ontology, particular attributes pertaining to the financial entities, including the dimension and volume of CFDs traded, were not deemed significant factors. Instead, the emphasis was placed on establishing a coherent and representative sample exclusively aligned with CFDs. This deliberate approach ensured that the sample selection process was primarily driven by the particularities of the CFDs, thus fostering a comprehensive depiction of their elements.

The selection of the KID sample from the CMVM website facilitate a comprehensive approach to gathering relevant information for the development of the KID ontology. By sourcing the KID sample from an authority and reliable source of information, allowed to obtain accurate and up-to-date information.

The information obtained from those documents was then converted into classes, subclasses, object property and data properties (further described in the section below).

The process of ontology reuse was taken into consideration for the KID ontology, however the lack of investigation within this domain did not allow to leverage the advantages that the reuse of an existing ontology would do for this project. This limitation consequently led to the dismissal of the ontology reuse process, as the lack of an established foundation within this specific domain of the KID significantly impeded the seamless integration and adaptation of pre-existing ontology structures. Therefore, the dissertation focused on developing a custom ontology tailored to the European KID regulatory framework.

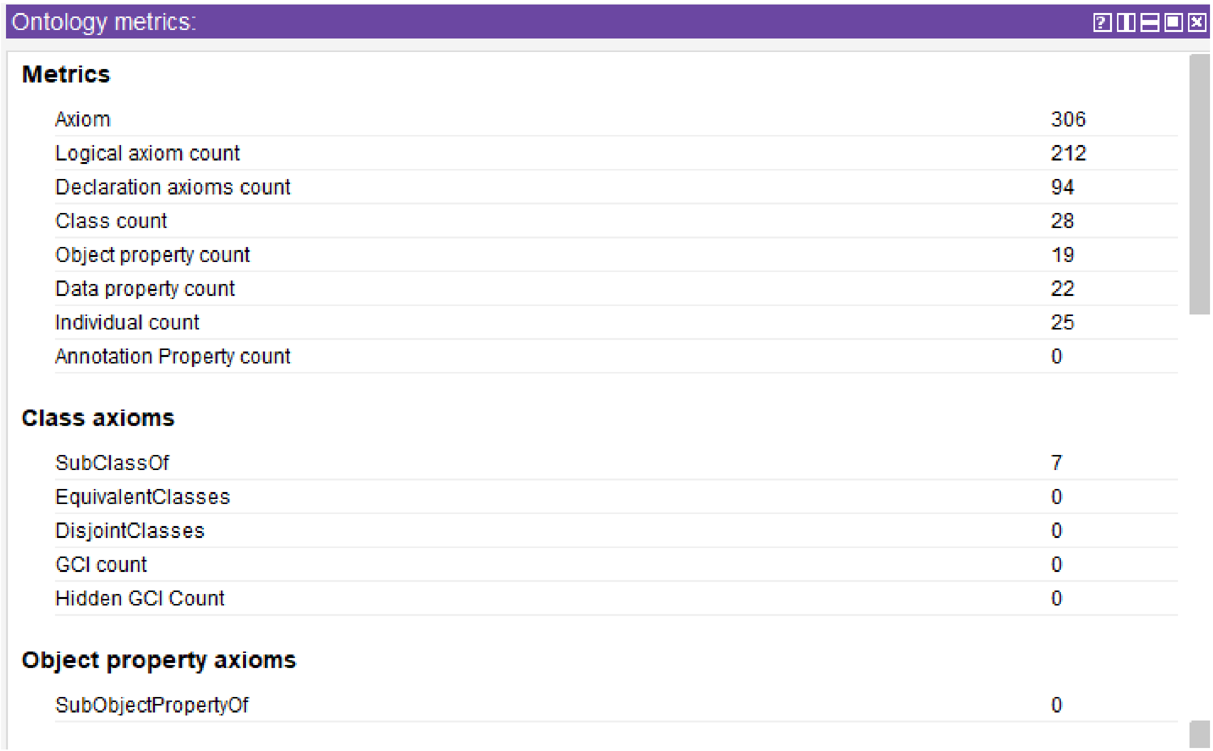
Most of the knowledge acquisition process and its analysis was focused on understanding the concepts in terms of classes and/or data properties, as well the existing connections between the terms as object properties by comparing the different KIDs used as a sample for this dissertation, the specification of those KIDs and the way from which a sentence is able to form by identifying what is a class (the subject of a sentence) and the relation between classes (the predicate of the sentence).

1.4. Implementation

By filtering the key knowledge that should be integrated in the KID ontology we have minimized the risk of redundancies and/or irrelevant information being considered. The relevant information was then compiled in a document where was represented the taxonomy, its structure and its relevant properties.

After the process of identifying and extracting the relevant concepts from template and samples, the Protégé editor was used to develop the KID ontology and represented in OWL, which provides a formal manner to describe the domain concepts.

The KID ontology counts with 28 classes, 7 subclasses, 22 data properties, 19 object properties and 25 individuals, as shown in Figure 12:



Ontology metrics:

Metrics

Axiom	306
Logical axiom count	212
Declaration axioms count	94
Class count	28
Object property count	19
Data property count	22
Individual count	25
Annotation Property count	0

Class axioms

SubClassOf	7
EquivalentClasses	0
DisjointClasses	0
GCI count	0
Hidden GCI Count	0

Object property axioms

SubObjectPropertyOf	0

Figure 12 - KID ontology metrics

1.4.1. KID Ontology classes

The implementation process of an ontology is a fundamental step in creating a structured and meaningful knowledge representation of a certain domain. This step is considered an iterative process that reflects the evolution of new requirements that may emerge during the process and consequently the need to revisit and update the class definitions and their links, beginning by identifying the key concepts that requires representation within the ontology. Each class should represent a distinct and significant entity within the domain that is being studied.

The information obtained from the KID template was converted into classes, datatype properties and object properties. By default, the KID ontology root domain is called "owl: Thing", which is the main class that represents all the subsequent classes, hence all succeeding classes are a subclass of "owl: Thing". The classes for the KID ontology are created by adding subclasses (Table 1):

Categories	Classes
Financial entities and their roles in the financial markets	Distributor_Supervisor
	Manufacturer_Supervisor

	Distributor
	Manufacturer
	Investor
	Markets
	Currency
Legal and regulatory aspects	Holding_Period
	Applicable_Law
	Competent_Courts
	Compensation_Scheme
	Jurisdiction
Financial Instruments and the risks involved	Financial_Instruments
	Risk_Factors
	Performance_Scenarios
	Product_Typology
	CFD
	Costs
Information and Communication	Information_Sites
	Complaint_channels

Table 1- Classes. Source: The author

Classes on OWL can be defined as representations of a concept about a certain domain. So, for example, a class named “Investor” encompasses everyone that invests in financial instruments. Classes can be further specified if needed. Such specifications of a given notion are referred to as subclasses and they introduce granularity to the concept that they are associated with. Building upon the previous example, the class “Investor” can be further specified to “Professional”, “Retail” and “Eligible Counterparty”, depending on their categorization, and these are represented as subclasses of the class “Investor”.

The classes that comprise the KID ontology stem from the KID documents and can be grouped into four main categories: (i) Financial entities and their roles in the financial markets; (ii) Legal and regulatory aspects; (iii) Financial Instruments and the risks involved; and (iv) Information and Communication related classes.

i. About the financial entities and their roles in the financial markets: These classes include “Product_Typology”, which categorizes financial products into different types, each with its own set of characteristics and risks. The classes

“Distributor_Supervisor” and “Manufacturer_Supervisor” represent entities that are responsible for regulating and supervising the distribution and manufacturing processes in the financial markets, specifically the CFD production and marketing processes to investors. The class “Investor” refers to individuals or organizations participating in the financial markets by investing in those products, while the “Distributor” and “Manufacturer” classes are entities involved in the process of creating and distributing financial products.

ii. The Legal and regulatory aspects: Many ontology classes of the KID ontology pertain to the legal and regulatory aspects of the financial markets. The class “Holding_Period” respects the duration for which an investment or financial instrument is recommended by the manufacturer and distributor to be held by investors in order to receive certain benefits or to comply with regulatory requirements. The classes “Applicable_Law” and “Jurisdiction” indicate the legal framework and geographical authority governing financial transactions and potential disputes. The class “Competent_Courts” refers to the authorized legal entities to address financial disputes that may arise from the negotiation of such financial instruments. The class “Compensation_Scheme” represents the legal mechanisms by which investors are compensated in case of any losses or disputes that arise from the negotiation of complex financial instruments, such as CFDs.

iii. Regarding the financial instruments and the risks involved category: The ontology includes classes related to financial instruments and risk management. The class “Financial_Instruments” encompasses a wide range of assets and contracts used for investment or hedging purposes, including products such as CFDs. The class “Risk_Factors” represents the factors that may affect the performance of the investments and the class “Performance_Scenarios” involves projections of assessments of how financial products may perform under certain circumstances. The class “Performance_Scenarios” is further specified into four subclasses that represents the four possible scenarios that the applicable legislation requires to be calculated and inserted in the KID, which are the “Favorable_Scenario”, the “Moderate_Scenario”, the “Stress_Scenario” and the “Unfavorable_Scenario”.

iv. In the information and communication category, there are classes that represent communication and information dissemination in the financial markets. The class "Information_Sites" encompasses sources of information about financial products and market conditions, where investors can obtain better understanding of the dynamics of the markets and make informed investment decisions. Finally, the class "Complaint_channels" indicates the means through which investors can raise concerns or file complaints regarding financial products or services.

The ontology classes outlined above offer a structured framework for categorizing and understanding the information contained in the KID documents of CFDs. These classes cover a certain range of elements, from financial product categorization to legal and regulatory aspects and communication channels. This structured approach not only facilitates clearer communication and information exchange but also supports risk management and compliance.

1.4.2. KID Ontology object properties

The classes represented in an ontology have properties that link them to other classes, known as "object properties", that form a relationship between them. For example, the class "Investor" has a location which is given by the class "Country".

The properties have a characteristic called inheritance, which means that if a class X have a connection with class Y by the property N, all the subclasses of X are linked to the subclasses of Y by the property N. By default, the object property root domain is called "owl:topObjectProperty" and the KID ontology is comprised of 19 object properties.

The table below provides a detailed description of the object properties defined for the KID ontology by displaying the object properties, its characteristics, domain and range (Table 2).

Object Property	Characteristic	Class Domain	Class Range
hasApplicableLaw	Functional	CFD	Legal_Framework
hasCompensationSchemes	Functional	CFD	CompensationSchemes
hasCompetentCourts	Functional	Legal_Framework	Court
hasComplaintChannels	Functional	CFD	Complaint_Channels
hasCosts	Functional	CFD	Costs
hasDistributor	Inverse functional	CFD	Distributor
hasDistributorSupervisor	Inverse functional	Distributor_Supervisor	Distributor

haFinancialInstruments	Functional	CFD	Financial_Instruments
hasHoldingPeriod	Functional	CFD	Holding_Period
hasInformationSites	Functional	CFD	Information_Sites
hasInvestor	Functional	CFD	Investor
hasJurisdiction	Functional	CFD	Jurisdiction
hasLocation	Functional	Country	Manufacturer Investor Distributor
hasManufacturer	Inverse functional	CFD	Manufacturer
hasManufacturerSupervisor	Inverse functional	Manufacturer	Manufacturer_Supervisor
hasMarkets	Transitive	CFD	Markets
hasPerformanceScenarios	Functional	CFD	Performance_Scenarios
hasProductTypology	Functional	CFD	Product_Typology
hasRiskFactors	Functional	CFD	Risk_Factors

Table 2 - Object properties assertions. Source: The author.

Although OWL provides a wide range of object properties (functional, inverse functional, symmetric and transitive), the KID ontology uses only the functional, inverse functional and transitive properties. These characteristics, which describe how the object properties behave, can be defined as follows: (i) functional, meaning that for a given individual, there can be at most one individual in the range associate with it; (ii) inverse functional, which is the reverse of the functional characteristic. This means that for a given individual in the range, there can be at most one individual associated with it; and (iii) transitive, which indicates that if an individual "A" is related to another individual "B" and the individual "B" is related to the individual "C", then there is an implied relationship between individual "A" and individual "C". The object property "hasHoldingPeriod" is an example of a functional link between the classes "CFD" and "Holding_Period", because a "CFD" should only have one "Holding_Period". Conversely, the object property "hasManufacturerSupervisor" is an example of an inverse functional link between the classes "Manufacturer" and "Manufacturer_Supervisor", given that the "Manufacturer" should only have one "Manufacturer_Supervisor". Finally, the object property "hasMarkets" is an example of a transitive connection between the classes "CFD" and

"Markets" because if a "CFD" is related to one market and that market is related to another market a link exists between the CFD and the second market.

In the class implementation process we've categorized the classes in four different main categories: (i) financial entities and their roles in the financial markets; (ii) legal and regulatory aspects; (iii) financial Instruments and the risks involved; and (iv) information and Communication related classes. As such, the object properties listed above have a crucial role in connecting the ontology classes within the same category.

The object properties "hasApplicableLaw", "hasJurisdiction" and "hasCompetentCourts" connect the CFDs to legal and regulatory aspects. For instance, "hasApplicableLaw" associates a CFD with the specific laws governing it, "hasJurisdiction" specifies the geographical jurisdiction under which the CFD falls under and "hasCompetentCourts" identifies the legal bodies responsible for handling disputes related to the CFDs. These properties enable the mapping of legal and regulatory aspects, ensuring that the CFD adhere to specific laws, operate within a specific jurisdiction and are subject to legal bodies for dispute resolution.

With regard to the object properties "hasCompensationSchemes" and "hasCosts", these link the CFD to their associated compensation schemes and costs. The property "hasCompensationSchemes" establishes the relationship between the CFD and the legal mechanisms by which investors are compensated in case of any losses. Meanwhile, the property "hasCosts" links the CFDs to its cost structure, including fees and charges, allowing investors to measure the financial impacts of their investments in an accurate manner.

Concerning the object properties "hasDistributor", "hasDistributorSupervisor", "hasManufacturer" and "hasManufacturerSupervisor", these establish a connection between the CFDs, and the entities involved in their production and marketing. The property "hasDistributor" and "hasManufacturer" link the CFD to the respective entities responsible for their creation and distribution, respectively. Conversely, the properties "hasDistributorSupervisor" and "hasManufacturerSupervisor" associate the CFD manufacturer and marketing entities with competent supervisory bodies.

The ”hasInformationSites", "hasComplaintChannels" and "hasLocation" object properties are fundamental to represent the communication and accessibility of information for investors. The "hasInformationSites" connects the CFD to sources where investors can access relevant information. The "hasComplaintChannels" establishes links to the channels through which investors can raise concerns or complaints, ensuring transparency and dispute resolution solutions between the involved parties. The "hasLocation" object property links the country to investors, manufacturer and distributor.

Finally, “hasFinancialInstruments”, “hasPerformanceScenarios”, “hasProductTypology” and “hasRiskFactors” object properties help describe the characteristics and attributes of the CFD. The “hasFinancialInstruments” object property links the CFD to an underlying specific financial instrument or contracts they represent. The “hasPerformanceScenarios” connects the CFD with projections or scenarios regarding their potential performance under certain conditions. The “hasProductTypology” categorizes the financial instruments into different types, according to their legal definition. Finally, the “hasRiskFactors” links the CFD to the factors that may affect their performance.

These object properties contribute to a more structured, organized and standardized representation of the KID documents.

1.4.3. KID Ontology data properties

The data property provides a relation to append an entity instance to a datatype value that is a measure of what that data property is about. By default, the data property root domain is called “owl:topDataProperty” and the KID ontology is comprised of 22 data properties.

The table below provides a detailed description of the data properties defined in the KID ontology by displaying the data properties, its characteristics, domain and range (Table 3).

Categories	Data Property	Characteristic	Domain	Range
Identification and description	CFD_name	Functional	CFD	xsd:string
	Cost_name	Functional	Costs	xsd:string
	Court_name	Functional	Court	xsd:string
	Distributor_name	Functional	Distributor	xsd:string
	Legal_Framework_name	Functional	Legal_Framework	xsd:string
	Manufacturer_name	Functional	Manufacturer	xsd:string
	Market_name	Functional	Markets	xsd:string
	PerformanceScenarios_name	Functional	Performance_Scenarios	xsd:string

	ProductTipology_name	Functional	Product_Typology	xsd:string
Quantitative information	HoldingPeriod_duration	Functional	Holding_Period	xsd:string
	Cost_value	Functional	Costs	xsd:decimal
	CFD_date	Functional	CFD	xsd:dataTime
Links and codes	Product_ISIN	Functional	CFD	xsd:string
	InformationSite_URL	Functional	Information_Sites	xsd:anyURL
	PRIIP_Code	Functional	CFD	xsd:integer
	Manufacturer_LEI	Functional	Manufacturer	xsd:string
	Distributor_LEI	Functional	Distributor	xsd:string
	ComplaintChannel_Contact	Functional	Complaint_Channels	xsd:string
Risk assessment	RiskFactor_detail	Functional	Risk_Factors	xsd:string
	SRI	Functional	Risk_Factors	xsd:string
	Investor_category	Functional	Investor	xsd:string
	CFD_Category	Functional	CFD	xsd:string

Table 3 - Data properties assertions. Source: The author.

The data properties listed above provides specific attributes ad values associated with the ontology classes and object properties represented in the KID ontology. These properties contribute to a more detailed and comprehensive representation of financial products, entities involved in the process and legislation.

The represented data properties can be categorized in four different categories: (i) identification and description; (ii) quantitative information; (iii) links and codes; and (iv) risk assessment:

i. The “CFD_name”, “Cost_name”, “Court_name”, “Manufacturer_name”, “Legal_Framework_name”, “Market_name” and “ProductTypology_name”, data properties provide labels associated with the classes and objects, enabling a clear *identification and description* of CFD, manufacturer, distributor, courts and costs, contributing to the enhancement of the ontology's usability.

ii. The "Costs_value" and "Holding_Period_duration" data properties provide quantitative information to the ontology. The "Cost_value" provides numerical *quantitative* values related to the costs and charges of the CFD. The "Holding_Period_duration" specifies the duration of a holding period, allowing the comprehension of the time frame recommended for a specific investment.

iii. The "InformationSite_URL" and "PRIIP_code" provide web links and unique identifying *codes* that allows a direct access to relevant information sites and specific codes aligned with the applicable legislation facilitating product tracking.

iv. The "RiskFactor_detail" and "SRI" data properties contribute to the *risk assessment*. Specifically, the "RiskFactor_detail" provides details about certain risk factors associated with the CFD and the "SRI" provides a summarized risk assessment indicator, which contributes to a simplified communication.

To put it concisely, the ontology data properties enrich its content by providing specific details, attributes and values associated with the CFD, involved entities and regulatory aspects. These properties enhance the ontology's utility be introducing precise identification and quantitative information for a certain investment assessment.

1.4.4. KID Ontology instances

The table below provides a detailed description of the data properties defined in the KID ontology by displaying the instances, the classes, object property and data property assertions (Table 4).

Individual	**Class**	**Object property**	**Data property assertions**
Banco_de_Investimento_Global	Distributor	hasDistributor	Distributor_LEI Distributor_name
Banco_de_Portugal	Distributor_Supervisor	hasDistributorSupervisor	InformationSite_URL
Bonds	Financial_Instruments	hasFinancialInstruments	ProductTipology

CFD_cryptocurrency	CFD	hasPerformanceScenarios	PerformanceScenarios_name
CFD_stocks	CFD	hasHoldingPeriod	Holding_Period_duration
CMVM	Distributor_Supervisor	hasDistributorSupervisor	InformationSite_URL
Commodities	Financial_Instruments	hasFinancialInstruments	ProductTipology_name
Complaints	Complaint_Channels	hasComplaintChannels	ComplaintChannel_contact
Credit_risks	Risk_Factors	hasRiskFactors	RiskFactor_detail
Danis_financial_supervisory_authority	Manufacturer_Supervisor	hasManufacturer Supervisor	InformationSite_URL
Deposit_Guarantee_Fund	Compensation_Schemes	hasCompensation Schemes	Legal_Framework_name
Entry_costs	Costs	hasCosts	Cost_name Cost_value
Exit_costs	Costs	hasCosts	Cost_name Cost_value
Investor_Compensation_Schemes	Compensation_Schemes	hasCompensation Schemes	Legal_Framework_name
Liquidity_Risks	Risk_Factors	hasRiskFactors	Risk_Factors_detail
Market_Risks	Risk_Factors	hasRiskFactors	Risk_Factors_detail
OTC	Markets	hasMarkets	Market_name
Performance_Costs	Costs	hasCosts	Cost_name Cost_value
Portugal	Country	hasLocation	N/A.
PRIIP_Regulation	Legal_Framework	hasApplicableLaw	Legal_Framework_name
Renato_Franco	Non_Professional	hasInvestor	Investor_category
Saxo_Bank	Manufacturer	hasManufacturer	Manufacturer_LEI Manufacturer_name
Transaction_Costs	Costs	hasCosts	Cost_name

			Cost_value
Tribunais_de_Comarca	Competent_Courts	hasCompetentCourts	Legal_Framework_name
Tribunal_de_Relação_de_Lisboa	Competent_Courts	hasCompetentCourts	Legal_Framework_name

Table 4 – Individuals. Source: The author.

The listed instances offer a representation of multiple entities, connections and attributes encountered within the KID domain. Below we explore in detail these instances.

The “Banco_de_Investimento_Global” is an instance of the “Distributor” class that represents an investment bank registered with the CMVM and authorized to provide investment services related to CFDs. It is linked to its Legal Identity Identifier (LEI code) through the “hasDistributor” object property. The LEI code is a unique identifier code that is attributed to financial institutions that helps to ensure transparency in financial transactions, such as CFD contracts negotiations.

The “Banco_de_Portugal” is an instance of “Distributor_Supervisor” linked to its information site URL through “hasDistributorSupervisor” object property.

The “Bonds” and “Commodities” are instances of “Financial_Instruments” representing the different underlying asset types that can be negotiated with CFD contracts. They are linked to their respective product typologies, in particular the “ProductTipology_name”, through the “hasFinancialInstruments” object property.

The “CFD_cryptocurrency” and “CFD_stocks” are instances of “CFD”, representing the complex financial products of CFD. The “CFD_cryptocurrency” is linked to the performance scenarios named “PerformanceScenarios_name” through the “hasPerformanceScenarios” object property. Regarding the “CFD_stocks” is linked to a holding period duration through the “hasHoldingPeriod” object property.

The "CMVM" is an instance of "Distributor_Supervisor" linked to its information site URL, because the legal framework requires that all distributors of CFD be registered with the CMVM and that information be accessible to the public domain.

The “Credit_risks”, “Liquidity_risks” and “Market_risks” are instances representing the multiple risk factors that any investor of CFD is subject to. They are linked to their respective risk factors through the “hasRiskFactors” object property.

The “Deposit_Guarantee_Fund” and “Investor_Compensation_Scheme” are instances of “Compensation_Schemes” representing different schemes that protect the investors in case of losses with the investment. Both are connected to their respective legal framework through the "hasCompensationScheme" object property.

The “Entry_costs” and “Exist_costs” represent different cost components associated with a certain financial transaction. They are connected to cost names “Cost_name” and cost values “Cost_value” through the “hasCosts” object property.

The “Portugal” is an instance that represents the country of Portugal within the ontology and it’s linked to its geographical location using “hasLocation” object property.

The “PRIIP_Regulation” is an instance of “Legal_Framework” representing a regulation related to PRIIPs and it’s linked to its legal framework name through the “hasApplicableLaw” property.

The instance “Renato_Franco” represents an example of investor of CFD associated with an “Investor_name”.

The “Saxo_Bank” is an instance of a financial institution that is linked to “Manufacturer” class and includes a LEI code given by “Manufacturer_LEI” and a name “Manufacturer_name”.

The “Transaction_costs” is an instance that represents costs associated with a certain financial transaction. It’s linked to cost names “Cost_name” and cost values “Cost_value” through the “hasCosts” object property.

Finally, the “Tribunais_de_Comarca” and “Tribunal_de_Relação_de_Lisboa” represent instances of competent courts to resolve any potential conflicts that may arise from the CFD negotiation. They are both linked to a respective legal framework “Legal_Framework_name” through the “hasCompetentCourts” object property.

1.4.5. Challenges encountered in the KID Ontology implementation process

A challenge that we have experienced during this phase was deciding whether to create a specific class for a certain concept or to set it as a property. For example, the Synthetic Risk Indicator for a certain PRIIP has two different dimensions given by the credit risk and the market risk. Initially, we have created a class named “SRI” in order to represent this concept. However, we had also created a class named “Risk factors” that had 17 subclasses of risks associated to the PRIIPs negotiation cycle. The credit risk and market risk were both represented as a subclass of the class “Risk factors”. Because of this redundancy and considering the scope of KID ontology, we made the decision to remove the class “SRI” and represent it as a data property related to the class “Risk factors”.

Another example of the challenge was that, initially, we represented the “Key Information Document” as a class of the KID ontology. However, during the implementation process, we observed that a class named “Key Information Document” didn’t provide any value to the ontology because it expressed the object from reality whose concepts were being formally represented and not a concept itself to be represented.

An additional challenge was related to the ambiguity between the terms. Ambiguity is an inherent part of the human language and happens when the same term has different meanings for the different contexts. Terms such as “Costs”, “Information Sites”, “Applicable Law, “Financial Instruments” and/or “Product Typology” may differ depending on who is interpreting the term. Another example of this issue manifests with the term “bond” since can refer to various types of debt-based securities, such as corporate bonds and/or government issued bonds. In the specific case of the KID ontology, some of these issues can be mitigated by adopting the definitions provided by supervisors and regulators that constitute a standard and stable understanding of a certain technical term.

It's relevant to state that the financial sector is an inherently complex environment, and the KID ontology classes may not fully capture the intricate nuances of the financial products, regulations and market conditions. The complexity around this matter is further increased by the fact that some European jurisdictions may adapt the PRIIP regulations according to their specific needs. Adhering to the specific nuances that are implemented in some jurisdictions can be a complex endeavor and impact reach of the KID ontology.

1.5. Evaluation

The evaluation phase is characterized by a technical perspective to assess the quality of the produced ontology. For this phase it used the inference system and SPARQL queries built-in Protégé editor for deduction purposes and exploration of the stability of the ontology knowledge and its overall consistency. By utilizing the inference system, some logical deductions were performed, helping to uncover potential inconsistencies and discrepancies within the ontology structure. Additionally, the integration of SPARQL queries within the Protégé editor facilitated the exploration of the ontology assertions.

1.5.1. Reasoning

For the evaluation of the ontology development process was used the reasoners available in the Protégé editor, specifically the Pellet reasoner, in order to ensure that no errors or inconsistencies were being made during the development process. The Pellet reasoner is one of the most used reasoners on Protégé editor, which is primarily designed to assist ontological reasoning and inferencing in OWL based ontologies. The Pellet reasoner allows the verification of inconsistencies in the ontologies by identifying axioms that lead to contradictory conclusions. Pellet also supports property reasoning, allowing the ontology to infer connections between individuals and properties. For instance, the object property "hasCosts" links the CFD to its associated costs and Pellet reasoning can infer the relationship for individuals without the need for explicit assertions.

The Pellet reasoner within Protégé is a valuable tool for performing reasoning of ontologies. It can help ensure ontology consistency, classify individuals, infer property relationships and detect inconsistencies within the ontology. These capabilities contribute to ensuring that the ontology is accurate, reliable and useful.

During the ontology development process the number of classes and properties represented was adapted throughout the ontology development process, since the logical inferences obtained from the execution of the Pellet reasoner allowed the observation of redundancies and inconsistencies between the originally identified classes and subclasses of the KID ontology. The solution of this issue paved the way for the simplification of the KID ontology that was originally drafted. For example, if the class "Manufacturer" is related to the class "CFD" and the class "CFD" is related to the class "Investor", the inference system can derive that the class "Manufacturer" is also related to the class "Investor" without the need to explicitly represent that fact.

With the assistance of the reasoners, it is possible to infer certain facts using axioms and inference rules based on a previous asserted set of facts. Put differently, the reasoner makes explicit the assertions that are only implicit in the formally represented concepts. Therefore, the inferred information is a corollary of the explicitly stated information.

In order for the inference engine correctly derive facts from the explicitly represented concepts, it was necessary to use the wide range of object properties available on the Protégé editor, mainly the functional, inverse functional and transitive.

The application of the Pellet reasoner to the KID ontology provided the extrapolation of inferred insights about the ontology, in particular:

i. Regarding the classification of Financial Instruments, the ontology has an instance called "Bonds" belonging to the class "CFD" and its necessary to classify it based on its product typology. Through the object property “hasProductTypology”, its inferred that “Bonds” is associated with a product typology and the reasoner can classify it being a product typology that a CDF can have as an underlying asset;

ii. Concerning the investor classification, the ontology has an instance called “Renato_Franco” belonging to the class “Investor” and we want to classify this investor based on its category. Using the data property “Investor_category” its possible to infer the investor category, which it is defined as “Non-professional” its possible to classify the investor “Renato_Franco” as a non-professional investor of CFD;

iii. Identification of competent courts. The ontology has an instance known as “Tribunais_de_Comarca” and “Tribunal_de_Relação_de_Lisboa” along with their associated legal framework. By using the object property “hasCompetentCourts” and associating each court with its relevant legal framework though data property assertions “Legal_Framework_name” it’s possible to determine which courts are competent considering the specific legal framework. For example, the “Tribunal_de_Comarca” and “Tribunal_da_Relação_de_Lisboa” has a data property assertation associated with the PRIIP regulation;

iv. Regarding the cost calculations, the ontology has instances of CFD with associated costs, such as “Entry_costs” and “Exit_costs”, and its required to calculate the total costs for a specific CFD including both the entry and exit costs. By using the object property “hasCosts” and data properties such as “Cost_name” and “Cost_value” its possible to perform such as calculations and infer the total costs of a certain CFD instance;

v. Concerning the market classification, the ontology has instances representing the financial market called “OTC”. By using the object property “hasMarkets” it’s possible to associate each financial instrument to the market;

vi. Regarding the performance scenarios, by using the object property “hasPerformanceScenarios” it’s possible to associate a CFD with its performance scenarios;

vii. Complaint channel contacts. The ontology has instances representing the complaint channels available, such as “Complaints”. By using the data property “ComplainChannel_Contact” it’s possible to specify the contact information for the complaint channel instances and its contact details;

viii. The ontology also has instances regarding the compensation schemes. By using the object property “hasCompensationSchemes” it’s possible to link CFDs to a specific compensation scheme which is given by the inference system.

The reasoning examples showcase the utility of the KID ontology regarding the identification of specific components that are present in the complex instruments KID, namely about the financial instruments, the manufacturers, the distributor, costs, performance scenarios, among others.

The KID ontology design, featuring classes such as “Financial_Instruments”, “Manufacturer”, “Distributor”, “Costs” and “Performance_Scenarios”, as well as their associated object properties and data properties, enables specific component identification. This allows to specify and relate elements within the KID documents content, namely the object property “hasCosts” connects a CFD to its cost components.

Furthermore, accurate and detailed information about performance scenarios, risk indicators and other components within KID documents is pivotal for assessing the potential risks associated to CFD financial instruments. The ontology helps the establishment of links between CFD and its relevant risk components, contributing to the identification of its risk factors.

However, considering the inherent complexity of the financial markets, the KID ontology could benefit with the addition of extra classes to represent a broader range of concepts within this domain and the enhancement of the links between those classes by adding more object properties and increasing the ontology complexity.

The KID ontology also lacks time-based properties in order to capture chronological aspects such as historical data and time-sensitive information.

1.5.2. Validation using SPARQL Queries

The validation process of an ontology is an important step in ensuring the consistency and accuracy of the represented domain by allowing the verification of asserted knowledge. SPARQL queries (Annex II) was used in order to validate the KID Ontology, allowing to interrogate the ontology knowledge and validate its alignment with the intended semantics.

The creation of SPARQL queries requires the use of a specific notation. The “Prefix” declarations serve the function of setting up and defining the namespaces employed within the query, enabling precise identification and retrieval of instances residing within the ontology. By utilizing the variable "?subject", it is possible to retrieve the instance matching the defined class within the KID ontology. By using the “select” in combination with the “distinct” keyword, the query ensures the extraction of the properties that are linked a specific class. The variable “?property” represents the properties, while the variable “?value” corresponds to the values associated with these properties.

The table below provides a description of the types of queries created for the purpose of validating specific aspects of the KID Ontology and the results obtained by running the queries (Table 5).

Query	Results
Lists the types of costs (Table 6, Annex II)	Performance_Costs Exit_Costs Transaction_Costs Entry_Costs
Lists the risks factors class (Table 7, Annex II)	Risk_Factors Market_Risks Credit_risks Liquidity_Risks
Lists the properties assertions of the performance costs (Table 8, Annex II)	Cost_name Cost_value hasCosts
Lists the compensation schemes class (Table 9, Annex II)	Investor_Compensation_Scheme Deposit_Guarantee_Fund
Lists the CFD with different underlying assets (Table 10, Annex II)	CFD_stocks CFD_cryptocurrency
Lists the holding period of CFD stocks (Table 11, Annex II)	No recommended holding period
Lists the market class (Table 12, Annex II)	Over-the-counter

Table 5 - SPARQL queries results. Source: The author.

1.5.3. Comparison with similar ontologies

The comparison process of the KID ontology with other existing ontologies within the financial domain revealed to be an unproductive endeavor because of its restricted scope. While there are a few financial ontologies available, there is no direct counterpart that mirrors the same level of detail and/or granularity in representing the KID documents of complex financial instruments, in particular CFD, and its components.

Furthermore, the KID ontology considers a regulatory environment that is specifically designed and applied within the European Union (EU), which sets it apart from many other financial ontologies that tend to adopt a more global approach or only focused on the United States of America regulatory environment. This focused attention on EU specific regulations, in particular the KID requirements under the PRIIPs regulation, caters to the specific needs of institutions, markets and investors operating within this region.

In summary, while there are financial ontologies available, the KID ontology focuses on a specific domain making it inviable to compare results through the comparison with other financial ontologies.

CHAPTER 5

Conclusions

1. General considerations

This chapter synthesizes the main conclusions obtained from the ontology development process. From the formalization of ontology classes, properties and instances, this final chapter reflects on the significance of this research within the broader landscape of ontology development process and financial markets supervision of KIDs. It provides a reflective analysis of the achieved goals and the extent to which research questions have been addressed and also outlines the limitations faced.

This dissertation aims to showcase the process of building an ontology applied to the capital markets, specifically regarding the key information documents drafted by issuers of complex investment products such as CFDs. The dissertation also highlights the challenges faced in the ontology development process and the impact of the inference system on its development.

The KID ontology goal is to create a formal knowledge representation of the KID required by the PRIIPs regulatory framework. The KID ontology is, however, limited to the CFDs. Despite the need to limit the scope of the dissertation to this specific financial instrument, CFDs are the most common derivative financial instrument traded by investors.

The ontology development process was mainly based on the Methontology methodology, an established approach that guides the systemic construction of ontologies, without relying on all the steps suggested by the said methodology. This ontology development methodology is characterized by a set of steps that includes the specification, implementation and evaluation phases.

As the descriptive potential of ontologies and the advantages of using the inference model were discovered, the inferred information allowed for the simplification of the ontology comparatively to the assertions made during its initial design phase. The initially designed ontology was large and complex and later we found to have redundancies and inconsistencies. In this regard, the reasoning process made it possible to identify and report errors in the ontology that otherwise would have passed unnoticed and update the KID ontology accordingly. In addition to the benefits of error identification and reports, the inference system also helped ensure the logical deduction from axioms that avoids descriptions that can be inferred. The reasoning process helped to alert for unpredicted interactions between the asserted classes. The ability to use the reasoners allowed us to focus on the knowledge acquisition and class description process and then test it for errors and inconsistencies identified by the inference system and then update it appropriately.

The KID ontology can also be used by artificial intelligence systems, including NLP, machine learning models and for knowledge sharing within this domain. Capital markets regulators across various jurisdictions, namely Europe, Asia, and the United States of America, are increasing their efforts to keep pace with financial innovation that was prompted mainly by the financial crisis of 2008. In this context, they have been testing NLP techniques to improve their legal compliance systems, namely by exploring the possibility to perform analysis on legal documents such as KID and increase the automation of their processes. To achieve this, ontologies like the KID ontology can help optimize those processes.

The dissertation is able to provide answers to the research questions stated above, in particular:

i. RQ1 - What is the advantage of using OWL to share and use knowledge as an ontology?

ii. RQ2 - Can ontologies capture the structure and semantics of classes of complex investment products such as PRIIPs?

Concerning the first question (RQ1), the use of OWL for sharing and utilizing knowledge in the form of an ontology offers multiple advantages. OWL's facilitates the representation of complex relationships and intricate domain-specific knowledge, fostering greater understanding of various concepts within the financial domain. OWL's formal semantics provide a standard framework for knowledge representation, allowing for the unambiguous interpretation of information about a particular domain and facilitating interoperability between various financial systems and applications. Additionally, OWL’s ability to describe axioms and logical constraints promotes reasoning capabilities, enabling the inference of new knowledge form existing data, which is particularly beneficial for making informed decisions and conduction comprehensive analysis within the financial industry. The use of the inference systems enabled us to avoid the redundancy of explicitly describing axioms that can be generated from the stated definitions and relationships between them. This allows simplifying the ontology development process, allowing for a more concise domain representation and avoiding the explicit definition of axioms. The inference systems’ ability to infer implicit links leads to a more efficient and enhanced representation of the knowledge domain. By generating logical reasoning and employing deduction tools, the inference system can generate knowledge and insights that is not explicitly defined within the ontology. This ability to derive new information from previously stated axioms reflects a form of computational intelligence that helps the ontology developers in ensuring the ontology coherence, accuracy and that it follows the established logical principles.

Concerning the second question (RQ2), ontologies, particularly those constructed using OWL, demonstrate a remarkable ability to capture the intricate structure and semantics of classes related to complex investment products such as PRIIPs. For instance, the ontology's "Risk_Factors class can be linked to specific details and subclasses of risk factors such as "Market_Risk" and "Credit_Risk", providing information about the risk profiles related to complex investment products. By leveraging OWL's modelling capabilities, ontologies can effectively represent the diverse attributes, relationships and characteristics of PRIIPs, including their underlying assets, performance scenarios, risk factors, legal frameworks, among others. OWL's support for defining class hierarchies, specifying properties and establishing logical connections allows for the comprehensive representation of the intricate components and features associated with PRIIPs, facilitating a more detailed and accurate understanding of these complex investment products. For instance, the ontology's representation of "Markets" and related properties such as "Market_name" allows to categorize the markets where the investment products are traded, providing information about the possible market segments where the derivatives instruments can be traded by the investors, which is usually OTC markets.

Additionally, OWL's formal representation capabilities enable the integration of PRIIPs related information from multiple sources, fostering a cohesive and standardized knowledge base for in-depth analysis and decision-making in the financial sector.

2. Practical considerations

The fact that ontologies enable the formal knowledge representation of a certain domain and data integration they have practical use for the institutions that operate in the capital markets sector. Specifically concerning the KID ontology, by structuring investment products relevant information such as risk factors, costs, performance scenarios, manufacturer, distributor and other information, the ontology is able to enhance data interoperability, support compliance management tools and regulatory oversight procedures. Furthermore, the utilization of Semantic Web technologies, namely OWL and SPARQL, contributes to the practical utility of these ontologies to the decision-making processes by facilitating data querying and the levering the inference mechanisms capabilities.

Regarding this matter, it is also relevant to mention that the authors of this dissertation have submitted a scientific paper (Annex II) and were invited to make an oral presentation at the ICST 2024: XVIII, International Conference on Semantic Technology to be held in Paris, France during April 11-12, 2024. The scientific paper highlights the relevance of ontologies like the KID ontology in the field of regulatory compliance through the convergence of ontological frameworks and NLP models applied to the capital markets.

3. Limitations and future work

While the developed ontology offers a certain degree of detail of the knowledge represented about the key information documents, it exhibits limitations and opportunities for future research. Some limitations include the lack of time related information, the lack of geographical information and market related information, which may impact specific transactional knowledge representation. The option to limit the scope of the dissertation to CFD derivative financial instruments may also impact the knowledge representation and data integration of other complex financial instruments that requires the disclosure of key information documents. Furthermore, the KID ontology does not accommodate the constant evolving regulatory framework, which may require ongoing updating to accommodate the changes that may arise.

The process of ontology reuse was taken into consideration, however the lack of research within this domain did not allow to leverage the advantages that the reuse of an existing ontology would do for this dissertation. This limitation consequently led to the dismissal of the ontology reuse process, as the lack of an established foundation within this specific domain of the KID significantly impeded the seamless integration and adaptation of pre-existing ontology structures.

Future research efforts could focus on refining the ontology to incorporate time related information, geographical and market information as well regulatory updates, enabling more dynamic and adaptive knowledge representation. Additional integration of other financial instruments may also help refine the ontology. Further exploration of advanced reasoning systems and machine learning techniques can enhance the ontology's capabilities in analysing, for instance the market trends, risk factors, the costs and performance scenarios, thereby providing more accurate insights for the market participants and supervisory bodies. Also, additional research can be directed towards the development of compliance tools and decision support systems that leverages the KID ontology. It is also possible to broaden the ontology scope in order to incorporate different complex financial instruments and more diversity of underlying assets.

The ontology’s importance for NLP models employed by capital markets supervisory bodies may assist them to promptly identify non-compliant or misleading disclosures in the key information documents, monitor market risks and ensure a more dynamic investor protection. The ontology’s integration with NLP models serves as a powerful tool for automating the analysis of key information documents, enhancing the supervisory activities and promoting transparency to financial markets.

References

Amzallag, A., Bagattini, G., and Linz, L. (2022). Parsing prospectuses: A text-mining approach, ESMA Report on Trends, Risks and Vulnerabilities.

Armstrong, P., and Harris., A (2019). RegTech and SupTech – change for markets and authorities, ESMA Report on Trends, Risks and Vulnerabilities, no. 1.

Berners-Lee, Tim & Hendler, James & Lassila, Ora. (2001). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. ScientificAmerican.com.

Borst, W.N. (1997). Construction of Engineering Ontologies for Knowledge Sharing and Reuse.

Chhim, P., Chinnam, R.B. & Sadawi, N (2019). Product design and manufacturing process based ontology for manufacturing knowledge reuse. J Intell Manuf 30, 905–916. https://doi.org/10.1007/s10845-016-1290-2.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., López-Cima, A. (2005). Building Legal Ontologies with METHONTOLOGY and WebODE. In: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds) Law and the Semantic Web. Lecture Notes in Computer Science, vol 3369. Springer.

Liddy, E.D. (2001). Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.

Fikes, R., & Farquhar, A. (1999). Large-scale repositories of highly expressive reusable knowledge. IEEE Intelligent Systems, 14(2), 73-79.

Financial Stability Board (2020), The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions.

G. G. Petrova et. al. (2017). Application of the Financial Industry Business Ontology (FIBO) for development of a financial organization ontology, J. Phys.: Conf. Ser. 803 012116.

Gordon, T.F. (2010). An Overview of the Legal Knowledge Interchange Format. In: Abramowicz, W., Tolksdorf, R., Węcel, K. (eds) Business Information Systems Workshops. BIS 2010. Lecture Notes in Business Information Processing, vol 57. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15402-7_30.

Graf, S., Kling, A (2020). PRIIP-KID: appearances are deceiving or why to expect the unexpected in a generic KID for multiple option products. Eur. Actuar. J. 10, 527–555. https://doi.org/10.1007/s13385-020-00243-0.

Guarino, Nicola. (1998). Formal Ontologies and Information Systems, Conference: FOIS'98 ConferenceAt: Trento, Italy.

Guarino, N. and Giaretta, P. (1995). Ontologies and Knowledge Bases. In: Towards Very Large Knowledge Bases, IOS Press, Amsterdam, 1-2.

Javed, M.A., Muram, F.U., Kanwal, S. (2022). Ontology-Based Natural Language Processing for Process Compliance Management. In: Ali, R., Kaindl, H., Maciaszek, L.A. (eds) Evaluation of Novel Approaches to Software Engineering. ENASE 2021. Communications in Computer and Information Science, vol 1556. Springer, Cham. https://doi.org/10.1007/978-3-030-96648-5_14.

Maple, C., Szpruch, L., Epiphaniou, G., Staykova, K., Singh, S., Penwarden, W., and Avramovic, P. (2023). The AI Revolution: Opportunities and Challenges for the Finance Sector. arXiv preprint arXiv:2308.16538.

Mika, Peter. (2007). Ontologies are Us: A Unified Model of Social Networks and Semantics. SSRN Electronic Journal. 5. 10.2139/ssrn.3199347.

Motik, B., Patel-Schneider, P., and Parsia, B. (2012). OWL 2 Web Ontology Language

Structural Specification and Functional-Style Syntax, https://www.w3.org/TR/owl2-syntax/#Axioms.

Perchet, Romain and Herambourg, Nicolas and Leote de Carvalho, Raul (2023). PRIIPs Regulation Unwrapped: Essential Aspects and Practical Implications. SSRN Electronic Journal. 10.2139/ssrn.4388183.

Pinto, H., Martins, J (2004). Ontologies: How can They be Built?. Know. Inf. Sys. 6, 441–464.

Thomas R. Gruber (1993). A translation approach to portable ontology specifications, Knowledge Acquisition, Volume 5, Issue 2.

Thomas R. Gruber (1993). A translation approach to portable ontology specifications, Knowledge Acquisition, Volume 5, Issue 2.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. The Knowledge Engineering Review, 11(2), 93-136. doi:10.1017/S0269888900007797.

Annex I

In the context of the dissertation, a set of SPARQL queries were conducted to demonstrate the consistency and accuracy of the ontology. Below we list the SPARQL codes used to test the ontology.

a) Lists the types of costs

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>

SELECT ?subject
        WHERE { ?subject a untitled-ontology-35:Costs }
```

Table 6 - Lists the types of costs. Source: The author

b) List the risks factors class

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>
SELECT ?subject
        WHERE { ?subject a untitled-ontology-35:Risk_Factors }
```

Table 7 - List the risks factors class. Source: The author.

c) List of properties assertions of the performance costs

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>

select distinct ?property ?value where {
untitled-ontology-35:Performance_Costs ?property ?value .
  filter ( ?property not in ( rdf:type ) )}
```

Table 8 - Lists the properties assertions of the performance costs. Source: The author.

d) List the compensation schemes class

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>
SELECT ?subject
        WHERE { ?subject a untitled-ontology-35:Compensation_Schemes}
```

Table 9 - List of compensation schemes class. Source: The author.

e) Lists the CFD with different underlying assets

.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>
SELECT ?subject
        WHERE { ?subject a untitled-ontology-35:CFD}
```

Table 10 - Lists the CFD with different underlying assets. The author.

f) List the holding period of CFD stocks

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>
SELECT  ?period
        WHERE {
untitled-ontology-35:CFD_stocks untitled-ontology-35:HoldingPeriod_duration ?period}
```

Table 11 - List the holding period of CFD stocks. Source: The author

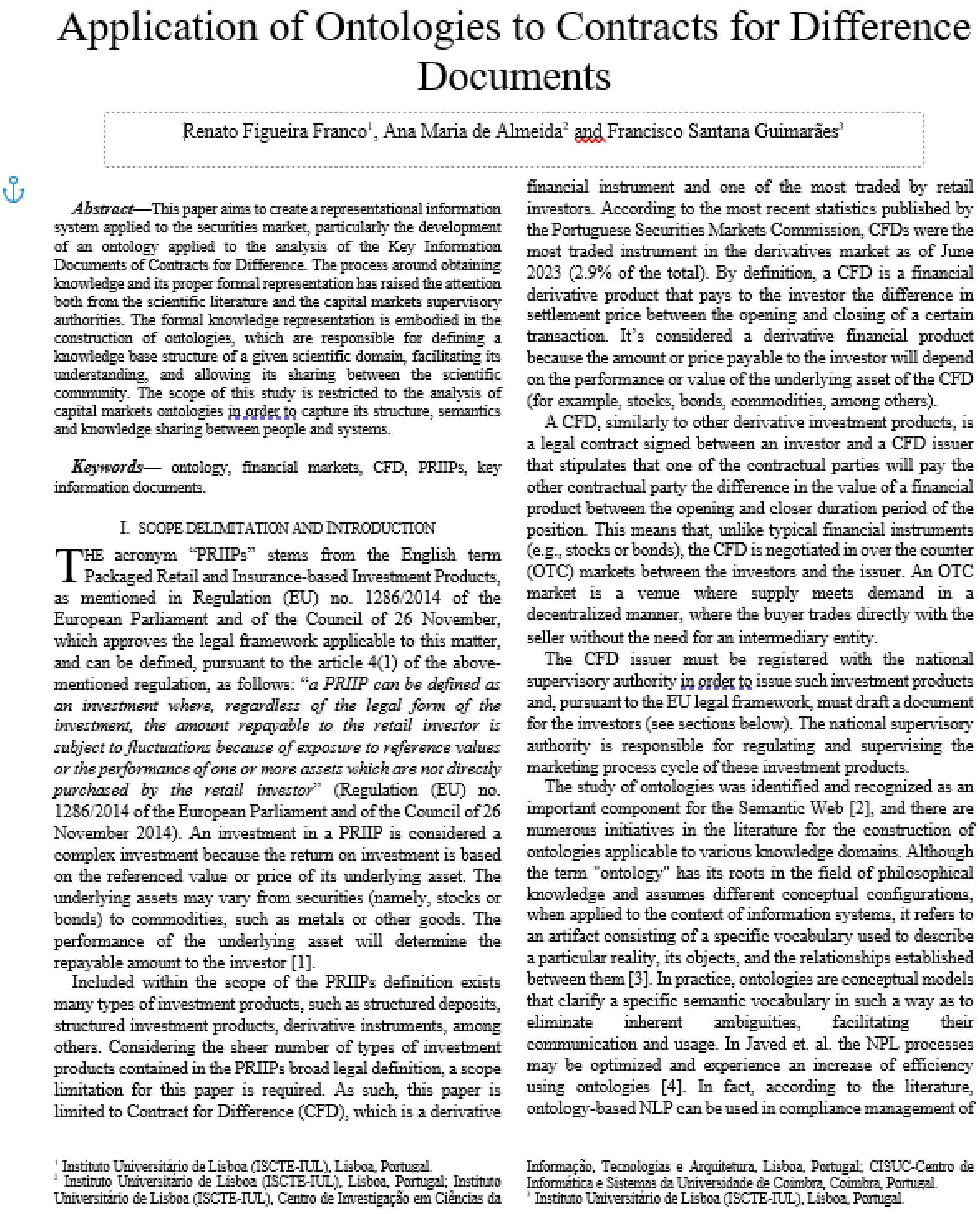
g) Lists the market class

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX untitled-ontology-35:
<http://www.semanticweb.org/renato/ontologies/2023/8/untitled-ontology-35#>
SELECT  ?name
        WHERE {
untitled-ontology-35:OTC untitled-ontology-35:Market_name ?name}
```

Table 12 - Lists the market class. Source: The author.

Annex II

Scientific paper submitted and accepted to an oral presentation at the ICST 2024: XVIII, International Conference on Semantic Technology to be held in Paris, France during April 11-12, 2024.



Application of Ontologies to Contracts for Difference Documents

Renato Figueira Franco[1], Ana Maria de Almeida[2] and Francisco Santana Guimarães[3]

Abstract—This paper aims to create a representational information system applied to the securities market, particularly the development of an ontology applied to the analysis of the Key Information Documents of Contracts for Difference. The process around obtaining knowledge and its proper formal representation has raised the attention both from the scientific literature and the capital markets supervisory authorities. The formal knowledge representation is embodied in the construction of ontologies, which are responsible for defining a knowledge base structure of a given scientific domain, facilitating its understanding, and allowing its sharing between the scientific community. The scope of this study is restricted to the analysis of capital markets ontologies in order to capture its structure, semantics and knowledge sharing between people and systems.

Keywords— ontology, financial markets, CFD, PRIIPs, key information documents.

I. SCOPE DELIMITATION AND INTRODUCTION

THE acronym "PRIIPs" stems from the English term Packaged Retail and Insurance-based Investment Products, as mentioned in Regulation (EU) no. 1286/2014 of the European Parliament and of the Council of 26 November, which approves the legal framework applicable to this matter, and can be defined, pursuant to the article 4(1) of the above-mentioned regulation, as follows: "*a PRIIP can be defined as an investment where, regardless of the legal form of the investment, the amount repayable to the retail investor is subject to fluctuations because of exposure to reference values or the performance of one or more assets which are not directly purchased by the retail investor*" (Regulation (EU) no. 1286/2014 of the European Parliament and of the Council of 26 November 2014). An investment in a PRIIP is considered a complex investment because the return on investment is based on the referenced value or price of its underlying asset. The underlying assets may vary from securities (namely, stocks or bonds) to commodities, such as metals or other goods. The performance of the underlying asset will determine the repayable amount to the investor [1].

Included within the scope of the PRIIPs definition exists many types of investment products, such as structured deposits, structured investment products, derivative instruments, among others. Considering the sheer number of types of investment products contained in the PRIIPs broad legal definition, a scope limitation for this paper is required. As such, this paper is limited to Contract for Difference (CFD), which is a derivative financial instrument and one of the most traded by retail investors. According to the most recent statistics published by the Portuguese Securities Markets Commission, CFDs were the most traded instrument in the derivatives market as of June 2023 (2.9% of the total). By definition, a CFD is a financial derivative product that pays to the investor the difference in settlement price between the opening and closing of a certain transaction. It's considered a derivative financial product because the amount or price payable to the investor will depend on the performance or value of the underlying asset of the CFD (for example, stocks, bonds, commodities, among others).

A CFD, similarly to other derivative investment products, is a legal contract signed between an investor and a CFD issuer that stipulates that one of the contractual parties will pay the other contractual party the difference in the value of a financial product between the opening and closer duration period of the position. This means that, unlike typical financial instruments (e.g., stocks or bonds), the CFD is negotiated in over the counter (OTC) markets between the investors and the issuer. An OTC market is a venue where supply meets demand in a decentralized manner, where the buyer trades directly with the seller without the need for an intermediary entity.

The CFD issuer must be registered with the national supervisory authority in order to issue such investment products and, pursuant to the EU legal framework, must draft a document for the investors (see sections below). The national supervisory authority is responsible for regulating and supervising the marketing process cycle of these investment products.

The study of ontologies was identified and recognized as an important component for the Semantic Web [2], and there are numerous initiatives in the literature for the construction of ontologies applicable to various knowledge domains. Although the term "ontology" has its roots in the field of philosophical knowledge and assumes different conceptual configurations, when applied to the context of information systems, it refers to an artifact consisting of a specific vocabulary used to describe a particular reality, its objects, and the relationships established between them [3]. In practice, ontologies are conceptual models that clarify a specific semantic vocabulary in such a way as to eliminate inherent ambiguities, facilitating their communication and usage. In Javed et. al. the NPL processes may be optimized and experience an increase of efficiency using ontologies [4]. In fact, according to the literature, ontology-based NLP can be used in compliance management of

[1] Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal.
[2] Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura, Lisboa, Portugal; CISUC-Centro de Informática e Sistemas da Universidade de Coimbra, Coimbra, Portugal.
[3] Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal.

software engineering processes to analyze standardized documents.

However, in the domain of financial markets, particularly in capital markets and especially in the field of complex investment products, the interest in developing ontologies has lacked enthusiasm. Despite this fact, it's critical to emphasize the growing interest of the stakeholders in the usage of natural language processing (NPL) models to analyze these types of documents [5]. New regulatory and supervisory technologies are being developed and enhanced in order to improve the detection of fraud capabilities of the supervisors and other issues, such as regulatory reporting information, risk management and data collection. In this regard, in the beginning of 2019, the European Securities and Markets Authority (ESMA) began exploring the application of NPL to the analysis of more than 20,000 KIDs, from 500 issuers, in 21 EU languages [6].

The importance of building ontologies that facilitate the communication of certain concepts becomes evident when seeking to build information systems that process information more efficiently, accurately, and quickly. In addition, ontologies facilitate and promote interoperability between information systems, allowing for a common understanding of a particular domain of reality that is intended to be described in such a way that it can be communicated by humans who come into contact with it, as well as by information systems that use it [7].

The knowledge base represented in the KID Ontology is the key information documents used as a vehicle to convey information about PRIIPs.

The Key Information Document (KID) is a pre-contractual information document, clearly distinct from promotional materials, which provides key information for non-professional investors to fully understand and compare the main characteristics, risks and returns, and costs of the investment product in which they intend to invest. The KID has a legally defined format and content, structured into sections. In addition to the initial sections entitled "Purpose" and "Product," and a final section called "Other Relevant Information," reserved for any additional relevant information, there are six more sections legally provided for.

Despite these current developments, we've observed that there is scarcity of ontologies related to capital markets, specifically focusing on the key information documents. As a result of these recent developments and shift in perspective regarding the usage of supervisory and regulatory technologies, this paper seeks to create an ontology for the key information documents.

This study is divided into two main phases. The first phase involved examining the text of the key information documents in order to extract its structure and semantics as the main relevant information. Specifically, it was necessary to thoroughly examine all the sections of the document to enable the representation of the entire knowledge within this specific domain, i.e., the domain of the key information documents required by the legislation to be drafted by a manufacturer of PRIIPs.

The second phase involves the creation of the ontology in OWL language, using the Protégé editor, based on the knowledge acquired in the previous phase. The design of the KIID ontology comprises five main activities: (i) specification; (ii) knowledge acquisition; (iii) formalization; (iv) implementation and (v) evaluation [8].

This specific paper is organized as follows. Section 2 provides a brief review of the related work on capital markets ontologies and an overview of the structure of the key information documents. Section 3 is dedicated to ontology development. The last section is dedicated to future work.

I. RELATED WORK ON CAPITAL MARKETS ONTOLOGIES

Traditionally, the term ontology has been identified as a category within the field of philosophical studies, dedicated particularly to the study of the ontic reality, mainly focused on the Being and its essential characteristics. More recently, this term has been applied by many authors to the context of information systems but with a different meaning. In this specific field of knowledge, ontologies are envisioned as an information representation tool capable of collecting, mapping, and disseminating knowledge of specific fields of study. Fundamentally, ontologies are abstract models that allow the architecture of a lexicon of technical expressions used in a particular scientific domain, enabling a language free of ambiguity which facilitates the transfer and usage of knowledge through all stakeholders.

In the literature there are several attempts to define the term "ontology", with some notable notions being that ontology consists of an explicit and formal specification of a shared conceptualization [9]. The specification refers to concepts, attributes, relations, and axioms that are explicitly predefined. According to this definition, formalization refers to its interpretability by the information systems, the conceptualization refers to the abstract model of a particular phenomenon in the real world that is being mapped, and sharedness refers to consensus in the community [10].

In Guarino and Giarreta [11], the authors take on a divergent approach from Gruber's. For them, ontologies are both a partial and explicit description of specific concepts. According to these authors, ontology fulfills two essential purposes: (i) conceptualization should be a true syllogism independent of the different subjects who are going to utilize it. This means that the ontology should be based on a conceptual construction independent of subjective dimensions, as it is intended to be used as a basis for sharing knowledge of a specific domain; and (ii) ontology should consist of a set of premises through which strict restrictions are developed according to inferences designed to be shared by users who agree with its conceptual construction.

In Fikes and Farquhar [12], ontology is the study of a specific domain that defines a lexicon of entities, classes, properties, predicates, functions, and a set of relationships that necessarily exist between such concepts. Such a definition is considered one of the most comprehensive in scientific literature as it clearly identifies all the attributes that are inherent to every ontology.

The research fields related to ontologies are spreading and thriving in computer science and encompass multiple areas of knowledge. There are plenty of benefits that can be obtained through the ontological formalization of a specific domain of knowledge, and ontologies are being used in many and diverse scientific domains, including natural language processing, knowledge representation, knowledge management, among others.

Despite the wide range of ontological notions that appear in academic literature, the formalization of a knowledge domain through an ontology will always result in a language that represents the existing knowledge about one specific domain.

The scientific literature identified a set of essential elements in the development process of an ontology, namely [13]:

I. Classes: organize concepts associated with a particular domain, constructed based on a taxonomy;

II. Relationships: represents the type of interaction established between classes in a particular domain;

III. Instances: examples or use cases of classes used to represent specific objects; and

IV. Competency questions: questions designed to be answered by the ontology. They help define the scope and characteristics of the ontology, specify the tasks and problems to be addressed.

Ontologies have thrived in various areas of expertise in scientific literature, namely within the legal domain. However, legal ontologies have certain features that the differs from the ontologies in other areas [14]. As legal rulings must be justified by reason and supported by solid evidence, legal ontologies are more inclined to cover epistemological concepts, such as norms, court, contract, legal and/or natural person, role, duties, rights, responsibility, property, crime, interpretation, sanction, delegation, legal documents, among others [15].

In this phase, it is relevant to emphasize some characteristics that can be observed in legal ontologies. The Law relies on documents to support the reasoning behind any legally binding decisions. That's why documents are the main infrastructure behind all legislative processes. Documents have a three main dimensions: (i) the physical dimension (which is the document); (ii) the representational dimension (the form in which the language is represented); and (iii) the cognitive dimension (which is the intended content by its author).

In knowledge and information management, there are a variety of types of documentation and its structures. Documents may range from narrative texts (stories, histories, case descriptions, testimony) via "non-narrative" texts (reports, articles, handbooks, instructions) to fully pre-structured filled-in forms. Also, they range from "primary sources of law", i.e., codes or regulations to legal instruments created to determine rights in private transactions, such as deeds or wills.

In the field of capital markets, not many ontologies are mentioned by the scientific literature. However, the Financial Industry Business Ontology (or FIBO) has been extensively cited among the authors that research this specific domain. The FIBO ontology provides relationships between financial constructs, provide high-level descriptions, and help its users to describe the financial business, namely regarding legal entities, market data, contracts, and the contractual obligations the arise from them and for many different financial instruments (e.g., Contracts for Difference, Swaps, Options, Futures, Forwards, and many others) [16].

One way to represent an ontology in FIBO is from a formal OWL description made with the Protégé ontology editor, which was an editor developed by a research team from Stanford University. The FIBO ontology can be used by anyone interested in working in the financial sector. As stated above, the FIBO ontology provides a large set of financial business-related notions, definitions, and relations between them with which organizations can use as a complement to their own models of the field.

FIBO can be more accurately described as an intricate web of ontologies rather than just a single master ontology. This web is divided into subcategories and some of them contain sets of shared ontologies that link to other subcategories. The ontologies that make up the FIBO "web" are based on the top-level ontology including groups called "sections". In turn, these sections contain a description of various types of fundamental constructs. Such formal models allow for separate description, application and extension of concept groups contained within separate modules.

Also, within the financial markets field of study, exists the Bank Regulation Ontology (BRO), the Financial Regulation Ontology (FRC) and the Legal Knowledge Interchange Format (LKIF). The BRO is a structured and comprehensive knowledge representation framework designed to capture and model the intricate landscape of regulatory guidelines, regulations and standards in the banking industry sector. It serves as a valuable resource for regulatory authorities, financial institutions, researchers, and policy makers to enhance their understating and compliance with the banking regulations. This ontology is particularly important in the context of a highly regulated industry where compliance with various regulatory frameworks is essential for maintaining financial stability and safeguarding the interests of the stakeholders.

At its foundations, the BRO categorizes and defines key concepts and connections related to the banking sector regulations. It encompasses high-level categories such as capital adequacy, risk management, consumer protection and reporting requirements. Each individual category is further refined to include specific regulations and guidelines issued by regulators at national and international levels. For example, it may include Basel III standards Dodd-Frank Act legal provisions and from the Financial Stability Board.

The ontology also captures temporal aspects by tracking the evolution of the regulations over time. In fact, it can express revisions and effective dates of regulatory documents in order to ensure the most up to date information concerning that specific topic. The BRO also incorporates semantic connections between regulations, such as dependencies, conflicts and hierarchical relations, enabling users to navigate the complex web of regulatory requirements and assess their impact on financial institutions.

Overall, the BRO serves as a valuable tool for regulatory compliance and risk management within the banking sector.

The ontology aims to provide transparency, consistency and efficiency regarding the compliance of regulatory requirements.

Regarding the FRO, these are specialized knowledge structures designed to systematically capture, represent and organize the area of financial regulations. These ontologies are essential in creating complex legal frameworks within the financial sector in a machine-readable and semantically rich format. The difference between the FRO and BRO mentioned above is that the FRO focuses on modelling legal frameworks beyond the banking sector and encompasses the sector of the financial markets. Contrarywise, the BRO focuses on modelling and representing the legal framework applicable to the banking sector.

The FRO encompasses a wide range of regulatory domains, including banking, insurance and capital markets. Each ontology is composed to represent a specific law or regulations issued by national and/or international regulators. Similarly, to the BRO, the FRO is constantly being updated in order to assure the most accurate and recent knowledge about the regulatory frameworks mapped and constantly refined to keep pace with evolving regulatory environments and emerging compliance challenges that arise within the finance sector.

Another example of relevant related work is the Legal Knowledge Interchange Format (LKIF), which is a specialized framework designed to assist the structured representation and exchange of legal knowledge and information withing the field of law and jurisprudence. LKIF is a standardized ontology that leverages semantic technologies to encode legal concepts, rules, regulations and legal documents in a machine-readable format. Its main objective is to enhance accessibility, interoperability and understanding of legal information, making it a valuable resource for legal professionals, scholars and policymakers.

The LKIF employs ontological concepts that define legal entities, namely laws, legal proceedings, judges, lawyers and other legal related concepts. Moreover, LKIF enables the integration of legal knowledge with other domains, such as natural language processing, which helps the integration with legal applications, reasoning systems, contract analysis tools and legal information capture systems. LKIF adherence to semantic web standards and principles ensures that legal information can be seamlessly integrated with other knowledge domains, fostering interdisciplinary collaborations and enhancing the capabilities of legal technology solutions.

In summary, the LKIF is a standardized ontology for encoding and sharing legal knowledge in a structured and machine-readable format, allowing the comprehension of legal concepts which makes it an important resource for legal research and legal practice.

A. KID analysis by using Natural Language Processing

The field of research of Artificial Intelligence (AI) is comprised of many subfields, one of which is Natural Language Processing (NLP). This subfield of AI aims to train and enable computers to comprehend, interpret, and generate human language, enabling to obtain efficiency gains in communication between humans and the machines by providing the means to the latter to read human language.

By using NLP models, it is possible to analyze text based on a predefined set of rules and techniques. This capability is being leveraged by the financial sector in order to acquire efficiency improvements in the services provided to the clients, such as chatbots for improving the customer experience, but also by helping the supervisory activity of the financial markets regulators by enabling the analysis of large volumes of documentation.

The regulatory activity of the financial market supervisors requires the analysis of several legal documentation provided by the financial market operators to the investors, such as KIDs, prospectuses, financial statements, policies and procedures, among other information. In order to address this issue, “the European Securities and Markets Authority (ESMA) began exploring the use of NLP to analyze the information of more than 20,000 KIDs, from more than 500 issuers, in 21 European languages”.

More recently, ESMA published a report where they presented the results of its endeavor to apply NLP methods to a dataset of 3,220 documents with more than 593,000 pages of text. The overall results were positive and ESMA concluded that the algorithms behind NLP solutions opens new possibilities for helping the analysis of large volumes of information and lengthy documents.

As mentioned above, in Javed et. al., an ontology-based NLP can be used in compliance management of software engineering processes to analyze standardized documents.

In this sense, ontologies play a determining role in the formalization of knowledge as pillar above which is going to be built the NLP systems that optimize compliance management software tools.

B. Key Information Document Structure

The aftermath of the 2008 financial crisis witnessed concerted efforts to develop KIDs for PRIIPs, with the overarching objective of improving consumer comprehension regarding financial products, specifically their risk, reward profiles and costs and charges associated with those financial instruments [17].

Insufficient understanding was identified as a significant factor contributing to the unanticipated losses experienced by certain investors during the financial crisis. Against this backdrop, the primary impetus behind PRIIP regulation has been to foster transparency and facilitate the comparison of diverse products for retail investors through the provision of a comprehensible pre-contractual document.

The pursuit of consistent transparency rules at the European Union level aimed to mitigate discrepancies and bolster investor protection, considering the variations that had previously existed in disclosure requirements across sectors and Member-States. These distinctions between the applicable legal framework of the countries impeded or even precluded the comparison of distinct financial instruments, resulting in an inequitable playing field among different investment products and distribution channels. The PRIIP regulation sough to resolve this issue by mandating improved disclosures in a uniform manner for all products, irrespective of their form or structure, with the ultimate objective of rebuilding retail investors’ confidence in financial markets.

Regulation no. 1286/2014 of the European Parliament and of the Council of 26 November 2014, set the framework applicable to PRIIPs. This legal document underscored the indispensability of disclosure requirements for investment products, enabling retail investors to comprehend the risks entailed when making investment decisions. Traditionally, both manufacturers and distributors drafted a prospectus for each product marketed to investors. Because these documents were too complex to be understood by the average investor, they heavily relied on the advice and explanations provided by distributors.

To mitigate this issue, the European Commission introduced a preliminary draft of regulatory technical standards for the KID. As mentioned above, the KID is a pre-contractual information document provided to retail investors prior to making any investment decision. To ensure that the retail investors make an informed decision regarding a specific investment product, the KID must be drafted with accuracy, correctness, and clarity.

The information conveyed through the KID must be consistent with the information contained in other documents, namely prospectus, particularly regarding the terms and conditions of the PRIIP.

The legal framework that regulates the drafting of the KID requires this document to be concise and succinct, consisting of maximum of three A4-sized pages, and should ensure the comparability between different PRIIPs. The entire document should be drafted with characters that are easily legible to any readers.

According to the PRIIP regulation, the KID aims to:

1. Provide general information regarding the investment product.
2. Identify the degree of the risk for each PRIIP in the form of a risk class by using a synthetic risk indicator (SRI).
3. Identify the performance scenarios; and
4. Identify all the costs and charges related to the PRIIP.

In terms of its structure, the PRIIPs Regulation stipulates that the KID should adhere to a standardized format, emphasizing brevity. This is crucial to avoid overwhelming the reader with excessive information overload. In this context, the KID should be drafted by the manufacturer in a concise way and without any unnecessary details that do not contribute to making well-informed investment decisions.

To address the issue of length, the PRIIPs Regulation has introduced a formal maximum of three A4-sized pages when printed. This measure was influenced by the experience obtained from implementing the summary prospectus pursuant to the EU Prospectus Directive.

To fulfill the information needs of retail investors and enable easy comparison of various PRIIPs, it is crucial that the document is prepared in a standardized format. This ensures consistent ordering of items and adherence to uniform headings across all KIDs.

In terms of quality of information requirements, the KID is expected to present precise, equitable, lucid, and non-misleading information that aligns with other binding documents and terms and conditions applicable to the specific PRIIP. Furthermore, the PRIIPs Regulation emphasizes that the KID should be an independent and unbiased document, clearly separated from any marketing materials. It should also refrain from including cross-references to promotional content.

The KID structure can be divided between 12 sections or areas, each one designed to include specific information:

i. Section 1 should contain a comprehension alert to the investor.

ii. Section 2 should contain general information about the product.

iii. Section 3 contains information about the PRIIP (e.g., type of product, supervisory authority, production date, contact details about the manufacturer.

iv. Section 4 contains information about the risks.;

v. Section 5 contains information about the performance scenarios.

vi. Section 6 contains information regarding the investor compensation scheme.

vii. Section 7 contains information about the costs and charges to investors.

viii. Section 8 contains information about the investment impacts.

ix. Section 9 contains detailed information about the costs and charges.

x. Section 10 contains information about the recommended detention period and penalties applied.

xi. Section 11 contains information about the complaint channels available to the investor; and

xii. Section 12 contains other relevant information.

I. Ontology development for key information documents

The KID ontology was developed using Protégé editor that supports a wide range of formats (e.g., RDF, OWL, and others). The domain concepts of this ontology are described in OWL Language, an XML-based language.

The ontology development process followed the main activities: (i) specification; (ii) knowledge acquisition; (iii) formalization; (iv) implementation and (v) evaluation. Using the Protégé editor allows for the creation of a hierarchy of concepts (classes), which can be further categorized and enables the definition of connections between those classes.

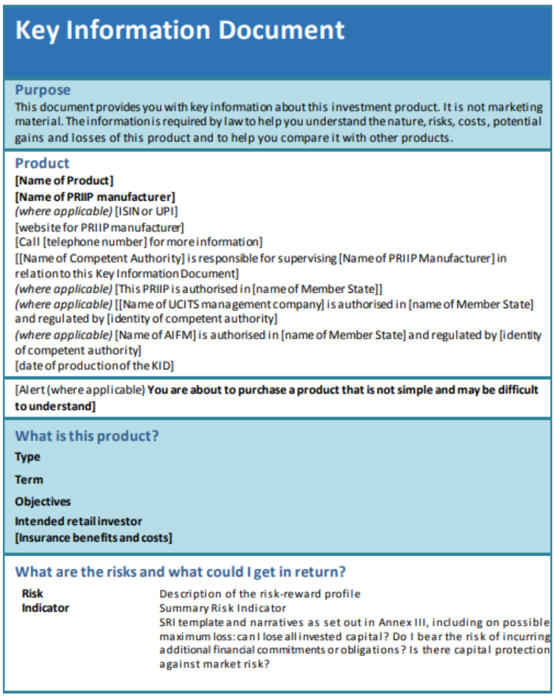
A. Specification

The purpose, scope and the degree of detail are essential parameters that need to be defined for any ontology development process.

The scope of this study includes the creation of an instantiated knowledge database that describes the concepts and the properties of the KID developed by any manufacturer of PRIIPs using the template provided by the legislation. As mentioned above, the instantiated knowledge is restricted to certain types of PRIIPs, particularly the CFDs which are the main derivatives instruments traded by retail investors.

A. Knowledge acquisition

This stage is characterized by the data extraction from various sources. The main source of knowledge used to produce the KID ontology structure is from the KID template structure that has been published in the Annex I of the Commission Delegated Regulation (EU) no. 2021/2268 of 6 September 2021, as shown in Fig. 1:



Key Information Document

Purpose

This document provides you with key information about this investment product. It is not marketing material. The information is required by law to help you understand the nature, risks, costs, potential gains and losses of this product and to help you compare it with other products.

Product

[Name of Product]
[Name of PRIIP manufacturer]
(where applicable) [ISIN or UPI]
[website for PRIIP manufacturer]
[Call [telephone number] for more information]
[[Name of Competent Authority] is responsible for supervising [Name of PRIIP Manufacturer] in relation to this Key Information Document]
(where applicable) [This PRIIP is authorised in [name of Member State]]
(where applicable) [[Name of UCITS management company] is authorised in [name of Member State] and regulated by [identity of competent authority]
(where applicable) [Name of AIFM] is authorised in [name of Member State] and regulated by [identity of competent authority]
[date of production of the KID]

[Alert (where applicable) **You are about to purchase a product that is not simple and may be difficult to understand]**

What is this product?

Type

Term

Objectives

Intended retail investor

[Insurance benefits and costs]

What are the risks and what could I get in return?

Risk Indicator

Description of the risk-reward profile
Summary Risk Indicator
SRI template and narratives as set out in Annex III, including on possible maximum loss: can I lose all invested capital? Do I bear the risk of incurring additional financial commitments or obligations? Is there capital protection against market risk?

Fig. 1 - KID Template: Commission Delegated Regulation (EU) no. 2021/2268

In this context, the relevant information was extracted from a sample of 13 KIDs of CFDs issued by Portuguese financial entities, actively marketed in Portugal, and regulated by the Portuguese Securities and Markets Commission (CMVM). The specific characteristics of the financial entities, such as dimension or volume of CFDs traded, did not matter for the purpose of the KID Ontology. The KID sample considered was limited to CFDs and obtained from the CMVMs website.

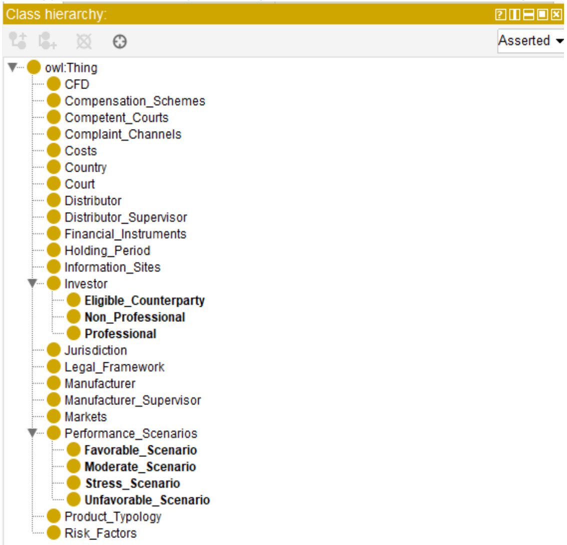
The information obtained from those documents was then converted into classes, subclasses, object property and data properties (described in the section below).

The process of ontology reuse was taken into consideration for the KID Ontology, however the lack of investigation within this domain, and specifically applied to the European framework, did not allow to leverage the advantages that the reuse of an existing ontology would do for this project. As a result of this limitation the process of ontology reuse was disregarded.

In this context, most of the knowledge acquisition and its analysis was focused on understanding the concepts in terms of classes and/or data properties, and the existing connections between the terms as object properties by comparing the different KIDs used as a sample for this paper, the specification of those KIDs and the way from which a sentence one is able to identify what is a class (the subject of a sentence) and the relation between classes (the predicate of the sentence).

B. Formalization

The information obtained from the KID template was converted into classes, datatype properties and object properties. By default, the KID ontology root domain is a class called "owl: Thing", this is the main class that represents all the subsequent classes, hence all succeeding classes are a subclass of "owl: Thing". The formalization of the classes for the KID ontology are created by adding subclasses, as shown in Fig. 2:



- owl:Thing
  - CFD
  - Compensation_Schemes
  - Competent_Courts
  - Complaint_Channels
  - Costs
  - Country
  - Court
  - Distributor
  - Distributor_Supervisor
  - Financial_Instruments
  - Holding_Period
  - Information_Sites
  - Investor
    - **Eligible_Counterparty**
    - **Non_Professional**
    - **Professional**
  - Jurisdiction
  - Legal_Framework
  - Manufacturer
  - Manufacturer_Supervisor
  - Markets
  - Performance_Scenarios
    - **Favorable_Scenario**
    - **Moderate_Scenario**
    - **Stress_Scenario**
    - **Unfavorable_Scenario**
  - Product_Typology
  - Risk_Factors

Fig. 2 - KID Ontology class hierarchy

Classes on OWL can be defined as representations of a concept about a certain domain. So, for example, a class named "Investor" encompasses everyone that invests in financial instruments. Classes can be further specified if needed. Such specifications of a given notion are referred to as subclasses and they introduce granularity to the concept they are associated with. Building upon the previous example, the class "Investor" can be further specified to "Professional", "Retail" and "Eligible Counterparty" and have these represented as subclasses.

The classes represented in ontology have properties that connect them to other classes, that make up a relation between them. For example, the class "Investor" has a location which is given by the class "Country". The properties have a characteristic called inheritance, which means that if a class X has a connection with class Y by the property N, all the subclasses of X are linked to the subclasses of Y by the property N. By default, the object property root domain is called "owl:topObjectProperty" and the KID Ontology is comprised of 19 object properties. Although OWL provides a wide range of object properties (functional, inverse functional, symmetric, and transitive), the KID Ontology uses the functional, inverse functional and transitive properties. These characteristics, which describe how the object properties behave, can be defined as follows: (i) functional, meaning that for a given individual, there can be at most one individual in the range associate with it; (ii) inverse functional, which is the reverse of the functional characteristic. This means that for a given individual in the range, there can be at most one individual associated with it; and (iii) transitive, which indicates that if an individual "A" is related to another individual "B" and the individual "B" is related to the individual "C", then there is an implied relationship between individual "A" and individual "C". The object property "hasHoldingPeriod" is an example of a functional link between the classes "CFD" and "Holding_Period", because a "CFD" should only have one "Holding_Period". Conversely, the object property "hasManufacturerSupervisor" is an example of an inverse functional link between the classes "Manufacturer" and "Manufacturer_Supervisor", given that the "Manufacturer" should only have one "Manufacturer_Supervisor". Finally, the object property "hasMarkets" is an example of a transitive connection between the classes "CFD" and "Markets" because if a "CFD" is related to one market and that market is related to another market a link exists between the CFD and the second market.

Regarding the terms "domain" and "ranges" they can be defined as follows: (i) the "domains" specify the class to which a certain object property applies, (ii) while "ranges" specify the class to which the object property points to.

The data property provides a relation to append an entity instance to a datatype value that is a measure of what that data property is about. By default, the data property root domain is called "owl:topDataProperty" and the KID Ontology is comprised of 22 data properties.

A challenge that we have experienced during this phase was deciding whether to create a specific class to a certain concept or to set it as a property. For example, the Synthetic Risk Indicator (SRI) for a certain PRIIP has two different dimensions that's given by its credit risk and its market risk. Initially, we have created a class named "SRI" in order to represent this concept. However, we had also created a class named "Risk factors" that had 17 subclasses of risks associated to the PRIIPs negotiation. The credit risk and market risk were both represented as a subclass of the class "Risk factors". Because of this redundancy and considering the scope of KID Ontology, we made the decision to remove the class "SRI" and represent it as a data property related to the class "Risk factors".

Another example of the above-mentioned challenge was that initially, we represented the "Key Information Document" as a class of the KID Ontology. However, during the formalization process, we realized that a class named "Key Information Document" didn't provide any value to the ontology, because it was the object from reality whose concepts were being formally represented and not a concept itself to be represented.

An additional challenge experienced during the KID Ontology was related to the ambiguity between the terms. Ambiguity is an inherent part of the human language and happens when people have different meanings for the same term. Terms such as "Costs", "Information Sites", "Applicable Law, "Financial Instruments" and/or "Product Typology" may differ depending on who is interpreting the term. In the specific case of the KID Ontology, this issue was mitigated by adopting the definitions provided by supervisors and regulators that constitute a standard and stable understanding of a certain technical term.

During the formalization process the number of classes and relations between them represented was reduced, since the inference obtained from the execution of the reasoners allowed the observation of redundancies and inconsistencies between the originally though classes and subclasses of the KID Ontology. The solution of this issue paved the way for the simplification of the KID ontology that was originally drafted. For example, if the class "Manufacturer" is related to the class "CFD" and the class "CFD" is related to the class "Investor", the inference system can derive that the class "Manufacturer" is also related to the class "Investor" without the need to explicitly represent that fact.

The inference engine on ontologies expressed in OWL and edited on Protégé is performed by an automatic reasoner built-in the editor. With the help of the reasoners, it is possible to infer certain facts using axioms and inference rules based on a previous asserted set of facts. Put differently, the reasoner makes explicit the assertions that are only implicit in the formally represented concepts. Therefore, the inferred information is a corollary of the explicitly stated information.

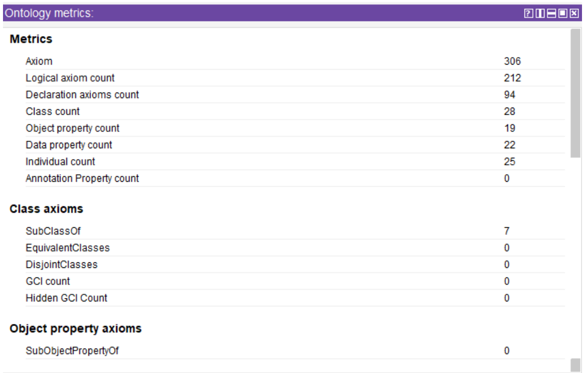
Finally, for the inference engine correctly derive facts from the explicitly represented concepts, it was necessary to use the wide range of object properties available on the Protégé editor, mainly the functional, inverse functional, transitive, symmetric and asymmetric.

A. Implementation

To build the KID Ontology, we began by identifying the relevant concepts contained in the template given by the legislators. By filtering the key knowledge that should be integrated in the KID Ontology we have minimized the risk of redundancies and/or irrelevant information being considered. The relevant information was then compiled in a document where was represented the taxonomy, its structure and its relevant properties.

After the process of identifying and extracting the relevant concepts from the above-mentioned template and samples, we used the Protégé editor to develop the KID Ontology and represented in OWL Language, which provides a formal manner to describe the domain concepts.

The KID Ontology has 28 classes, 22 data properties and 19 object properties, as shown in Fig. 3.



Ontology metrics:

Metrics	
Axiom	306
Logical axiom count	212
Declaration axioms count	94
Class count	28
Object property count	19
Data property count	22
Individual count	25
Annotation Property count	0
Class axioms	
SubClassOf	7
EquivalentClasses	0
DisjointClasses	0
GCI count	0
Hidden GCI Count	0
Object property axioms	
SubObjectPropertyOf	0

FIG. 3 - KID ONTOLOGY METRICS AS AT 1 OCTOBER 2023

A. Evaluation

The evaluation phase is characterized by a technical perspective to assess the quality of the produced ontology. For this phase we used the Pellet inference system and SPARQL queries built-in Protégé editor for deduction purposes and exploration of the stability of the ontology knowledge and its overall consistency.

However, due to lack of equivalent ontologies in this domain, it was not possible to compare the results with other ontologies form the same domain.

II. FUTURE WORK

This paper aims to showcase the process of building an ontology applied to the capital markets, specifically regarding the key information documents developed by issuers of complex investment products such as CFDs. The paper also highlights the challenges faced in the ontology development process and the impact of the inference system on its development.

The inferred information allowed for the simplification of the ontology comparatively of the assertions made during its initial design phase. The initial designed ontology was large and complex and later we found to have redundancies and inconsistencies. In this regard, the reasoning process made it possible to identify and report errors in the ontology that otherwise would have passed unnoticed and update the KID Ontology accordingly. The reasoning process helped to alert for unpredicted interactions between the asserted classes. The ability to use the reasoners allowed us to focus on the knowledge acquisition and class description process and then test it for errors and inconsistencies identified by the inference system and then update it appropriately.

The KID Ontology aims to create a formal knowledge representation of the KID required by the PRIIPs regulatory framework. The KID Ontology is, however, limited to the CFDs. Despite the need to limit the scope of the paper to this specific financial instrument, CFDs are the most common derivative financial instrument traded by investors.

The KID Ontology can also be used by artificial intelligence systems, including NPL, machine learning models and for knowledge sharing within this domain. As previously mentioned, capital markets regulators across various jurisdictions, namely Europe, Asia, and the United States of America, are increasing their efforts to keep pace with financial innovation that was prompted mainly by the financial crisis of 2008. In this context, they have been testing NPL techniques to improve their legal compliance systems, namely by exploring the possibility to perform analysis on legal documents such as KID and increase the automation of their processes. To achieve this, ontologies like the KID Ontology can help optimize those processes.

Acknowledgment

This project would not have been possible without the guidance and orientation from Professor Ana Maria de Almeida from Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação em Ciências da Informação, Tecnologias e Arquitetura, Lisboa, Portugal AND CISUC-Centro de Informática e Sistemas da Universidade de Coimbra, Coimbra, Portugal and Professor Francisco Guimarães from Instituto Universitário de Lisboa (ISCTE-IUL).

References

[1] Perchet, Romain and Herambourg, Nicolas and Carvalho, Raul Leote de, PRIIPs Regulation Unwrapped: Essential Aspects and Practical Implications, p.4.

[2] Berners-Lee, Tim & Hendler, James & Lassila, Ora. (2001). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. ScientificAmerican.com.

[3] Guarino, Nicola. (1998). Formal Ontologies and Information Systems.

[4] Javed, M.A., Muram, F.U., Kanwal, S. (2022). Ontology-Based Natural Language Processing for Process Compliance Management. In: Ali, R., Kaindl, H., Maciaszek, L.A. (eds) Evaluation of Novel Approaches to Software Engineering. ENASE 2021. Communications in Computer and Information Science, vol 1556. Springer, Cham. https://doi.org/10.1007/978-3-030-96648-5_14

[5] Financial Stability Board (2020), The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions.

[6] Armstrong, P., and Harris., A (2019), RegTech and SupTech – change for markets and authorities, ESMA Report on Trends, Risks and Vulnerabilities, no. 1.

[7] Pinto, H., Martins, J. Ontologies: How can They be Built?. Know. Inf. Sys. 6, 441–464 (2004).

[8] Chhim, P., Chinnam, R.B. & Sadawi, N. Product design and manufacturing process-based ontology for manufacturing knowledge reuse. J Intell Manuf 30, 905–916 (2019). https://doi.org/10.1007/s10845-016-1290-2

[9] Gruber, T.R. (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, 199-200.

[10] Borst, W. N. (1997). Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Enschede: Centre for Telematics and Information Technology (CTIT).

[11] Guarino, N. and Giaretta, P. (1995) Ontologies and Knowledge Bases. In: Towards Very Large Knowledge Bases, IOS Press, Amsterdam, 1-2.

[12] Fikes, Richard and Adam Farquhar. “Large-Scale Repositories of Highly Expressive Reusable Knowledge.” (1999).

[13] Gruber, T.R. (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, 199-200.
[14] Uschold, Michael & Grüninger, Michael. (1996). Ontologies: Principles, methods, and applications. The Knowledge Engineering Review. 11.
[15] Corcho, O., Fernández-López, M., Gómez-Pérez, A., López-Cima, A. (2005). Building Legal Ontologies with METHONTOLOGY and WebODE. In: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds) Law and the Semantic Web. Lecture Notes in Computer Science, vol 3369. Springer.
[16] G G Petrova et. al 2017 J. Phys.: Conf. Ser. 803 012116.
[17] Graf, S., Kling, A. PRIIP-KID: appearances are deceiving or why to expect the unexpected in a generic KID for multiple option products. Eur. Actuar. J. 10, 527–555 (2020). https://doi.org/10.1007/s13385-020-00243-0.