INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Application of Semantic Web Techniques in Data Warehouse and Business Intelligence Systems

António Lorvão Ferreira Antunes

PhD in Information Science and Technology

Supervisors:

Doctor Elsa Alexandra Cabral da Rocha Cardoso,
Assistant Professor, ISCTE - Instituto Universitário de Lisboa

Doctor José Eduardo de Mendonça Tomás Barateiro,
Assistant Professor, Universidade do Algarve

April, 2024

Department of Information Science and Technology

Application of Semantic Web Techniques in Data Warehouse and Business Intelligence Systems

António Lorvão Ferreira Antunes

PhD in Information Science and Technology

Jury:

Doctor João Marques Guerreiro, Associate Professor, ISCTE - IUL

Doctor Orlando Belo, Associate Professor with Habilitation, University of Minho

Doctor António Palma dos Reis, Full Professor, ISEG - University of Lisbon

Doctor Luís Nunes, Associate Professor, ISCTE - IUL

Doctor José Barateiro, Assistant Professor, University of Algarve

April, 2024

Obrigado Mami, sem ti não seria...

Acknowledgment

This research would not have been possible without the support and encouragement of the people who have surrounded me over the last few years. Their invaluable contributions have been fundamental to the completion of this PhD. I would like to express my sincere gratitude to each and every one of them.

First of all, I would like to express my gratitude to my supervisors, Professor Elsa Cardoso and Professor José Barateiro, for taking on the challenge of supervising and guiding me throughout my doctoral thesis. Thank you for believing in my capabilities and for always encouraging me to do more and better, for all the coffees and Port wine!

To LNEC - National Laboratory for Civil Engineering, the institution that has welcomed me since 2018, for all the institutional support, all the resources and the research environment that have been provided. I would like to express my thanks to all the staff and collaborators at LNEC, with a special mention to Eng. Juan Mata, Sofia Cerqueira and João Rico, and the Information Technology in Civil Engineering Unit.

To ISCTE - IUL, which has been my second home since 2012, and which has seen me grow as a person, teacher and researcher. I would like to thank all the professors I have met during my academic career, with special thanks to Professor Luís Nunes, a member of my monitoring committee, whose feedback over the last four years has been essential, and to Professor Ricardo Ribeiro.

I would like to thank the FCT - Foundation for Science and Technology for the financial support that enabled me to carry out this work (Grant 2021.07134.BD).

On a personal note, heartfelt thanks to my family, including my aunts and uncles, especially Aunt Cati and Uncle Luís João, who, among other things, gave me the opportunity to be a student in Lisbon, my cousins, near and far, and to my ever-youthful grandmother, Nini, my number one fan on all social networks. Special thanks to my third supervisor, Uncle Rui Grilo, for his unwavering support in and beyond studies.

To all my friends who have always looked out for me, who have diverted me from the path (often at my request) when necessary and especially when not. From those I've known for over 25 years to those I've met more recently, to the "Bandoleros" from Leiria, the "Velhos" who have always been, and all members of "Tatu às 6" — thank you. A special nod to my housemates, Lourenço Maltez, Catarina Rodrigues, and Bernardo Ribeiro.

Lastly, to my mother, father, and brother — for everything.

Resumo

Devido à sua formalização semântica e qualidades de inferência, as técnicas da Web Semântica (SW), como as ontologias, são utilizadas nos sistemas de informação para fazer face à necessidade crescente de partilha e reutilização de dados e conhecimento em várias áreas de investigação. A integração de artefactos baseados no conhecimento em sistemas de apoio à decisão (DSS), como os sistemas de Data Warehouse e Business Intelligence (DW/BI), pode fornecer novas fontes de informação, permitir novas capacidades analíticas e melhorar o processo de tomada de decisão.

Esta tese de doutoramento segue uma Metodologia de Investigação Design Science, com uma solução centrada em objectivos, compreendendo quatro iterações de desenho e desenvolvimento e propondo seis artefactos. A principal contribuição é o desenho e desenvolvimento do artefato Integration Framework, que permite a representação, a exploração e a validação de informação específica de domínio, como a estratégia, em ambientes de BI. A representação e exploração de conhecimento específico do domínio, incluindo os seus conceitos e relações, é assegurada através da conceção e desenvolvimento de duas ontologias. A primeira ontologia é dedicada ao domínio das estruturas de engenharia civil, enquanto a segunda se foca nos elementos da estratégia e nas suas relações, seguindo uma abordagem de Balanced Scorecard. A demonstração e a avaliação dos artefactos da investigação são realizadas em dois estudos de casos reais: um centrado na integração entre Building Information Modeling (BIM) e os Sistemas de Gestão de Activos (AMS), e outro na análise da estratégia possibilitada pelo DW/BI.

A Integration Framework permite que o DSS seja enriquecido com dados de fontes externas. Os dados de outras fontes, tais como sistemas operacionais ou documentos (por exemplo, relatórios estratégicos), são formalizados e integrados semanticamente com o DSS, fornecendo conhecimento adicional que pode ser utilizado por qualquer aplicação de BI para melhorar o processo de tomada de decisão.

Ao nível estratégico, a integração de técnicas de SW em sistemas DW/BI reforça a análise de desempenho, fornecendo contexto estratégico ao ambiente de BI e permitindo o apuramento automático de valores de indicadores de desempenho. Estas melhorias aumentam a eficiência, a fiabilidade e a conveniência (no momento certo) da tomada de decisão, proporcionando aos gestores um ambiente analítico, baseado em dados, para a tomada de decisão alinhada com a estratégia organizacional.

Palavras-chave: Data Warehouse; Business Intelligence; Sistemas de Apoio à Decisão; Web Semântica; Ontologias; Building Information Modeling (BIM); Gestão de Activos; Balanced Scorecard; Gestão da Estratégia; Integração Semântica.

Abstract

Due to their semantic formalization and inference qualities, Semantic Web (SW) techniques, such as ontologies, are used in Information Systems to cope with the growing need for sharing and reusing of data and knowledge in various research areas. Integrating knowledge-based artifacts into Decision Support Systems (DSS), such as Data Warehouse and Business Intelligence (DW/BI) systems, can provide new sources of information, enable new analytical capabilities, and enrich the decision-making process.

This doctoral thesis follows a Design Science Research Methodology, with an objective-centered solution, comprising four design and development iterations and proposing six artifacts. The main contribution is the design and development of the Integration Framework artifact, enabling the representation, exploration, and validation of domain-specific information, such as strategy, in BI environments. The representation and exploration of domain-specific knowledge, including its concepts and relationships, is ensured through the design and development of two ontologies. The first ontology is dedicated to the domain of civil engineering structures, while the second focuses on the elements of strategy and their relationships, following a Balanced Scorecard approach. The demonstration and evaluation of research artifacts are conducted in two real-world case studies: one focusing on the integration between Building Information Modeling (BIM) and Asset Management Systems (AMS), and another on DW/BI enabled strategy analysis.

The Integration Framework enables the DSS to be enriched with data from external sources. Data from other sources, such as operational systems or documents (e.g., strategy reports), is formalized and semantically integrated with the DSS, providing additional knowledge that can be used by any BI application to enhance the decision-making process.

At the strategic level, the integration of SW techniques in DW/BI systems enriches performance analysis by providing strategic context to the BI environment and enabling the automatic retrieval of performance indicator values. These improvements increase the efficiency, reliability and timeliness (at the right time) of decision-making, providing managers with an analytical and data-driven environment for their decision-making aligned with the organizational strategy.

Keywords: Data Warehouse; Business Intelligence; Decision Support Systems; Semantic Web; Ontologies; Building Information Modeling (BIM); Asset Management; Balanced Scorecard; Strategy Management; Semantic Integration.

Contents

List of Abbreviations

AI	Artificial Intelligence
AMS	Asset Management System
API	Application Programming Interface
BEAM	Business Event Analysis & Modelling
BIM	Building Information Modelling
BI	Business Intelligence
BSC	Balanced Scorecard
BSO	Balanced Scorecard Ontology
CoDEC	Connected Data for Effective Collaboration
CEDR	Conference of European Directors of Roads
DSRM	Design Science Research Methodology
DSS	Decision Support Systems
DW	Data Warehouse
ETL	Extract, Transform, and Load
EUROTL	European Road Object Type Library
IFC	Industry Foundation Classes
IS	Information Systems
JA	Journal Article
KPI	Key Performance Indicator
LDWOWL	Light Data Warehouse Ontology
LNEC	National Laboratory for Civil Engineering
NRA	National Road Authorities
O&M	Operational and Management
OOPS!	OntOlogy Pitfall Scanner!
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SHACL	Shapes Constraint Language
SPARQL	SPARQL Protocol and RDF Query Language

SW	Semantic Web
SWRL	Semantic Web Rule Language
W3C	World Wide Web Consortium

CHAPTER 1

Introduction

This chapter lays the foundation for this doctoral thesis by outlining its core motivation and research goals. It introduces the methodological approach employed throughout the research and details the key contributions made, including the research artifacts developed and the list of publications.

Fundamental background concepts related to the main research fields are introduced, namely concerning Data Warehouse/Business Intelligence (DW/BI) systems and the Semantic Web (SW). To demonstrate and assess the effectiveness and applicability of the contributions of this research, two real-world case studies are employed. One case study pertains to civil engineering structures, while the other is related to the field of organizational strategy. These case studies are described in this chapter and will be used to demonstrate the practical implications of the research contributions. Both the background concepts and the case studies' information are further explored in depth in the publications originated from this doctoral thesis.

This thesis adopts an article structure. This chapter elaborates on the key research communications and their relevance to the research contributions, providing a guide for readers. It also presents a summary of tools and software utilized in this research. Lastly, it outlines the structure of the remaining chapters of this document.

1.1. Motivation and Goals

Business Intelligence (BI) is a term introduced in the mid-1990s and has since become a fundamental component in many enterprises [11]. It is an "umbrella" term that refers to applications, infrastructures, tools, and methodologies aimed at enhancing decision-making processes and performance by leveraging access and analysis of data and information. DW/BI systems represent data-driven Decision Support Systems (DSS) [55], offering integrated repositories (DW) to provide analytical and decision capabilities to business users [31].

Despite their proficiency in managing and analyzing structured transactional data, these systems face challenges in handling the expanding array of unstructured data [54]. Moreover, the reliance on SQL-based data access provided by DW/BI systems is proving inadequate for accommodating the diverse data types and latest algorithms utilized in Artificial Intelligence (AI) and Data Science analyses [25]. The need to extract information and gather insights from various sources is ever-increasing in the era of Big Data, where data is generated in high volume, velocity, and variety of formats and structures [19].

Nonetheless, while recent literature highlights the significance of unstructured data exploration, business activities typically generate structured data related to their business processes and transactions. Structured data analysis is crucial and has substantial business value, with the importance and efficacy of DW/BI systems in its analysis remaining undeniable [55]. Key Performance Indicators (KPIs) are readily available as structured data, such as sales figures and product quantities. Additionally, structured historical data plays a critical role in the development of descriptive, predictive, and prescriptive analyses.

DW/BI systems are designed and used to support the analytical requirements across various departments or business sectors within an organisation, ensuring a unified understanding of data and providing a 'single version of the truth'. For this reason, it is essential to establish common vocabularies or terminologies that allow business users to communicate with each other and with the development team and ensure alignment with other organizational systems [32]. There has been a growing emphasis on Open[1] and FAIR[2] data principles in Information Systems (IS) research, with a central focus on interoperability and data sharing. These principles are increasingly being integrated into diverse research domains, facilitating the circulation and accessibility of data and information, as evidenced by initiatives like the European Open Data Portals[3]. Knowledge representation formalisms, such as ontologies, are being developed to facilitate researchers' access to data, information, and knowledge related to their respective fields of study. Ontologies are used in the Semantic Web (also referred to as World Wide Web 3.0) and other fields of study to encode data, enabling the sharing and reuse of knowledge, and, more importantly, ensure its machine readability and processability [24, 51]. Various research and application fields, such as biology and computer science, have embarked on initiatives to enhance knowledge discovery and utilization [51]. The shared semantics provided by these knowledge representations are fundamental in avoiding misunderstandings or errors, particularly in scenarios where natural language plays a key role, such as during requirement gathering, data source analysis (context and meaning of each entity), or data analysis and exploration phases.

The primary goal of this work is to explore the use and integration of SW techniques, such as ontologies, with DSS, namely DW/BI systems, to enhance the decision-making process. Due to their semantic, formalization, and inference qualities, the integration of ontologies into DSS can provide new sources of information and enable new analytical possibilities or facilitate the decision-making process for the business user. The main contribution of this research is the development of a framework that enables the support, navigation, and validation of domain-specific information, such as strategy, in BI environments. The representation and exploration of each domain, including its concepts and relationships, is ensured through the design and development of two distinct ontologies.

[1]Open data handbook - http://opendatahandbook.org/
[2]Go fair initiative - https://www.go-fair.org/
[3]https://data.europa.eu/

The first ontology is dedicated to the domain of civil engineering structures, while the second focuses on the elements of strategy and their interrelationships, with an emphasis on the Balanced Scorecard. The validation and demonstration of these research artifacts are conducted by means of real-world case studies, allowing a careful assessment of their effectiveness and applicability.

1.2. Background

This section introduces core background concepts essential for this research, covering DSS, particularly DW/BI systems, alongside key aspects of the SW and ontologies.

1.2.1. Decision Support Systems

DSS are interactive computer-based systems designed to aid business users in identifying and resolving problems, as well as facilitating the decision-making process [47]. DSS should provide managers and business users with quick and interactive information assistance, ensuring that they receive the "right information at the right time, in the right format." [62]. The Association for Information Systems Special Interest Group on Decision Support Systems adopts a classification of DSS proposed by Power [47], which classifies DSS according to the type of components they use [55]: (a) **Communication-driven or Group DSS**: DSS that feature communication, collaboration and sharing (through technology) as their decision-making support; (b) **Data-driven**: DSS focusing on the access, analysis and manipulation of data. DW/BI systems and business process management systems are some examples of data-driven DSS; (c) **Document-driven**: DSS that emphasize the use (or retrieval), storage, management and analysis of documents; (d) **Knowledge-driven**: DSS that use knowledge bases and artificial intelligence (e.g., Expert systems, Data Mining); (e) **Model-driven**: DSS that focus on the use of quantitative models; (f) **Compound DSS**: Hybrid DSS that combine two or more of the previous components.

1.2.2. Data Warehouse/Business Intelligence Systems

DW/BI systems are data-driven DSS comprised of two primary subsystems, as shown in Figure 1.1: data warehousing, focused on "getting data in," and business intelligence, focused on "getting data out" [64]. The objective of data warehousing is to extract, transform, and load data from diverse source systems into an integrated repository known as the DW. The inherent diversity of data across these source systems poses integration challenges, such as varying formats or representations of the same entities, which are addressed through the Extract, Transform, Load (ETL) process. BI, on the other hand, leverages the DW to retrieve data and provide data-driven decision support to business users. This support can take various forms, including data analysis and exploration through reporting tools and dashboards, as well as the utilization of data mining models to extract predictions and insights from analytical data. Through these mechanisms, BI enables users to access, analyze, and derive actionable insights from the data stored within the DW, ultimately facilitating informed decision-making processes in organizations.

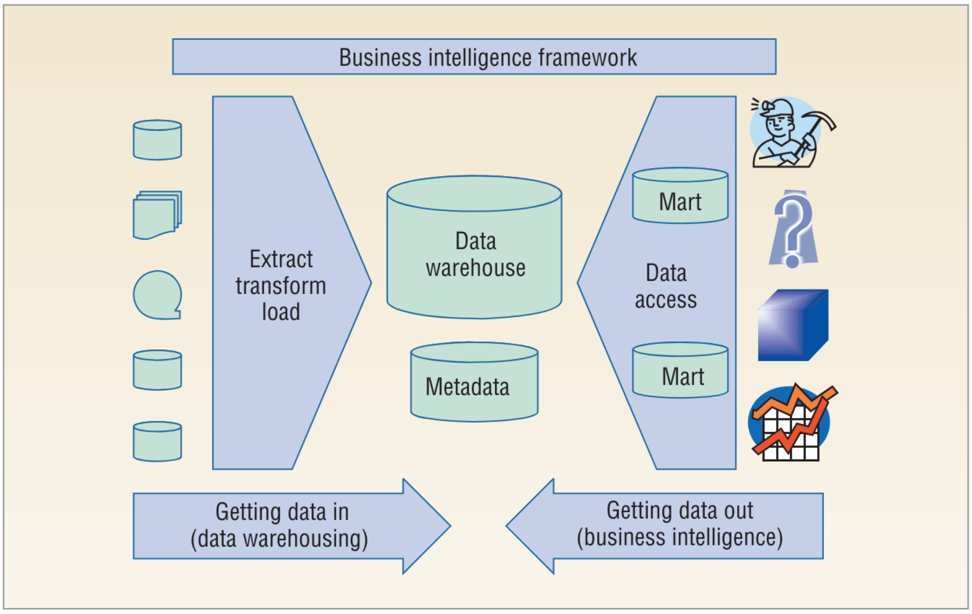


Figure 1.1. DW/BI System. Retrieved from Watson and Wixom [64].

Dimensional modeling offers an intuitive and high-performance approach for aggregating, retrieving, and analyzing historical data in DW/BI systems [31]. Unlike traditional operational systems, that adhere to normalization rules, analytical systems follow dimensional modeling, where data is typically stored in either star schemes or cubes, also known as multidimensional databases [3, 31]. Central to dimensional modeling is the concept of distinguishing between facts and dimensions. Facts typically represent numeric and additive measurements of key processes (e.g., sales quantity, sales dollar amount), while dimensions provide context to these facts by representing various business entities (e.g., Client, Date, Vendor). Dimensions are utilized to filter or aggregate facts, enabling users to explore and analyze data from different perspectives. Hierarchies within dimensions describe the possible aggregation paths, utilizing parent-child relationships between attributes to drill up (remove detail) or drill down (add detail). This hierarchical structure facilitates the exploration and analysis of data at various levels of granularity, allowing users to navigate through different levels of detail to gain insights into specific contexts. For example, information on monthly sales can be drilled down to daily sales or aggregated (drilled up) to yearly sales, providing a comprehensive view of the data.

1.2.3. Semantic Web and Ontologies

The term "ontology" refers to a branch of philosophy that studies the nature and structure of things/objects, properties, events, and relations [56]. However, in Information Science, "ontology" denotes a computational artifact that encodes knowledge about a specific domain [57]. The most widely accepted definition in computer science for the term "ontology" was proposed by Studer et al. [58, p.25], which defines ontology as "a formal, explicit specification of a shared conceptualization". A conceptualization is "an

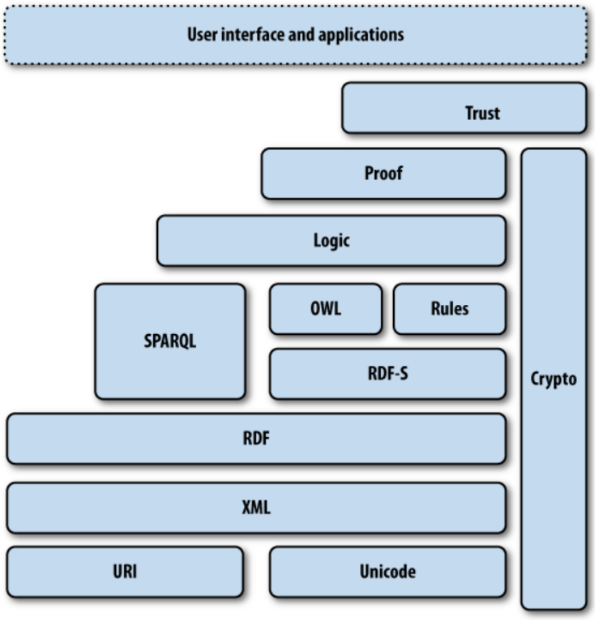


FIGURE 1.2. W3C Semantic Web Technology Stack. Retrieved from Hebeler et al. [23].

abstract, simplified view of the world that we wish to represent" [16, p. 1], i.e., an abstract model with the relevant concepts of something. An explicit specification entails that concepts, relationships, and constraints are clearly defined and encoded within the ontology. Furthermore, the formalization of an ontology ensures its machine-readability, facilitating computational processing. Ultimately, an ontology should reflect a community's agreed-upon conceptualization of a domain, emphasizing the importance of a shared understanding within the community [58].

The purpose of ontologies is to facilitate the sharing, reuse, and analysis of knowledge, ultimately promoting interoperability and heterogeneity [42]. According to the World Wide Web Consortium (W3C)[4], these qualities make ontologies an indispensable component of the SW. The W3C defines the Resource Description Framework (RDF), RDF Schema (RDFS), SPARQL and the Ontology Web Language (OWL) as standards for the Semantic Web (see Figure 1.2). RDF is the recommendation for the "creation, exchange and use of annotations on the Web" [17, p.72]. The resources are described in the form of triples (*subject property object*) [44], for example, "Professor" "rdfs:subClassOf" "Faculty Staff". SPARQL is a query language for accessing and manipulating data stored in RDF format, commonly used for querying SW data and knowledge graphs. The RDFS provides a vocabulary for RDF introducing the concepts of classes and hierarchies, together with the necessary inference rules. OWL enhances expressiveness by incorporating elements such as disjointness and cardinality, and defines properties as either object (relationships between classes) or data (attributes) properties. There are three OWL sublanguages/types: Lite, DL and Full, with different levels of expressiveness. Normally, the choice of an OWL sublanguage depends on the specific problem domain and modeling

[4]https://www.w3.org/

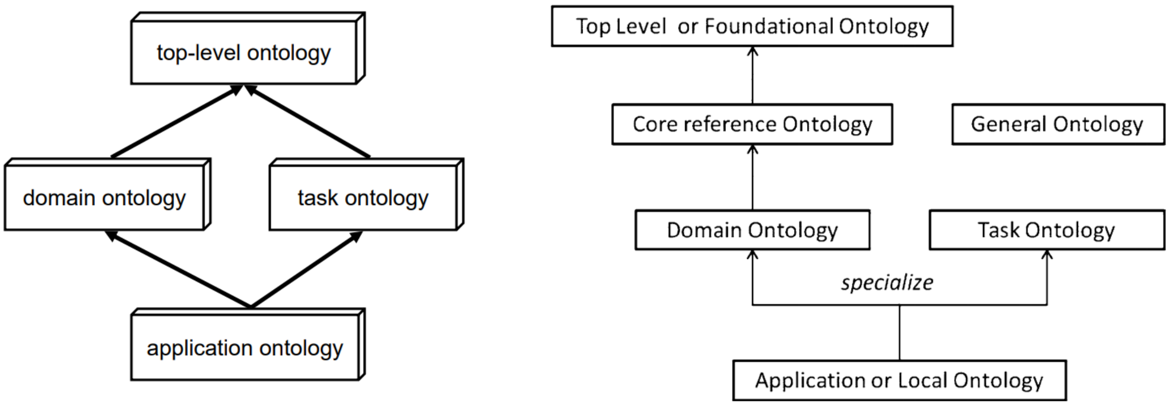


FIGURE 1.3. Ontology types hierarchy based on Scope. Retrieved from Stephan et al. [57] (left) and Roussey et al. [53] (right).

requirements, with a trade-off between expressiveness and inference capabilities (reasoning), i.e., the more expressive a language is, the less inference it is capable of [60, 59]. When populated with individual instances, an ontology is called a Knowledge Base [17].

Stephan et al. [57] and Roussey et al. [53] present ontology classifications based on their scope, illustrated in Figure 1.3. Top-level ontologies, characterized by their generality, encompass abstract concepts applicable across diverse domains and applications. They encapsulate fundamental notions such as objects, events, and processes, serving as foundational pillars for other ontologies. Domain and Task ontologies specialize in knowledge pertinent to specific domains or tasks. The conceptualization of a domain should be independent of tasks (e.g., a biology ontology should be separated from a diagnostic task ontology). Application or Local ontologies, with the narrowest scope, cater to specific tasks within distinct domains, leveraging both domain and task ontologies to achieve their objectives. Additionally, Roussey et al. [53] introduces two supplementary ontology types: the Core Reference ontology, which allows different communities to have different domain ontologies aligned and integrated with a standard, core, reference ontology; and the General ontology, which is not dedicated to a specific domain or field.

1.3. Case Studies

Two real-world case studies are used throughout this research. These case studies are used to demonstrate and evaluate the efficacy and applicability of the developed artifacts in practical settings. The different domains and decision-making applications are key to exploring the practical implications of the research artifacts, offering insights into their utility and potential impact within BI environments.

The first case study demonstrates the applicability of the research contributions in a domain-specific scenario of civil engineering, where users access specialized (domain-specific) applications to explore information and support decision-making at an operational level. This case study, detailed in Section 1.3.1, is developed in the context of European highways industry international research.

The second use case illustrates the versatility and adaptability of the research contributions through the integration of SW techniques with a typical DW/BI system. The resulting integration framework aims to enable the support, navigation, and validation of domain-specific information within BI environments while leveraging the DSS data management component (DW). This case study, presented in Section 1.3.2, aims to support the strategic analysis of a public organization.

1.3.1. Connected Data for Effective Collaboration Project

Physical infrastructures, such as buildings, bridges and roads, can be modeled and managed across the whole asset lifecycle using Building Information Modelling (BIM). The 3D visualization provided by BIM tools allows stakeholders to collaborate, share, and exchange information for decision support during the whole asset lifecycle. However, the use of BIM in transport infrastructures is still far from its wide application in the building industry [13]. Furthermore, despite its potential application in all phases of the infrastructure life cycle, BIM use during the operational and management (O&M) phase in transport infrastructures is currently limited [65].

Across Europe, National Road Authorities (NRAs) have implemented Asset Management Systems (AMS) to oversee the maintenance, management, and structural safety during the operational phase of engineering structures in European highways. These systems store operational asset data and information in various formats, including sensor readings and inspection results. Ideally, information should be shared between BIM models and AMS so that more efficient and informed decisions can be taken during the operational phase of these engineering structures. However, existing BIM data standards primarily focus on the design and construction phases, lacking emphasis on integration with AMS for operational phases. Due to the increasing number of solutions for asset monitoring (sensor technology and Internet-of-Things), the interoperability between BIM and AMS systems becomes increasingly crucial for timely decision-making in an integrated environment. Linking 3D model data with asset management data allows access to an integrated view of information, reduces errors, and saves time and costs, while also enhancing compliance, safety, and risk mitigation during the operational phase [65].

Although efforts such as the AM4INFRA standardization initiative [38, 33] have been made in Europe, data management practices remain largely tailored to the individual AMS within each NRA. While some countries have developed object-type libraries, there is a noticeable gap in the availability, extent, and content of data dictionaries for highway assets, which hinders the effective use of data, especially within a BIM environment [7, 8].

The Connected Data for Effective Collaboration (CoDEC) research project[5], was funded by the Conference of European Directors of Roads (CEDR). The project aimed to implement BIM principles in the European Highways Industry, focusing on data exchange between BIM and AMS to manage asset data during the operational phase. A "Master Data Dictionary" developed within the project served as a foundational structure

[5]CEDR Call 2018:BIM, from October 2019 to September 2021. https://www.codec-project.eu/

for integrating diverse information management systems, incorporating both legacy AMS data and sensor/scanner data. This shared conceptualization was key to provide standard data formats that can be used between Europe's NRA and their systems. By adopting semantic web and linked data principles, CoDEC aimed to link operational data with BIM environments, facilitating decision-making by providing standardized data formats across Europe's NRAs and their systems. As part of this research[6], an ontology was designed to model and represent structures, structural elements, and operational data, such as sensor information or legacy data, allowing for the development of a single format for information exchange [22] and enhances decision-making during the operational phase of these assets.

1.3.2. DW/BI Strategic Analysis in a Public Organization (LNEC)

Strategic management is a process undertaken by public organizations or other entities to formulate, execute, and evaluate strategies aimed at attaining long-term objectives and goals [9]. The effectiveness of strategic management significantly impacts organizational performance in both public and private sectors, with the formalization of strategic processes shown to enhance performance outcomes [14]. However, most public organizations typically use strategic management systems with low comprehensiveness or formality and are usually decentralized [10]. Moreover, Manes-Rossi et al. [37] only found that 8% of the works in their literature review explore non-financial reporting from a strategy management perspective in the public sector. Král [34] identifies research directions in this field, including continued performance evaluation, use of official quantitative data, and clear and understandable (to policymakers, managers, and stakeholders) performance management systems.

Kalampokis et al. [27] presents SW technologies as one of the emerging technologies in the public sector, highlighting the advantages of their usage, specifically the shared semantics, and interoperability. Ontologies can be used to create a common semantic data model, which can be useful to define unified report methods (beneficial for both reporters and readers) [37, 21], integrate or transfer data between public organizations [27], and automate processes using the formalization of knowledge from heterogeneous sources [26]. SW technologies have also been used to foster openness and transparency in the public sector [40].

Since its introduction in 1992 [29], the Balanced Scorecard (BSC) remains the most well-known approach for performance assessment, due to the balance between non-financial and financial indicators across various perspectives [34, 61]. A BSC approach was used to define the Portuguese National Laboratory for Civil Engineering (LNEC)'s strategy for 2021-2027, including the definition of strategic objectives and indicators used to monitor its execution [6]. LNEC was established in 1946 to provide specialized services in civil engineering. As a public laboratory, it has been involved in national projects (e.g.,

[6] Part of this doctoral thesis was developed in the context of the LNEC's participation in the CoDEC project consortium.

dams, communication routes, river and sea hydraulics, large structures) and international collaborations, performing scientific and technical works in almost fifty (50) countries. Over the years, LNEC expanded its competencies, becoming a hub for research, experimentation, postgraduate education, and community/local services. As a public institute, LNEC has the legal responsibility to report on its activities and performance, including the evaluation of indicators that ensure alignment with strategic objectives.

By formalizing the BSC framework, ontologies can assist in the validation of the formulated strategy, evaluation of performance indicators, and validation of cause-and-effect relationships between strategic objectives. Furthermore, ontologies can provide a semantic layer to facilitate the integration, alignment, and traceability of strategic models with organizational information systems, bridging the gap between strategy management and data related to the BSC framework, and enhancing the capacity of organizations to make informed data-driven decisions, efficiently allocate resources, and effectively navigate the intricate landscape of the public sector. This is a crucial contribution, given the growing importance of leveraging data in strategic decision-making processes in an evolving business environment [15].

1.4. Research Methodology

This research follows the Design Science Research Methodology (DSRM), proposed by Peffers et al. [45]. The methodology was instantiated as shown in Figure 1.4 with four iterations. This section outlines the contributions of the research within each iteration and phase of the DSRM.

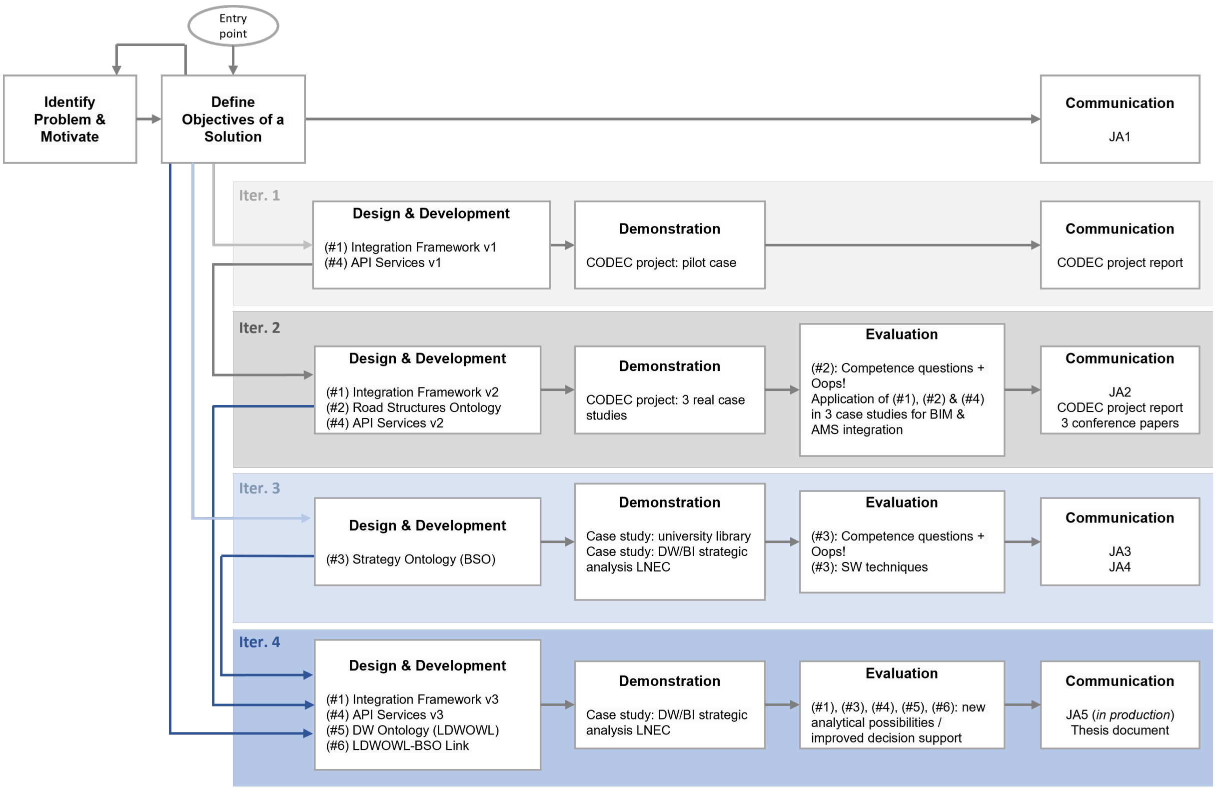


Figure 1.4. DSRM Process Instantiation.

The analysis of the current use and impact of knowledge representation techniques in DW/BI systems can be seen as the entry point of this DSRM research, corresponding to an objective-centered solution. Due to their semantic, formalization, and inference qualities, ontologies are used in IS to cope with the growing need for sharing and reusing data and knowledge in various research areas. The integration of these knowledge-based artifacts into DSS can provide new sources of information, enable new analytical possibilities, or facilitate the decision-making process for the business user.

The first iteration of DSRM (Iter. 1) focuses on the design and development of an integration solution, named Integration Framework (version 1 of artifact #1), providing a proof-of-concept to study the feasibility of using ontological knowledge within a DSS environment. A first set of API Services (version 1 of artifact #4) was also developed. This proof-of-concept is demonstrated with simulated data from a pilot project in the CoDEC project context (see Section 1.3.1). The second DSRM iteration (Iter. 2) enhances the design and development of the artifacts Integration Framework (version 2 of artifact #1) and API services (version 2 of artifact #4). Additionally, this iteration also creates the artifact #2, a domain-specific ontology named Road Structures Ontology. This iteration was performed specifically aiming at supporting the BIM-AMS integration applied in a real-world case study in civil engineering (European highways).

The third research iteration (Iter. 3) comprises the design and development of the artifact #3, named Strategy Ontology, with the objective of representing and exploring the domain-specific concepts related to organizational strategy analysis. Artifact #3 is demonstrated and evaluated, as a proof-of-concept, in a case study from a public university library. Further demonstration and evaluation of this artifact are performed in another case study related to the DW/BI Strategy Analysis in a Public Organization (LNEC) (see Section 1.3.2).

Finally, the fourth DSRM iteration (Iter. 4) describes the design and development of the final versions of artifact #1 (version 3) and artifact #4 (version 3), and of two new artifacts: the DW Ontology (LDWOWL) (artifact #5) and the LDWOWL-BSO Link (artifact #6). Version 3 of the Integration Framework (artifact #1) integrates all the other artifacts. The demonstration of this version is performed in the case study related to the DW/BI Strategy Analysis in a Public Organization (LNEC), presented in Section 1.3.2.

1.4.1. Identify Problem and Motivate

In the current state-of-the-art [5], ontologies are used to support DW/BI tasks, such as Dimensional Modeling, Requirement Analysis, ETL, and BI Application Design. Authors present a variety of motivations for ontology-driven solutions in DW/BI, such as eliminating or solving data heterogeneity/semantics problems, increasing interoperability, facilitating integration, or providing semantic content for requirement and data analysis. However, the integration of domain knowledge in DSS to enhance BI exploration was an identifiable gap in DSS research. Organizations already have DSS in place, either specialized systems or typical DW/BI systems, to address their decision-support needs. Semantic information provided by domain-specific ontologies can be used to support the analysis and exploration of these existing systems.

The analysis of the current state-of-the-art resulted in the first communication related to this doctoral thesis, titled "Incorporation of Ontologies in Data Warehouse/Business Intelligence Systems - A Systematic Literature Review", published in 2022 and presented in Chapter 2. The research presented in this doctoral thesis addresses the following five research questions:

RQ1: How can Semantic Web technologies complement current BI systems?
RQ2: How can the interoperability between DSS and other systems enhance decision-making using Semantic Web technologies at strategic and operational levels?
RQ3: To what extent can the use of Semantic Web technologies improve the interoperability between DSS and operational systems?
RQ4: How can the strategic elements of a BSC be formalized, ensuring their alignment with the various organizational levels (strategic, tactical, and operational)?
RQ5: To what extent can the use of Semantic Web technologies improve the interoperability between DW/BI systems and the organizational strategy?

1.4.2. Define Objectives of a Solution

The primary goal of this research is to explore the use and integration of SW techniques, such as ontologies, with DW/BI systems to enhance the decision-making process. Two main objectives were identified and addressed during this research: a) Represent and explore complex domain-specific concepts (related to the real-world case studies - civil engineering structures and organizational strategy); and, b) Take advantage of this knowledge in a DSS, ensuring new analytical possibilities or enhanced decision-support capabilities fostered by the alignment and integration between the DSS and other sources.

1.4.3. Design and Development

Six artifacts were designed and developed during this research. As stated before, the main contribution of work is the development of a framework that enables the support, navigation, and validation of domain-specific information, such as strategy, in BI environments. This model artifact, Integration Framework (artifact #1), delineates the essential components and prerequisites required to integrate ontological knowledge with DSS, as seen in Figure 1.5. The Integration Framework is designed to enable ontology-supported analyses and exploration of existing DSS data and information and improve decision support. Many organizations have already implemented DSS, ranging from specialized systems to typical DW/BI systems, to meet their decision-support requirements. The Integration Framework leverages SW techniques to complement existing DSS with a semantic layer and enhance the analysis and exploration capabilities of these existing systems with domain-specific knowledge.

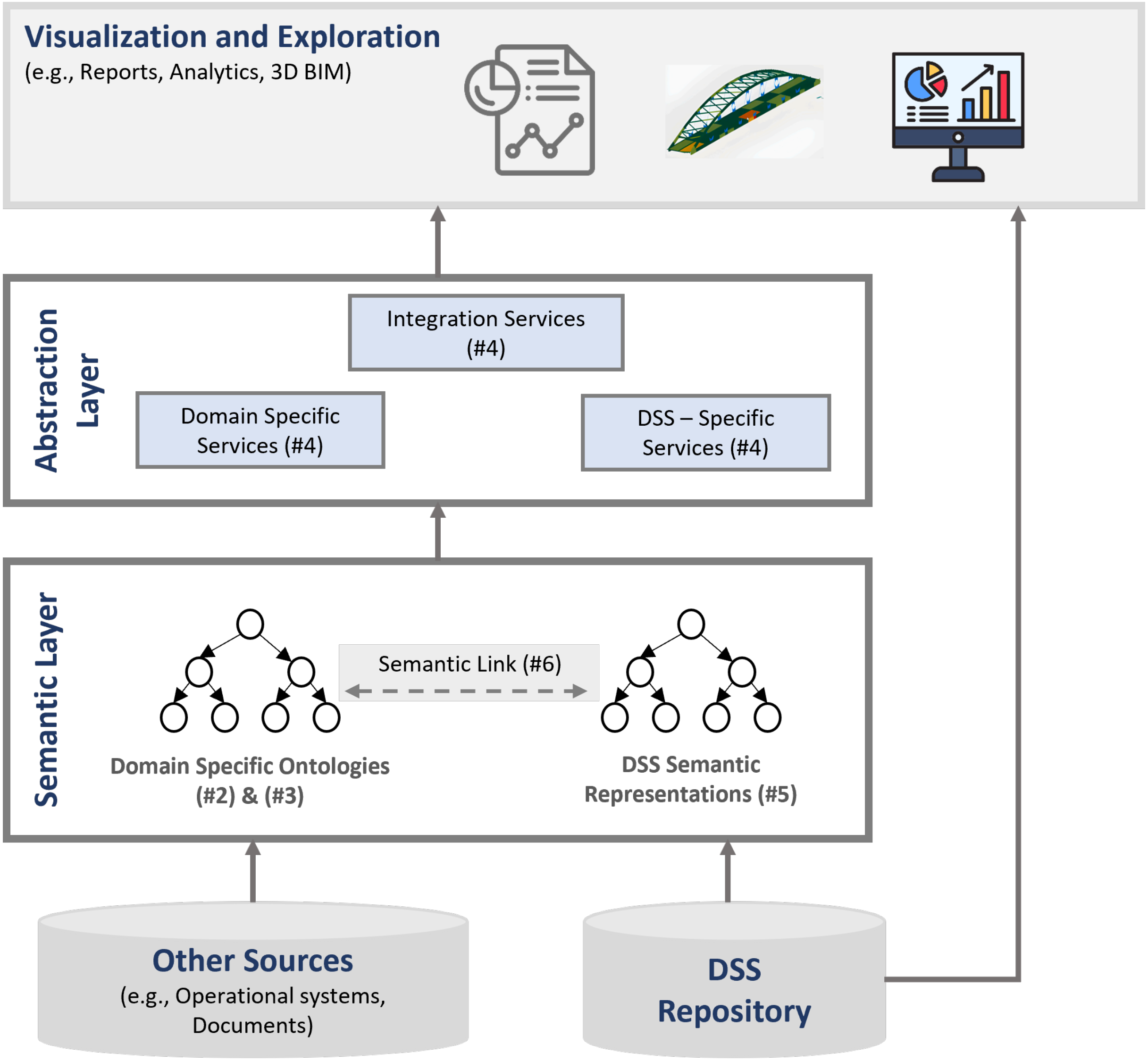


FIGURE 1.5. Integration Framework.

This work presents three versions of the Integration Framework (artifact #1), evolving throughout the research due to its application in real-world case studies. Version 1 is designed and developed during the DSRM's Iter. 1, providing a proof-of-concept for the integration solution in a 3D visualization environment. Version 2 is designed and developed during Iter. 2 to support decisions at an operational level, using BIM 3D visualization as the decision-support environment for users and decision-makers. Finally, version 3 is designed and developed during Iter. 4 to support decision at strategic-level, using a BI application tool. The same integration principles presented in Figure 1.5 are applied in each version of the framework, as will be demonstrated throughout this document, namely in Chapters 3 and 6.

The API Services (artifact #4) defines a set of methods, represented in the abstraction layer in Figure 1.5, providing an abstraction layer and ensuring accessibility for users and external applications. The API Services is intentionally designed to facilitate future expansion and enhancement in response to evolving needs, providing flexibility for various services and use cases associated with the underlying ontologies. Each version of the Integration Framework (artifact #1) adds a new set of services to support either domain-specific analyses, DSS-specific analyses, or services related to the integration between the domain-ontology and the DSS semantic representation.

To showcase the advantages of integrating SW techniques in DSS, such as DW/BI systems, two domain-specific ontologies were developed within the scope of this research, namely the Road Structures Ontology (artifact #2) and the Strategy Ontology (artifact #3). These model artifacts formalize the necessary domain knowledge pertinent to each case study, demonstrating the benefits of such formalization, including enhanced semantics, interoperability, and knowledge inference/reasoning capabilities. By employing these artifacts as part of the Integration Framework (artifact #1), DW/BI systems and other existing DSS can capitalize on these SW techniques, improving and facilitating their decision-making processes, ultimately empowering business users to make more informed and effective decisions. The Road Structures Ontology (artifact #2) is developed during the DSRM's Iter. 2 and is used as part of the Integration Framework (version 2 of artifact #1) related to the CoDEC project context (see Section 1.3.1), while the Strategy Ontology (artifact #3) is used in the DW/BI Strategy Analysis in a Public Organization case study (see Section 1.3.2). These ontologies are populated with data from external sources, such as operational systems (e.g., AMS) or unstructured data (e.g., strategy reports).

Finally, Iter. 4 describes the design and development of the final versions of the Integration Framework (version 3 of artifact #1) and its corresponding API Services (version 3 of artifact #4), and two new artifacts: the DW Ontology (LDWOWL) (artifact #5) and the LDWOWL-BSO Link (artifact #6). These artifacts are developed as integral components of the Integration Framework (version 3). The DW Ontology (LDWOWL) provides a semantic representation of the DSS repository related to the case study (i.e., the DW), necessary to provide an "anchor" point for the domain-specific ontologies. In the DW/BI Strategy Analysis in a Public Organization case study (see Section 1.3.2), the domain-specific ontology Strategy Ontology (artifact #3) is linked to the DW Ontology (LDWOWL) through a semantic link, named LDWOWL-BSO Link, which ensures the integration between the domain-specific ontology and the DSS.

1.4.4. Demonstration and Evaluation

As shown in Figure 1.4, research artifacts are demonstrated and evaluated throughout this research, providing insights for the next iteration of design and development and leading to new artifact versions.

The design and development of version 1 of the Integration Framework (artifact #1) is demonstrated in an initial pilot case related to the CoDEC project (see Section 1.3.1) during the Iter. 1. This demonstration uses information related to the maintenance of light posts, available from the INTERLINK project[7]. This first approach aimed at assessing the feasibility of using ontological knowledge within a BIM environment while providing tools for asset managers to access and analyze this information, as shown in Figure 1.6. The BEXEL Manager[8] was utilized as the BIM environment, where users interacted with an Industry Foundation Classes[9] (IFC) model containing a 3D representation of light posts. However, in this first version of the artifact, there is no semantic integration between this IFC model and the maintenance data inside the INTERLINK ontology.

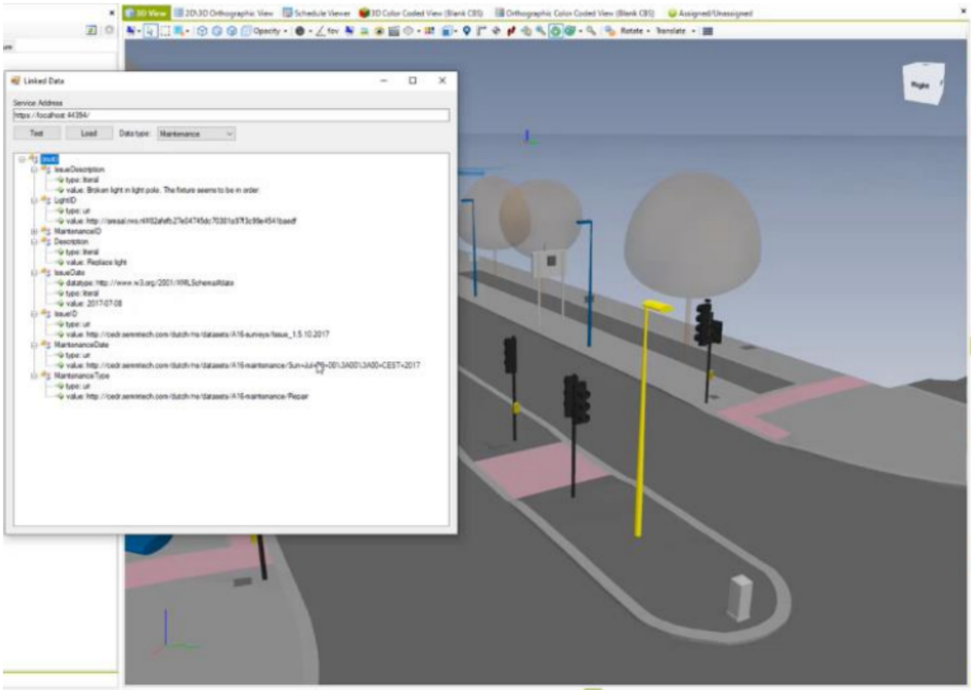


FIGURE 1.6. Ontological knowledge in a BIM environment. Retrieved from [1].

The applicability of ontological research artifacts is demonstrated by populating ontologies with data from different case studies. These knowledge bases are then utilized to answer a set of competency questions (defined with the help of domain experts), offering a comprehensive understanding of their capabilities and evaluating their effectiveness. Ontologies and their formalization are also validated using an online ontology validation tool (OOPS! - Ontology Pitfall Scanner! [46]).

During DSRM's second iteration (Iter. 2), the Road Structures Ontology (artifact #2) is demonstrated and evaluated with data related to three different pilot projects (tunnels, bridges, and pavements) from the CoDEC case study (see Section 1.3.1). The ontology is also evaluated as a base for integration between BIM and AMS, taking advantage of

[7]INTERLINK, INformation management for European Roads using LINKed data is a previous CEDR project. The main deliverable of this project was the European Road Object Type Library (EUROTL). More information related to the EUROTL is presented in Chapter 3

[8]https://bexelmanager.com/

[9]ISO 16739-1:2018 (2018). Industry foundation classes (IFC) for data sharing in the construction and facility management industries — part 1: Data schema. Geneva, CH: Standard, International Organization for Standardization.

the Integration Framework (version 2 of artifact #1) and the corresponding API Services (version 2 of artifact #4).

The Strategy Ontology (artifact #3) is initially formally demonstrated and evaluated based on a public university library use case, during the DSRM's third iteration (Iter. 3). Further, a second demonstration phase is presented during this DSRM iteration, populating the ontology with knowledge related to the DW/BI Strategy Analysis in a Public Organization (LNEC) case study (see Section 1.3.2). The Strategy Ontology is once more evaluated, taking advantage of SW techniques to assess strategy formulation and execution.

Lastly, version 3 of the Integration Framework (artifact #1) is demonstrated and evaluated, during Iter. 4, regarding its capacity to support the integration, alignment, and traceability between strategic models and the organizational information systems. This integration is necessary to provide data to the BSC's performance indicators, effectively bridging the gap between strategy definition, provided by the Strategy Ontology (artifact #3), and data. The artifacts designed and developed during Iter. 4, namely API Services (version 3 of artifact #4), DW Ontology (LDWOWL) (artifact #5), and the LDWOWL-BSO Link (artifact #6), are demonstrated and evaluated as part of the Integration Framework (version 3 of artifact #1).

1.4.5. Communication

The work produced within the scope of this doctoral thesis led to the publication of three articles in international journals, and the submission of a fourth article, which are presented in this manuscript as a compilation of key communications. The following journal articles (JA) were produced:

JA1 Lorvão Antunes, A., Cardoso, E. & Barateiro. J. (2022). **Incorporation of Ontologies in Data Warehouse/Business Intelligence Systems - A Systematic Literature Review**. International Journal of Information Management Data Insights, 2(2), 100131
DOI: https://doi.org/10.1016/j.jjimei.2022.100131
Website: https://www.sciencedirect.com/science/article/pii/S266709682200074X
ISSN(s): 2667-0968 (print); 2667-0968 (online)
Index: Scopus CiteScore (2023, Provisional): 18.7, **Q1[T5]** [Library and Information Sciences, Management Information Systems, Information Systems, Information Systems and Management], Q1[T10] [Artificial Intelligence]; Scimago Journal Ranking (2022): Q1[T5] [Library and Information Sciences], Q1[T10] [Management Information Systems, Artificial Intelligence, Information Systems, Information Systems and Management]
Times Cited: Scopus: 16; Google Scholar: 28 (April 2024)

JA2 Lorvão Antunes, A., Barateiro. J., Marecos, V., Petrović, J. & Cardoso, E. (2024). **Ontology-based BIM-AMS Integration in European Highways**. Intelligent Systems with Applications, 200366.

DOI: https://doi.org/10.1016/j.iswa.2024.200366
Website: https://www.sciencedirect.com/science/article/pii/S2667305324000425
ISSN(s): 2667-3053 (print); 2667-3053 (online)
Index: Scimago Journal Ranking (2022): **Q1** [Computer Science Applications; Computer Science (miscellaneous)]

JA3 Lorvão Antunes, A., Barateiro. J. & Cardoso, E. **The Balanced Scorecard Ontology: A Semantic Approach to Enhance Strategy Management**.
Status: Awaiting approval, in review process (minor reviews submitted/answered in January 2024)
ISSN(s): 0360-8581 (print); 0360-8581 (online)
Index: Scopus CiteScore (2023, Provisional): 7.3, **Q1** [Management of Technology and Innovation, Strategy and Management]; Scimago Journal Ranking (2022): Q2 [Management of Technology and Innovation, Strategy and Management]

JA4 Lorvão Antunes, A., Barateiro. J. & Cardoso, E. (2024). **Strategic Analysis in the Public Sector Using Semantic Web Technologies**. Digital Government: Research and Practice.
DOI: https://doi.org/10.1145/3656587
Website: https://dl.acm.org/doi/10.1145/3656587
ISSN(s): 2639-0175 (print); 2639-0175 (online)
Index: Scimago Journal Ranking 2022: **Q2** [Information Systems, Computer Science Applications]

JA1 is a systematic literature review article aiming to identify the problem and provide motivation. It analyzes publications related to the use of ontologies in DW/BI systems. JA2 reports on developments related to the second iteration of DSRM (Iter. 2), discussing the domain-specific ontological artifact Road Structures Ontology (artifact #2) and the API Services (version 2 of artifact #4), as part of the second version of the Integration Framework (version 2 of artifact #1). JA3 and JA4 describe the outcomes of the third iteration of DSRM (Iter. 3). JA3 introduces the Strategy Ontology (artifact #3), presenting its design and development, demonstration and evaluation, based on a public university library use case. JA4 focuses on the application of the Strategy Ontology, together with other SW techniques, in the real-world case study of the DW/BI Strategy Analysis in a Public Organization (see Section 1.3.2).

A fifth article (JA5) is currently in production, focusing on automating the integration of SW with DW/BI systems to enhance the decision-making process (Iter. 4). A semantic representation of the DSS component is necessary to implement the Integration Framework (artifact #1) and seamlessly integrate DW/BI data with domain-specific ontologies, as illustrated in Figure 1.5. In the DW/BI Strategy Analysis in a Public Organization case study, the DW Ontology (LDWOWL) (artifact #5) ensures the representation of the DW. Establishing a semantic link between LDWOWL and the Strategy Ontology (artifact #3), ensured by the development of LDWOWL-BSO Link (artifact #6), enables the integration of real performance data from the DW with strategic information. The use of these artifacts as part of the third version of the Integration Framework provides managers with enhanced decision-making capabilities within their BI environment. The work concerning the design and development of these artifacts, including the necessary API Services (version 3 of artifact #4), will be reported in this manuscript in Chapter 6). Table 1.1 outlines each JA's communications contribution in relation to this doctoral thesis's outcomes.

Table 1.1. JA in relation to the thesis's outcomes.

	Communications				
Research Outcomes	**JA1**	**JA2**	**JA3**	**JA4**	**JA5**
SLR	X				
Integration Framework (artifact #1)		X			X
Road Structures Ontology (artifact #2)		X			
Strategy Ontology (artifact #3)			X	X	
API Services (artifact #4)		X			X
DW Ontology (LDWOWL) (artifact #5)					X
LDWOWL-BSO Link (artifact #6)					X

Furthermore, the following set of conference articles were published, related to this research:

CA1 Antunes, António Lorvão; Lopes, Miguel; Barateiro, Jose; and Cardoso, Elsa, "**GELCO: Gamified Educational Learning Contents Ontology**" (2023). In Proceedings of the 23rd Conference of the Portuguese Association for Information Systems (CAPSI 2023) (pp. 295-316). Beja, Portugal: Association for Information Systems. http://dx.doi.org/10.18803/capsi.v23.295-316

CA2 Oliveira, B., Henriques, A., Oliveira, Ó., Duarte, A., Santos, V., Antunes, A. and Cardoso, E. (2023). **A measure data catalog for dashboard management and validation.** In Proceedings of the 12th International Conference on Data Science, Technology and Applications - DATA. (pp. 381-389). Rome, Italy: SciTePress. 10.5220/0012088400003541

CA3 Biswas, S., Proust, J., Andriejauskas, T., Wright, A., van Geem, C., Kokot, D., Antunes A., Marecos V., Barateiro, J., Bhusari, S. & Jovanovic, U. (2021, November). **CoDEC: Connected Data For Road Infrastructure Asset Management.** In IOP Conference Series: Materials Science and Engineering (Vol. 1202, No. 1, p. 012002). Riga, Latvia: IOP Publishing. 10.1088/1757-899X/1202/1/012002

CA4 Biswas, S.; Andriejauskas, T.; van Geem, C.; Kokot, D.; Marecos, V.; Barateiro, J.; Lorvão Antunes, A.; Bhusari, S. & Petrovic, J. (2022). **Demonstrating Connectivity and Exchange of Data Between BIM and Asset Management Systems in Road Infrastructure Asset Management.** In International

Road Federation World Meeting & Exhibition (pp. 379-392). Dubai: Springer, Cham. https://doi.org/10.1007/978-3-030-79801-7_27

CA5 Marecos, V.; Antunes, A.; Petrovic, J.; Barateiro, J. (2022). **Nova Abordagem Para Sistemas De Gestão De Ponte Usando BIM.** 10º Congresso Rodo-ferroviário Português

Lastly, during this work, the following reports were published, related to the real-world case studies:

- **Pilot projects report and consolidated implementation resources**. CoDEC Project Deliverable D3A. (2021)
- **Enquadramento Estratégico: Proposta de Mapa Estratégico**[10]. LNEC Technical Report (0102/1310/20796). Barateiro J., Couto P., Antunes A., Sebastião A., Santos M.A. (2021)

1.5. Software and Tools

This section describes the main set of tools utilized in the research's execution and demonstration of its outcomes. Protégé[11], a widely-used ontology development environment, served as the primary platform for designing and developing the Road Structures Ontology (artifact #2), Strategy Ontology (artifact #3), and DW Ontology (LDWOWL) (artifact #5). These ontologies were then populated with data using Cellfie[12], a Protégé plugin specifically designed for this purpose. Cellfie's ability to import data from Excel spreadsheets, following Manchester syntax rules[13], facilitated the process of populating the ontologies with domain-specific information.

GraphDB[14], a graph database, played a crucial role in storing and managing the knowledge bases created from the ontology population process. These knowledge bases serve as repositories of structured information, accessible for querying via APIs. Python[15], coupled with the FastAPI[16] package, was instrumental in developing the API Services artifact, providing an abstraction layer for users and external applications to interact with the knowledge bases stored in GraphDB.

Lastly, different exploration environments were utilized to visualize and analyze the integrated data in each version of the Integration Framework. BEXEL Manager[17], a BIM 3D visualization tool, was employed in one version, offering a comprehensive visual representation of the integrated data related to operational-level decision support. PowerBI[18], a business intelligence tool, served as the exploration environment in another version, providing rich visualizations and analytical capabilities for strategic decision support. These

[10]In Portuguese. "Strategic Framework: Strategy Map Proposal".

[11]https://protege.stanford.edu/

[12]https://github.com/protegeproject/cellfie-plugin

[13]https://www.w3.org/TR/owl2-manchester-syntax/

[14]https://graphdb.ontotext.com/

[15]https://www.python.org/

[16]https://fastapi.tiangolo.com/

[17]https://bexelmanager.com/

[18]https://powerbi.microsoft.com/

tools complemented the Integration Framework, offering intuitive interfaces for users and decision-makers to interact with the integrated ontological knowledge and make informed decisions.

1.6. Thesis Structure

This thesis follows an article structure, with each chapter presenting key publications, particularly the JAs detailed in the preceding communications section (Section 1.4.5). In line with open research initiatives, all significant publications produced and accepted within the scope of this research have been or will be published in open access. Consequently, Chapters 2 and 3 feature the version of record (already published) for JA1 and JA2, respectively. Chapter 4 showcases the original manuscript for JA3 submitted to an international journal, pending acceptance, with minor reviews since January 2024. Additionally, Chapter 5 presents the accepted manuscript for JA4, already accessible online on the journal's webpage. It is worth noticing that the acceptance/publication dates are contingent on the journal's schedule. Hence, the displayed dates may not align with the chronological sequence of the work.

The final communication pertaining to this doctoral thesis is currently being produced. Chapter 6 describes the final contributions associated with this research, namely research artifacts DW Ontology (LDWOWL) (artifact #5) and LDWOWL-BSO Link (artifact #6), to be communicated in JA5. Each chapter begins with an introductory section delineating its contribution to the doctoral thesis. Finally, Chapter 7 presents the thesis conclusions, with a research summary and a discussion of contributions, current limitations, and avenues for future research.

CHAPTER 2

Journal Article 1

This chapter presents a systematic literature review (SLR) that aims to survey the existing literature regarding the use of SW in DW/BI systems and how SW can be used to improve the quality of insights from structured data. Specifically, the goal is to understand the how, where and why ontologies are being used to improve DW/BI systems' analytical capabilities or simplify processes within the DW/BI lifecycle.

This article (JA1) directly contributes to the Identify and Motivate and Define Objectives of a Solution phases of the DSRM, as seen in Figure 2.1. The analysis of knowledge representation techniques' current use and impact in DW/BI systems serves as this doctoral thesis research entry point, allowing the identification of research gaps for which the remaining works will contribute, namely the use of ontologies to support and enhance the analysis and exploration of existing DSS.

Article details:

- **Title**: Incorporation of Ontologies in Data Warehouse/Business Intelligence Systems - A Systematic Literature Review;
- **DOI**: https://doi.org/10.1016/j.jjimei.2022.100131;
- **Date**: 2022;
- **Journal**: International Journal of Information Management Data Insights;
- **Publisher**: Elsevier.

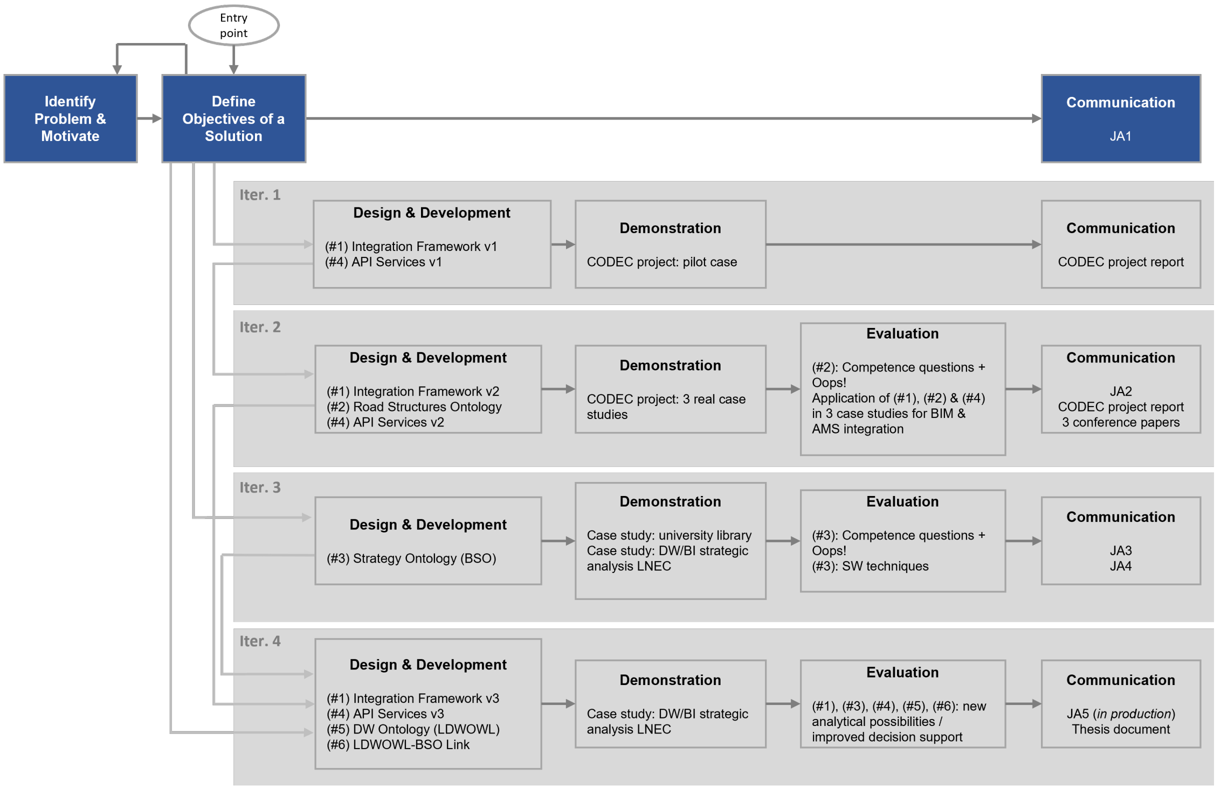


Figure 2.1. DSRM's JA1 Communication.

International Journal of Information Management Data Insights 2 (2022) 100131

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Review

Incorporation of Ontologies in Data Warehouse/Business Intelligence Systems - A Systematic Literature Review

António Lorvão Antunes[a,b,*], Elsa Cardoso[b,c], José Barateiro[a]

[a] LNEC- National Laboratory for Civil Engineering, Av. do Brasil 101, Lisbon 1700-075, Portugal
[b] ISCTE- Instituto Universitário de Lisboa, Department of Information Sciences and Technology, Av. Forças Armadas, Lisbon 1649-026, Portugal
[c] CIES-Iscte

ARTICLE INFO

Keywords:
Ontologies
Semantic web
Data warehouse
Business intelligence
Systematic literature review

ABSTRACT

Semantic Web (SW) techniques, such as ontologies, are used in Information Systems (IS) to cope with the growing need for sharing and reusing data and knowledge in various research areas. Despite the increasing emphasis on unstructured data analysis in IS, structured data and its analysis remain critical for organizational performance management. This systematic literature review aims at analyzing the incorporation and impact of ontologies in Data Warehouse/Business Intelligence (DW/BI) systems, contributing to the current literature by providing a classification of works based on the field of each case study, SW techniques used, and the authors' motivations for using them, with a focus on DW/BI design, development and exploration tasks. A search strategy was developed, including the definition of keywords, inclusion and exclusion criteria, and the selection of search engines. Ontologies are mainly defined using the Ontology Web Language standard to support multiple DW/BI tasks, such as Dimensional Modeling, Requirement Analysis, Extract-Transform-Load, and BI Application Design. Reviewed authors present a variety of motivations for ontology-driven solutions in DW/BI, such as eliminating or solving data heterogeneity/semantics problems, increasing interoperability, facilitating integration, or providing semantic content for requirements and data analysis. Further, implications for practice and research agenda are indicated.

1. Introduction

Business Intelligence (BI) is a term introduced in the mid-'90s, by the Gartner Group (Burton et al., 2006) and is now used as a cornerstone in most enterprises. It is seen as an "umbrella" term that encompasses applications, infrastructures, tools and practices used to improve and optimize decision-making and performance, through the access and analysis of data and information. Data Warehouse/Business Intelligence (DW/BI) systems are data-driven Decision Support Systems (DSS) (Sharda, Delen, Turban, Aronson, & Liang, 2015) that provide analytical and decision support capabilities to business users using an integrated repository (called DW) (Kimball & Ross, 2013). While these systems excel at handling and analysing structured, transaction-based data, they are not prepared to face the increasing variety of unstructured data Sawadogo & Darmont (2021). In addition, the SQL-based access to data typically provided by DW/BI systems is becoming inadequate for the types of data and the most recent algorithms used in Artificial Intelligence (AI) and Data Science analysis (Inmon, Levins, & Srivastava, 2021).

The need to extract information and gather knowledge from various sources is ever-increasing in a Big Data (BD) world, where data is created every second in countless shapes and forms (Gupta, Kar, Baabdullah, & Al-Khowaiter, 2018). Healthcare, Services and Financial Management, Public administration and governance, and (real-time) decision support systems are some of the Emerging Management Disciplines where BD and its analysis play a key role (Kushwaha, Kar, & Dwivedi, 2021). Organizations have started adapting the Data Lake (DL) Architecture as the primary storage for BD collection in their Information Systems (IS) (Inmon, 2016). When fully integrated and organised, this data can be used by data scientists and business users to power Data Science, BD Analytics, and BI tools and algorithms, thus realising their business value.

Data inside a DL can be divided into structured, textual, as well as other unstructured data (Inmon et al., 2021). Business activities typically generate structured data related to their business processes and transactions. Unstructured data is divided into textual data and data from other sources, such as sensors, images and video. Although there is an emphasis on unstructured data research in recent literature (Kumar, Kar, & Ilavarasan, 2021; Singh, Devi, Devi, & Mahanta, 2022), the importance and impact of structured data and DW/BI techniques in its analysis cannot be denied (Sharda et al., 2015). Due to its representation of business transactions, structured data analysis is crucial and has high

* Corresponding author.
E-mail addresses: antonio_lorvao@iscte-iul.pt, alfas@iscte-iul.pt (A.L. Antunes).

https://doi.org/10.1016/j.jjimei.2022.100131
Received 4 May 2022; Received in revised form 16 September 2022; Accepted 12 October 2022

business value. For example, most transaction-related Key Performance Indicators (KPIs) are available as structured data (e.g., sales value and product quantities). Moreover, structured historical data is also instrumental in developing descriptive, predictive, and prescriptive analysis, as recently demonstrated by Mishra, Urolagin, Jothi, Nawaz, & Haywantee (2021a) by applying machine learning methods to structured data and obtaining predictions about tourist arrivals to each country. Other recent examples of structured data analysis can be found in healthcare (Young & Steele, 2022), insurance analysis (Rawat, Rawat, Kumar, & Sabitha, 2021), and economics (Altuntas, Selim, & Altuntas, 2022).

DW/BI systems are designed, developed, and used to support the analytical needs of various departments or business areas within an organisation, providing a 'single version of the truth'. For this reason, it is essential to have common vocabularies or terminologies that allow business users to communicate with each other and with the development team (Kimball, Ross, Thornthwaite, Mundy, & Becker, 2008). IS researchers have increased their focus on Open[1] and FAIR[2] data, with interoperability and data sharing being a focal point in current research. Open and FAIR data principles are being integrated into several research areas to allow data and information to circulate and be accessible to those who need it (e.g., European Open data portals[3]). Knowledge representation formalisms, such as ontologies, are being developed to ensure that researchers have easier ways to access more data, information and knowledge in their fields of study. During the last years, the Internet evolved into the World Wide Web 3.0, also known as Semantic Web (SW) (Hitzler, 2021), in which data is encoded in a way that allows it to be shared, reused, and, most importantly, become machine-readable. Research and application fields, such as biology or computer science, have initiated efforts to facilitate the discovery and use of knowledge (Ristoski & Paulheim, 2016). The vast knowledge and value gained from integrating data across content, applications and systems is currently largely untapped (Gandon, 2018).

This shared semantics is fundamental to avoid misunderstandings or errors in situations in which natural language plays a key role, such as during the requirement gathering phase, data source analysis (context and meaning of each entity), or DW data analysis and exploration. Due to their semantic, formalisation and inference qualities, the integration of ontologies into DW/BI systems could help gather this knowledge at an organisational level and help mitigate or solve some of these problems. Ontologies could also provide new sources of information for the system, enriching data and providing new knowledge to the business user that would not otherwise be available within the organisation. Furthermore, ontology interoperability could be vital to link the DW/BI system and structured data to other DSS systems (inside or outside the organisation), with different knowledge bases or with DL-based architectures. This solution should also allow the integration of structured and unstructured data, either within the same ecosystem or in different IS, allowing communication between two different architectures (DW and a Data Lake, for example).

This systematic literature review (SLR) aims to survey the existing literature regarding the use of SW in DW/BI systems and how SW can be used to improve the quality of insights from structured data. Specifically, the goal is to understand the how, where and why ontologies are being used to improve the analytical capabilities of DW/BI systems or to simplify processes within the DW/BI lifecycle.

The remainder of this paper is structured as follows: Section 2 introduces background concepts from both DSS, DW/BI systems and ontologies. This section also introduces previous reviews with similar scope. The SLR methodology is presented in Section 3, defining the research questions, keywords, search engines and other criteria necessary for a SLR, followed by the preliminary results in Section 4. Section 5 presents the findings of the SLR and literature analysis, while Section 6 outlines the discussion, including its practical implications and research directions. Finally, conclusions are found in Section 7.

2. Background

This section presents background concepts needed for this systematic review. The section is divided into DSS, DW/BI systems and Ontologies.

2.1. Decision support systems

DSS are interactive computer-based systems intended to help business users identify and solve problems and assist in the decision-making process Power (2009). A DSS should offer quick and interactive information support to managers and business users, providing the "right information at the right time, with the right format" (Turban, Sharda, & Delen, 2010). The Association for Information Systems Special Interest Group on Decision Support Systems (AIS SIGDSS) adopts a classification of DSS proposed by Power (2009), which classifies DSS according to the type of components they use (Sharda et al., 2015): (a) **Communication-driven or Group DSS**: DSS that feature communication, collaboration and sharing (through technology) as their decision-making support; (b) **Data-driven**: DSS focusing on the access, analysis and manipulation of data. DW/BI systems and business process management systems are some examples of data-driven DSS; (c) **Document-driven**: DSS that emphasize the use (or retrieval), storage, management and analysis of documents; (d) **Knowledge-driven**: DSS that use knowledge bases and artificial intelligence (e.g., Expert systems, Data Mining); (e) **Model-driven**: DSS that focus on the use of quantitative models (such as any simulation model); (f) **Compound DSS**: Hybrid DSS that combine two or more of the previous components.

2.2. Data warehouse/business intelligence systems

As data-driven DSS, DW/BI systems are divided into two major subsystems: data warehousing ("getting data in") and business intelligence ("getting data out") (Watson & Wixom, 2007). The goal of data warehousing is to extract, transform and load data from different source systems into an integrated repository, the DW. The fact that data is distributed across heterogeneous source systems leads to various integration issues and challenges (e.g. different formats or representations of the same entities) that are addressed by the ETL process. BI retrieves data from the DW providing data-driven decision support to business users. Data can be presented and explored using reporting tools and dashboards or fed into data mining models to derive predictions and insights from analytical data.

Dimensional modeling is used in DW/BI systems, which unlike traditional data modeling (e.g., entity-relationship modeling), enables an intuitive and high-performance aggregation, retrieval and analysis of historical data (Kimball & Ross, 2013). In DW/BI systems data can be stored in star schemes or in cubes, also called multidimensional databases (Adamson, 2010; Kimball & Ross, 2013). The backbone of a dimensional model is the distinction between facts and dimensions. Facts are usually numeric and additive (although not all facts are additive) and represent important measurements of a given process (e.g., sales quantity, sales dollar amount). Dimensions represent the business entities that provide context to facts (e.g., Client, Date, Vendor), and are used to filter or aggregate the facts. Hierarchies are used to describe possible aggregation paths within a dimension. They use parent-child relationships between the dimension's attributes to drill up (i.e., remove detail) or drill down (i.e., add detail), allowing exploration of a certain context. For example, information about monthly sales of a company can be drilled down to a lower level of detail, such as daily sales, or aggregated (drill-up) to higher levels, such as semesterly or yearly sales.

According to Kimball et al. (2008), an enterprise DW corresponds to the union of subject-oriented subsets called data marts, if the following

[1] Open data handbook - http://opendatahandbook.org/
[2] Go fair initiative - https://www.go-fair.org/
[3] European open data portals -RT https://data.europa.eu/

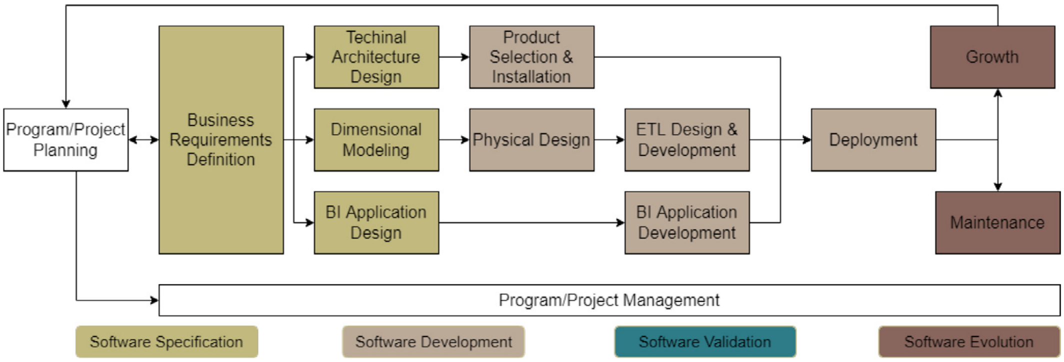


Fig. 1. Kimball's DW/BI lifecycle methodology. Adapted from Kimball et al. (2008).

conditions are met: each data mart must store granular data in dimensional models (i.e., with the lowest level of detail) and use conformed dimensions and facts (i.e. dimensions and facts share the same meaning across all data marts). Typically, a data mart is related to a single business process.

2.2.1. DW/BI system development

According to Sommerville (2011), software development involves four fundamental activities: Software specification, development, validation, and evolution. These activities are integral to most software process models, such as the waterfall model, incremental development or reuse-oriented software engineering. Agile methods were adopted and favored by software engineers in recent years to cope with the need for rapid system development and requirement changes during the software development process. Agile methods are also currently used in IS design, development and analysis (Siau et al., 2022). Agile methods focus on incremental deliveries with high customer involvement, simplicity, and change accommodation. They are used in DW/BI systems development to deal with the inherent high complexity of these integrated systems (Hughes, 2012).

Kimball's Lifecycle is a methodology to develop DW/BI projects. It can be described as a roadmap for effective DW design, development and deployment (Kimball et al., 2008). Fig. 1 displays the sequence of high-level tasks required for developing these systems. The iterative cycle includes tasks such as Business Requirements Definition, Dimensional modeling, ETL Design & Development, and BI Application Design and Development. It also presents a mapping between these tasks and the typical software development activities. Note that there is no task focused on validation, however, there are validation sub-process within most of the high-level tasks. For example, the ETL Design & Development process has its own lifecycle with specification, development, and validation activities.

A Planning phase is required to examine if the organization has the right elements and conditions for a successful implementation of a DW/BI system. A compelling business motivation for a DW, feasibility (from a technical, resources and data perspectives), IT-Business relationship, and current analytical culture are important factors when assessing the organizational readiness for the development of the DW/BI system. Business Sponsors that understand and believe in the project are also critical when transmitting the vision and impact of the DW project. The Planning phase also includes scope definition, benefit and cost estimations, staff selection and the development of a project plan. The Business Requirements Definition phase is connected to the Planning phase, and aims to understand the analytical needs and priorities of the business/organization. Requirements should be collected at both the organizational level (called the program level perspective) and for each business process (called the project level perspective). Business requirements impact every phase of the design, development and deployment of a DW/BI system.

The Dimensional Modeling phase comprises the design of conceptual data models following the dimensional approach. Subsequently, the Physical Design phase defines how data is physically structured in a database environment (i.e., indexing, partitioning, aggregation). The ETL process is responsible for extracting, cleaning, conforming and delivering source data to the DW. This process is critical within a DW/BI system, adding value and structuring the source data for later use by the BI applications. BI applications are designed and developed (using proprietary BI tools or in-house applications) to present an interface suitable to the user's needs for data presentation, exploration and analysis (e.g., reporting tools, dashboards, ad hoc queries, data mining).

Technical Architecture Design defines the overall architecture framework and vision based on business requirements, technical environment, and planned strategic directions. Once this framework is defined, tools and technologies for each component are evaluated and selected during the Product Selection and Installation task.

The Deployment phase begins when all the previous tasks have been completed. However, the DW/BI system still needs to be maintained, evolved and grown. The Maintenance phase ensures, among other things, continuous support for the business users and the correct operation of the system. The Growth task enables agile development of the DW/BI system, i.e., once a project is completed, the lifecycle can start over with new requirements for a new business process or data mart. Finally, Project Management ensures the correct tracking of each task, monitoring project status, issues, and change management.

2.3. BI and unstructured data

As shown in Fig. 2, BI has evolved over the years. The first generation of BI employed IT-generated reports and dashboards, while the second generation focused on self-service tools and analytical platforms (Ereth & Eckerson, 2018). The third and current generation of BI will be heavily affected by Artificial Intelligence, leading to the generation of more useful insights and making it easier for business users to interact with BI tools.

While the first and second generations of BI depended on data warehousing, using dimensional modeling to enable IT-Generated reports and dashboards or provide self-service analytics, the third generation will need different architectures to deal with unstructured data storage and analysis. The value of unstructured data analysis is proven in recent literature. For example, Neogi, Garg, Mishra, & Dwivedi (2021) present a sentiment analysis of Twitter posts (textual data) to study international public opinion related to the protests in India. A similar approach was used by Mishra, Urolagin, & Jothi (2020) to develop a recommendation system based on user reviews of tourists' points of interest. Another example is provided by Aggarwal, Mittal, & Battineni (2021), who surveyed the literature for different applications of Generative Adversarial Networks (a deep learning algorithm), such as 3D object generation, image processing, face detection, traffic control, and other image-based

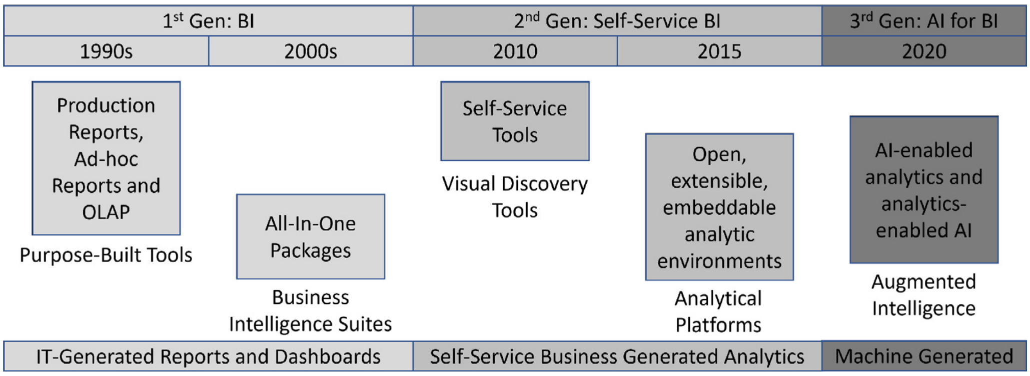


Fig. 2. Evolution of BI. Adapted from Ereth & Eckerson (2018).

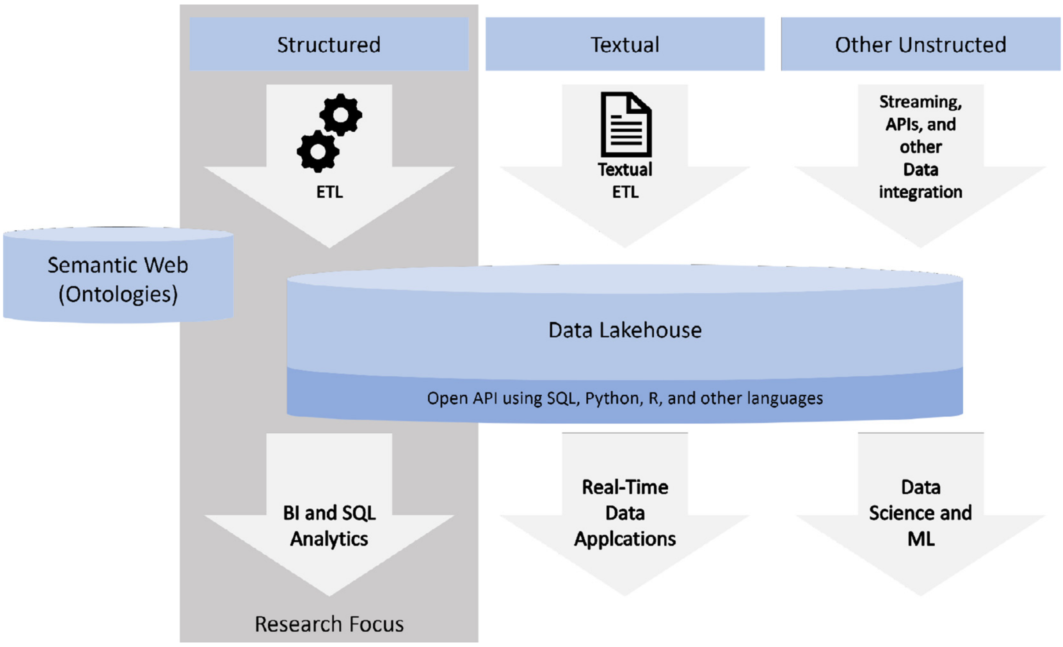


Fig. 3. Data Lakehouse Architecture. This study will focus on the impact of SW on structured data and its analysis. Adapted from Inmon et al. (2021).

applications. In most industries, however, BI can take advantage of both unstructured and structured data analysis. For example, Arjun, Kuanr, & Suprabha (2021) presented research on the banking industry where, depending on the banking sales process, the type of data used in its analysis differs. Structured data is used for customer loyalty/advocacy and purchase/service analysis, while unstructured data is used for purchase intention analysis.

Data Lakes are used to store raw, unfiltered data with cheap storage solutions for later analysis. This solution benefits the exploration and analysis of unstructured data, retrieved from social media, IoT, *etc.* Data is extracted from the DL via API and other data access services, which define and validate the structure, integrity and format of files as requested (which makes the DL a highly flexible solution). However, data fidelity and consistency are pointed out as the main disadvantages of a data lake (Sawadogo & Darmont, 2021).

The Data Lakehouse, an evolution of the DL architecture proposed by Inmon et al. (2021) in 2021 (see Fig. 3), still uses DW/BI techniques such as ETL (Extract, Transform, Load), BI and SQL Analysis when dealing with structured data. Sawadogo & Darmont (2021) propose that the DW should be seen as a part of the DL, or that the DL should be a data source for the DW. According to Ravat & Zhao (2019), the integration of DL architectures into IS as a DSS is still a subject of debate. While some authors advocate that the DL architecture is an "advanced version of DW", in contrast, Ravat & Zhao (2019) defend that both architectures should coexist in the same ecosystem, supported by the fact that DL and DW generally have different objectives and users.

2.4. Ontologies

Originally coined in 1613, the term "Ontology" refers to a branch of philosophy that studies the nature and structure of things/objects, their properties, events and relations (Smith, 2003). In Information Science, however, ontology refers to a "computational artefact" that encodes knowledge about a certain domain (Stephan, Pascal, & Andreas, 2007). While the meaning of ontology in computer science has been debated throughout the years, the most accepted definition was presented by Studer, Benjamins, & Fensel (1998, p.25): "An ontology is a formal, explicit specification of a shared conceptualization". A conceptualization is "an abstract, simplified view of the world that we wish to represent" (Gruber, 1993, p. 1), i.e., an abstract model with the relevant concepts of something. An explicit specification means that concepts, their relationships and constraints are explicitly defined and encoded. Moreover, the formalization of an ontology allows it to be machine-readable. The ontology should reflect an agreed-upon domain conceptualization in a community, i.e., a shared conceptualization (Studer et al., 1998).

The Resource Description Framework (RDF) was developed as a recommendation by the World Wide Web Consortium (W3C) to allow the "creation, exchange and use of annotations on the Web" in the form of

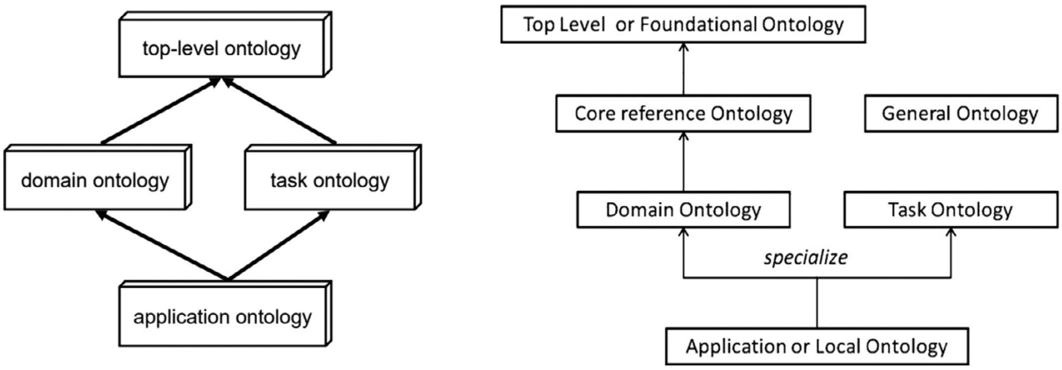


Fig. 4. Ontology types hierarchy based on Scope. Retrieved from Stephan et al. (2007) (left) and Roussey et al. (2011) (right).

triples (subject property object) (Pan, 2009). RDF Schema (RDFS) and Ontology Web Language (OWL) were developed on top of RDF and are used as standards in the Semantic Web effort. RDFS introduced class and hierarchy concepts, while OWL provides additional vocabulary and expressiveness (e.g., disjointedness, cardinality, object and data properties). There are three OWL sublanguages/types: Lite, DL and Full, with different levels of expressiveness. Normally, the choice of a language depends on the problem domain and modeling requirements, with an identified trade-off between expressiveness and inference capabilities (reasoning) (Lukasiewicz, 2008).

2.4.1. Ontology classifications

Ontology classifications are presented by Stephan et al. (2007) and Roussey, Pinet, Kang, & Corcho (2011) with different hierarchy paths between ontology levels (with lower ontologies specializing and inheriting concepts from the above). While slightly different, both classifications identify an application (or local) ontology as the most specialized ontology, followed by domain and task ontologies, and culminating in a top level (or foundational) ontology (see Fig. 4).

A summary of these ontology types is presented: (a) **Top level** ontologies are generic ontologies, with abstract and general concepts that can be used across domains and applications. They can be perceived as meta-ontologies and contain basic notions like objects, events and processes that are used in other ontologies. (b) **Domain and Task** ontologies contain knowledge about a certain domain or a certain task. The conceptualization of a domain should be independent of tasks (e.g., a biology ontology should be separated from a diagnostic task ontology). (c) **Application or Local** ontologies have the narrowest scope and support the resolution of a certain task in a certain domain. This means that they make use of both domain and task ontologies to fulfill their purpose. Roussey et al. (2011) classification introduces two additional types: the **Core Reference** ontology, which allows different communities to have different domain ontologies aligned and integrated with a standard, core, reference ontology; and the **General** ontology, which is not dedicated to a specific domain or field.

2.5. Overview of similar reviews

Other reviews have been published in recent years with a similar research objective. This section contains an analysis of these works to better understand the positioning and scope of the SLR presented in this paper.

Abelló et al. (2014) introduce Exploratory On-Line Analytical Processing (OLAP) as a way to ”discover, acquire, integrate and analytically query new external data.” The paper aims to survey how SW technologies can serve as a foundation for Exploratory OLAP, their feasibility and benefits, and identify future challenges. Challenges are found in three areas of research: (1) Schema Design (e.g., mapping, lack of SW tools, ontology evolution, and versioning), (2) Data Provisioning (e.g., ETL automation, complex semantic-aware integration), and (3) Semantic and Computational (e.g., reasoning at the instance level, expressiveness/inference trade-off). Future work includes SW-supported multidimensional querying and resolving scalability issues.

Laborie, Ravat, Song, & Teste (2015) present a survey of research results and outline future research challenges in BI and SW domains. Scalability, complexity, and heterogeneity of SW data are some of the main challenges that emerge when combining BI with SW to enhance BI analysis with web data and allow SW data analysis in BI tools. Two types of approaches are identified in the survey, OLAP-analysis oriented and Multidimensional modeling oriented. The first approach focuses on storing SW data in OLAP cubes to facilitate the analysis of information published on the web. The second approach provides compatible multidimensional modeling solutions that allow you to perform OLAP analysis directly on SW data (trying to overcome highly complex and time-consuming ETL processes). Due to the dynamic nature of web-published data, availability and consistency problems can emerge. Freshness can be partly forfeited in exchange for querying efficiency and data quality when materializing SW data in the DW. This trade-off and the automatic integration of SW data in the OLAP cube (automatically defining mappings at both schema and instance levels) are pointed out as the main future research directions.

Finally, Hussain, Al-Turjman, & Sah (2020) present a similar SW and OLAP integration analysis from Laborie et al. (2015). Furthermore, the authors discuss how different methods of integration can handle Big Data and the benefits from cloud computing application in BI in terms of scalability, cost effectiveness, data sharing, and reliability.

The abovementioned reviews, although relevant contributions, cannot be considered SLR since they analyze a small set of articles obtained without resorting to a research protocol indispensable to an SLR. The 2019 review by Wisnubhadra, Baharin, & Herman (2019), however, offers a survey strategy to analyze modeling and query of spatiotemporal multidimensional data on SW. Regarding the integration of ontological data in a DW, the authors mention the consistency of Linked Open Data in the DW as the main challenge, while acknowledging the proven advantages of OLAP.

This paper will present a systematic review with a comprehensive methodology and selection criteria of recent literature, with a focus on DW/BI design, development and exploration tasks, allowing a more specific analysis of ontology usage, integration and impact on each task. Each work will be classified based on the field of case studies, SW techniques used, and the authors' motivations for using them.

3. SLR methodology

This section introduces the research questions, the review protocol (see Fig. 5), and methods employed in this SLR, following the work presented by Budgen & Brereton (2006). To identify the relevant literature, a search strategy was developed, including the definition of keywords

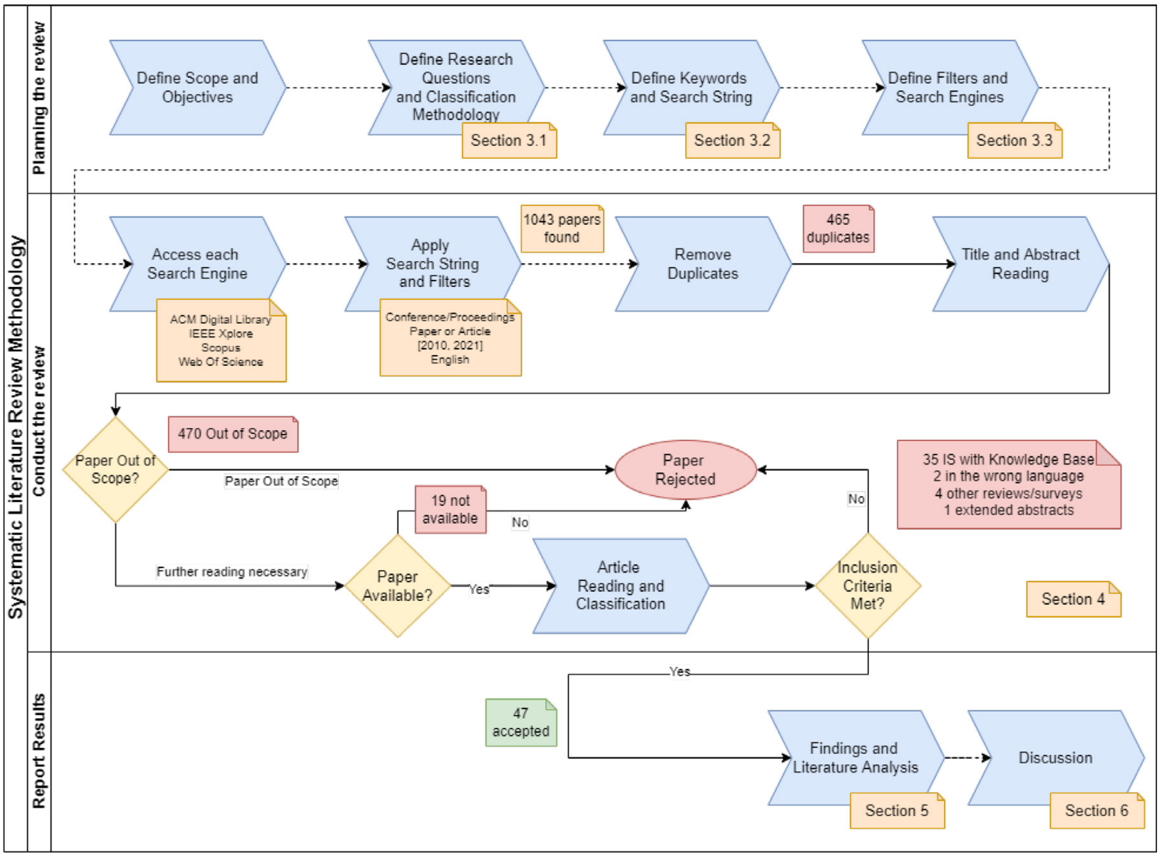


Fig. 5. SLR methodology.

(and search string), inclusion and exclusion criteria, and the selection of search engines.

3.1. Defining research questions and classification methodology

As previously stated, the main goal of this research is to gain insight into the existing literature concerning the use of ontologies in DW/BI systems. The following research questions are presented to guide the research:

RQ1: How are ontologies/knowledge bases being incorporated/integrated into DW/BI systems?

This research question looks to understand how SW techniques are being used to improve the quality of insights obtained from structured data in DW/BI systems. Information about ontology language and type is collected to gain insight into the use of SW techniques in each paper. Ontology type will be based on its scope. When omitted by the authors, ontologies are classified following the terminology presented in Section 2.4.1 and distinguished with an (*).

RQ2: In which high-level tasks of DW/BI system development are ontologies being used?

To better understand the impact of ontologies in DW/BI systems, works will be classified and analysed following a reference terminology for DW/BI development. Kimball's DW/BI lifecycle (see Section 2.2.1) is a well-known and well-established methodology (Cavalheiro & Carreira, 2016; Lukić, Radenković, Despotović-Zrakić, Labus, & Bogdanović, 2016) that was chosen to provide a classification reference terminology for DW/BI Task. The impact of the ontology should be limited to a task or part of the DW/BI lifecycle, such as Business Requirements Definition, Dimensional Modeling, and ETL Design & Development. Any exploratory task, such as data mining or OLAP, will be classified as BI Application Design.

RQ3: What are the reasons/gains presented for the utilization of SW techniques in DW/BI systems?

This research question seeks to identify the main advantages of the integration/incorporation of ontologies in DW/BI systems. The application scenario (or application field) is also collected to obtain a clearer vision of the impact of these techniques on DW/BI systems.

3.2. Defining keywords and search string

For the definition of keywords and search string, the recommendations of Silva & Neiva (2016) were followed. To fulfill the main goal of this research, which is to observe the impact of ontologies in DW/BI systems, synonyms and similar key terms were selected. To this end, keywords were divided in two groups.

Group 1 includes keywords related to DW/BI, specifically: "Data Warehouse", "Data Mart" and "Star Schema", and keywords related to the tasks from the DW/BI framework, such as "Dimensional Modeling" and "ETL". The keywords "Requirements", "Facts" and "Dimensions" were also added due to their relevance in DW/BI systems. Keywords such as "Decision Support System" were initially considered but then removed during the refinement process since any expert system based on ontologies is a knowledge-based DSS, leading to several out-of-scope papers being found.

Group 2 is comprised of keywords related to ontologies, such as "Ontology"/"Ontologies", "Ontological", "Knowledge Representation" and "Knowledge Base". "Semantic Web" was also added since is commonly used to refer to these types of techniques.

The search string will screen paper titles for the logical conjunction of any keyword in group 1 with any keyword in group 2 (see Table 1):

Table 1
Keywords in the search string.

Group 1	Business Intelligence; Data Warehouse(s); Data Warehousing; Data Mart; OLAP; Star Schema; Multidimensional; Dimensional Model(l)ing; ETL; Requirements; Facts; Dimensions
Group 2	Ontology; Ontologies; Ontological, Knowledge Representation; Knowledge Base; Semantic Web

Table 2
Results per Search Engine.

Search Engine	# of results
ACM Digital Library	31
IEEE Xplore	122
Scopus	562
Web of Science	328
Total	1043

Table 3
Results according to Accepted/Rejected outcome.

# of results	
Accepted	**47**
Rejected	**997**
Duplicates	465
Out-of-Scope (Title and Abstract reading)	470
IS with Knowledge Base	35
Not Available	19
Wrong Language	2
Other Reviews	4
Extended Abstract	1
Total	**1043**

Title:(”Business Intelligence” OR ”Data Warehouse” OR ”Data Warehouses” OR ”Data Warehousing” OR OLAP OR ”Data Mart” OR ”Dimensional Modeling” OR ”Dimensional Modelling” OR ”Star Schema” OR ”Multidimensional” OR ETL OR Requirements OR Facts OR Dimensions) AND Title:(Ontology OR Ontologies OR Ontological OR ”Knowledge Base” OR ”Knowledge Representation” OR ”Semantic Web”)

3.3. *Defining filters and search engines*

Under the university’s (blind information) network access agreement, the search string was used to gather research from the following search engines: ACM Digital Library (hdl.acm.org), IEEE Xplore (ieeexplore.ieee.org), Scopus (scopus.com) and Web of Science (webofknowledge.com). In addition to the search string, three filters were employed in the search, as follows: (a) document type: conference/ proceedings paper, article; (b) publication year: [2010, 2021]; and (c) language: English.

4. Conducting the SLR

This section introduces the preliminary outcomes of the SLR, following the methodology presented in Fig. 5. In total, 1043 documents were obtained from the different search engines (see Table 2), and applying the filters mentioned previously. Several duplicates were found in this phase, with a large overlap of papers between Scopus and other search engines.

From this initial set of documents, a first analysis was obtained by reading the title and abstract from each work. The main objective here was to identify out-of-scope works, which include research that does not mention DW/BI systems or any similar concepts in its title or abstract. Due to the use of keywords such as Requirements, a substantial set (470) of works were rejected in this phase. Ontologies are used in works related to requirements and software engineering due to their semantics and inference. However, analysis and requirements elicitation in generic software was considered out of scope for this SLR, explaining the high number of papers rejected in this first classification.

The remaining 108 works were fully analyzed to confirm that the documented research added to the scope and objectives of this SLR. Table 3 presents the main results of these analyses, presenting counts from the different outcomes (i.e., Accepted, and Rejected due to several reasons). The main reasons for rejections in the second analysis phase were the unavailability of the document and the research being out of scope for this SLR, in particular, IS with Knowledge Base. Despite the filters used in the search engines, a small number of documents still did not meet the necessary criteria for acceptance (e.g., papers not written in English). In the end, 47 documents were selected for further analysis and classification.

5. Findings

This section contains the main results and findings from the SLR. It is divided into two sections Bibliometrics, where year-wise and other statistics are presented, and Literature Analysis, which includes the outcome of the classification methodology.

5.1. *Bibliometrics*

Fig. 6 presents an evolution of works published per DW/BI task throughout the analyzed years (2010–2021). Three main conclusions can be drawn out: (a) There was a peak of publications in or before 2010, (b) the number of annual publications decreased between 2010 and 2013, stabilizing thereafter (with the exception of 2017), and (c) in the last few years the main focus of application of Semantic Web techniques was on BI application design tasks.

Of the 47 papers analysed, 36 were Conference Papers (76%), with only 11 works being published in journals. The International Conference on Information and Knowledge Management, with four works, and the International Convention on Information and Communication Technology, Electronics and Microelectronics, with three, are the conferences from which more research originated.

5.2. *Literature analysis*

Looking at Table 4, we can see a diverse set of research and application fields (e.g., Academic, Healthcare, Sales) where SW technologies are being used in conjunction with DW/BI systems. This was to be expected since both areas have abundant and overlapping fields of application. The use of OWL (SW standard) and its sub-languages (Full, Lite and DL) by most papers is also an expected result. The use of non-standardized ontologies may undermine their potential as it hinders their interoperability. The widespread use of domain- and task-specific ontologies is inevitable when there is a need to capture business and process detailed context, something for which generic ontologies, with abstract and broad concepts, are usually not suitable.

The remainder of this section divides results based on the Kimball’s DW/BI lifecycle tasks where ontologies are being used. Since no research was found on activities such as Maintenance and Project Management, these tasks were not considered. The primary motivation of each work is collected and presented in Table 5. Fig. 7 presents a distribution of the number of works per DW/BI task. There is a clear focus of research on Dimensional Modeling and BI Application Design. It is important to

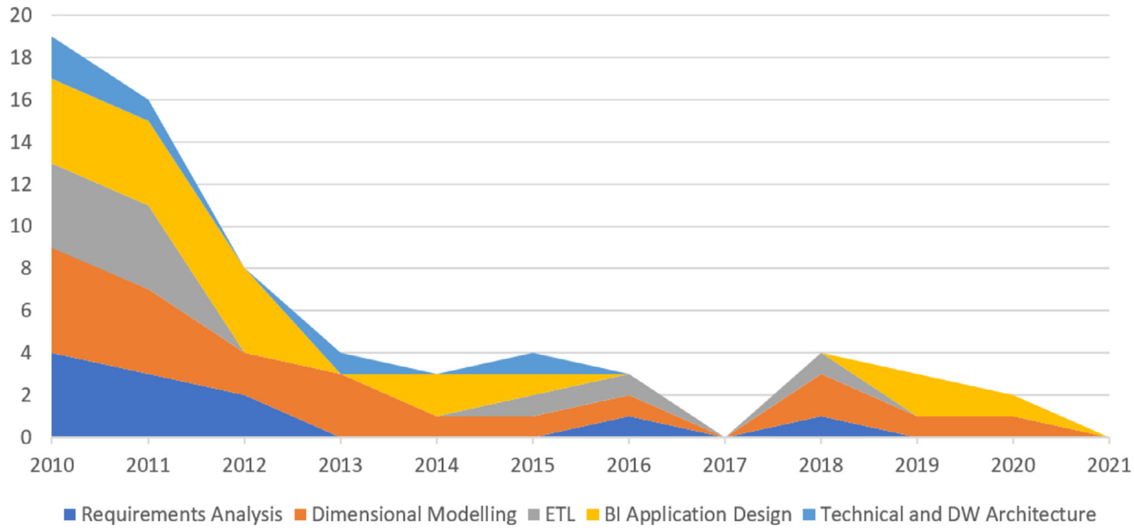


Fig. 6. Evolution of works published per DW/BI Task.

Table 4
Results classification.

Ref.	Year	Source	Case Study	Ont. Lang.	Ont. Type
Jiang et al. (2010)	2010	Scopus; IEEE	Health Care	OWL	Domain
Romero & Abelló (2010)	2010	Scopus; WoS	Car Rental	OWL-DL	Domain
Khouri & Ladjel (2010)	2010	ACM; Scopus	N/A	OWL	Global/Local
Kurze et al. (2010)	2010	Scopus; WoS; IEEE	Sales	OWL	Core
Nimmagadda et al. (2010)	2010	Scopus; IEEE	Human Ecosystem	N/A	Domain
Limongelli et al. (2010)	2010	Scopus; WoS	Academic	N/A	(*) Domain
Nicolicin-Georgescu et al. (2010)	2010	IEEE	N/A	OWL	(*) Task
Nicolicin-Georgescu et al. (2010)	2010	Scopus; WoS	N/A	OWL	(*) Task
Taa et al. (2010)	2010	Scopus	Academic	OWL	(*) Task
Simitsis et al. (2010)	2010	Scopus; WoS	N/A	OWL-DL	Domain/Application
Tanuska et al. (2010)	2010	Scopus; IEEE	Academic	UML	(*) Domain
Wu et al. (2010)	2010	Scopus; WoS	N/A	N/A	(*) Application
Abelló & Romero (2010)	2010	ACM	Car Rental	OWL	Domain
Zaharie et al. (2011)	2011	Scopus; WoS	Sales	OWL	Domain/Application
He et al. (2011)	2011	Scopus; IEEE	N/A	N/A	Domain
Ta'a & Abdullah (2011)	2011	Scopus; WoS	Natural Gas Distribution	OWL	(*) Task
Taa et al. (2011)	2011	Scopus	Natural Gas Distribution	RDF/OWL	(*) Task
Nimmagadda et al. (2011)	2011	Scopus; WoS; IEEE	(E-)Health Care	N/A	Domain
Villanueva Chávez & Li (2011)	2011	Scopus; IEEE	Auto parts company	OWL	Domain
Neumayr et al. (2011)	2011	Scopus; WoS	Health Insurance	OWL	Domain
Vanea & Potolea (2011)	2011	Scopus; WoS	Medicine	N/A	Domain
Wu et al. (2011)	2011	Scopus; WoS	Electronic Sales	N/A	Domain/(*) Local
Aymoré Martins. et al. (2012)	2012	Scopus	N/A	N/A	Upper
Fernandes et al. (2012)	2012	Scopus; WoS; IEEE	Planning and Budget	N/A	Task/Application
Prat et al. (2012b)	2012	ACM; Scopus	Agriculture	OWL-DL	(*) Global
Neumayr et al. (2012)	2012	ACM; Scopus	Health Care	N/A	(*) Domain
Prat et al. (2012a)	2012	Scopus; IEEE	Spatiotemporal data	OWL-DL	(*) Upper/Fundation
Bellatreche et al. (2012)	2012	Scopus; IEEE	N/A	UML	Domain
Tria et al. (2014)	2013	Scopus; WoS	Products Wholesale	N/A	Domain
Bargui et al. (2011)	2012	Scopus; WoS	Sales	N/A	Domain
Liu & Iftikhar (2013)	2013	Scopus; WoS	Sales	OWL	Domain
Gulic (2013)	2013	Scopus; WoS; IEEE	Invoices	OWL Lite	(*) Domain
Nimmagadda & Dreher (2014)	2014	Scopus; WoS; IEEE	Petroleum	OWL	Domain
Etcheverry et al. (2014)	2014	Scopus	Sales	RDF	(*) Domain
Szwed et al. (2015)	2015	Scopus; WoS	Insurance	OWL	(*) Global
Matei et al. (2015)	2015	Scopus	Energy Consumption	RDF	(*) Domain
Moreira et al. (2015)	2015	Scopus	National Electric System	OntoUML	Foundational/Domain
Oliveira & Belo (2016)	2016	Scopus; WoS	N/A	OWL	(*) Task
Aadil et al. (2016)	2016	Scopus; WoS; IEEE	Waste Management	OWL	Global / Local
Ren et al. (2018)	2018	Scopus; IEEE	Health Care	N/A	Domain
Pticek & Vrdoljak (2018)	2018	Scopus; WoS; IEEE	N/A	RDF	Local
Laadidi & Bahaj (2018)	2018	ACM; Scopus; WoS	N/A	OWL	N/A
Brahmi (2019)	2019	Scopus; WoS; IEEE	Sales	N/A	Domain
Amaral & Guizzardi (2019)	2019	Scopus; WoS	Education	OntoUML	Fundational
Namnual et al. (2019)	2019	Scopus	Higher Education	OWL	Domain
Quamar et al. (2020)	2020	ACM; WoS	Healthcare	OWL	Domain
Chakiri et al. (2020)	2020	Scopus; WoS	Local Governance	OWL	Global / Local / Domain

Table 5
Authors' Motivations.

Ref	Year	Motivation
Jiang et al. (2010)	2010	Eliminate data heterogeneity
Romero & Abelló (2010)	2010	Support end-user requirements elicitation and DW's design tasks / Identify and elicit unknown analysis capabilities from data sources
Khouri & Ladjel (2010)	2010	Querying DW in a semantic level and allowing integration with other DWs
Kurze et al. (2010)	2010	Provide the vocabulary for the integration of different OLAP applications
Nimmagadda et al. (2010)	2010	Knowledge sharing and reuse, ensuring concept interoperability across web sources
Limongelli et al. (2010)	2010	Develop an OLAP technique to help teachers to analyze Learning Objects stored in web repositories
Nicolicin-Georgescu et al. (2010)	2010	Improve service levels by managing DW cache allocations with autonomic computing
Nicolicin-Georgescu et al. (2010)	2010	Improving the allocation of shared resources
Taa et al. (2010)	2010	Obtain ETL process specification from DW requirements and business semantics / Solve limitations in modeling and designing DW systems
Simitsis et al. (2010)	2010	Assist in the collection and validation of metadata for ETL processes' conceptual design
Tanuska et al. (2010)	2010	Define the base classes to determine the influential factors in student failures
Wu et al. (2010)	2010	Support the mining process by reducing user involvement in query formulation and submission
Abelló & Romero (2010)	2010	Discover meaningful IDs from domain ontologies
Zaharie et al. (2011)	2011	Increase DW's responsiveness and adaptability to the information needs from the decision-making process
He et al. (2011)	2011	Formalize the users' needs into a conceptual model with semantic information and solve heterogeneity problems
Ta'a & Abdullah (2011)	2011	Reconciliation of the user semantics toward the modeling of the DW
Taa et al. (2011)	2011	Resolve user requirements ambiguity and semantic heterogeneity problems during data integration and transformation
Nimmagadda et al. (2011)	2011	Solve connectivity, communication and interaction problems and facilitate data interpretation
Villanueva Chávez & Li (2011)	2011	Automate extraction and categorization of data sources, generation of logical and physical data models and generation and data storage routines
Neumayr et al. (2011)	2011	Provide comparative data analysis and guide the business user through different kinds of knowledge
Vanea & Potolea (2011)	2011	Obtaining a semantically enhanced DW, with a flexible environment for query submission
Wu et al. (2011)	2011	Provide an active knowledge re-discovering mechanism, with better data mining models, fewer ineffective patterns dissemination and able to discover new concept rules
Aymoré Martins. et al. (2012)	2012	Integrate heterogeneous information concepts in a collaborative BI environment
Fernandes et al. (2012)	2012	Fast and automatic implementation of the BI system
Prat et al. (2012b)	2012	Leverage OWL-DL reasoning to ensure the reliability of OLAP analysis (e.g., summarization correctness)
Neumayr et al. (2012)	2012	Define and represent business analysts' hierarchical and multidimensional concepts
Prat et al. (2012a)	2012	Represent the multidimensional model as an OWL-DL ontology, increasing formalization and inference
Bellatreche et al. (2012)	2012	Make user requirements persistent into DWs and identify SQL queries for each business goal
Tria et al. (2014)	2013	Automatically integrate different schemas and solve syntactical/semantic inconsistencies
Bargui et al. (2011)	2012	Automation of analytical requirements elicitation, overcoming lack of domain knowledge
Liu & Iftikhar (2013)	2013	Describe semantics of big dimensions and automate the modeling process
Gulic (2013)	2013	Facilitate analysis of semantic data sources
Nimmagadda & Dreher (2014)	2014	Support data integration and information sharing; Facilitate data mining, visualization and interpretation
Etcheverry et al. (2014)	2014	Represent multidimensional models in the SW
Szwed et al. (2015)	2015	Provide a formal description of DW architectures
Matei et al. (2015)	2015	Model distributed multidimensional SW data, increasing interoperability of OLAP frameworks
Moreira et al. (2015)	2015	Increase the semantic expressiveness of the multidimensional modeling
Oliveira & Belo (2016)	2016	Support and enable the configuration and instantiation of ETL patterns
Aadil et al. (2016)	2016	Support a combination of need-driven and data-driven DW design
Ren et al. (2018)	2018	Optimize DW requirement analysis process and eliminate semantic heterogeneity
Pticek & Vrdoljak (2018)	2018	Enrich NoSQL database contents, allowing integration with traditional DWs
Laadidi & Bahaj (2018)	2018	Automatically identify multidimensional concepts in OWL sources
Brahmi (2019)	2019	Reduce system resource consumption and improve the mining process efficiency
Amaral & Guizzardi (2019)	2019	Improve semantic expressiveness of multidimensional models, improving communication and interoperability
Namnual et al. (2019)	2019	Enhance digital entrepreneurs' competencies for higher education
Quamar et al. (2020)	2020	Explore and obtain insights from a dynamic and intuitive conversational system interaction
Chakiri et al. (2020)	2020	Integrate data sources with existing requirement multidimensional schemes and minimize misconceptions or misunderstandings between different stakeholders

note that, in each work, the use of ontologies might cover more than one task.

For example, when ontologies are used in Requirement Analysis tasks, most of the time (10 out of 11 works), their impact on other tasks, such as Dimensional Modeling (6) or ETL (4), is also mentioned. On the other hand, when ontologies are used for BI Application Design, works usually only cover the impact of the ontology in this specific task. This disparity is expected since Requirement Analysis impacts most or all other development tasks. In contrast, BI applications design, which describes any information retrieval or exploration task, is done after the data is already in place and does not impact other design and development tasks.

Prior to the analysis of ontological impact on each DW/BI task, word clouds were obtained using Python's *wordcloud* package[4]. The abstracts of each study were used to generate the word clouds, after removing the keywords used in the SLR (see Table 1) and the word "Paper". Fig. 8 includes word clouds for all abstracts, as well as for each of the DW/BI tasks, in which only relevant documents to each task were used.

Starting by analysing all the abstracts, keywords such as "data", "design" and "system" are highlighted as they are employed in more than half of the abstracts. References to the (multi)dimensional model or DW model explain the frequent use of "model". Some authors present an ontology-"based" "approach", "process" or "method", words also typically used to describe research artifacts. The words "semantic", "information" and "knowledge" are also frequent, which is coherent with the area of research. Interestingly, "decision support" and "interoperability" do not seem to describe the type of systems or tools presented by the authors.

When observing the remaining word clouds, some keywords appear more frequently depending on the task. Requirement Analysis focuses on "conceptual" design and processes and "business" "users". "Data

[4] https://pypi.org/project/wordcloud/

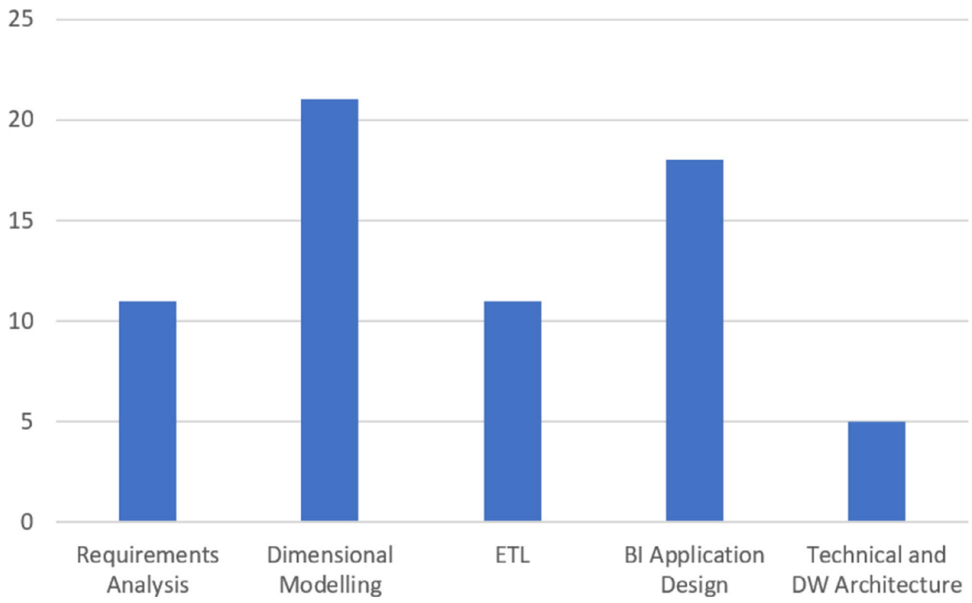


Fig. 7. Number of works per Kimball's DW/BI lifecycle task.

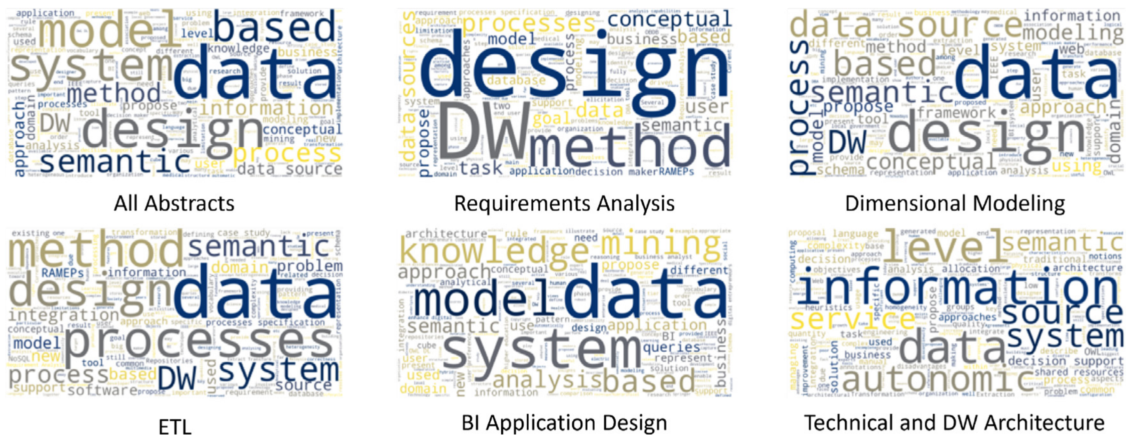


Fig. 8. Abstracts Word clouds.

Sources" also appear as keyword since they are analysed during the requirements phase. Identical keywords are used for Dimensional Modeling and ETL. The word cloud for Dimensional Model's word cloud, "data source" and "data" appear with higher frequency, with "domain" also appearing as an important keyword, related to the type of ontology used in some of the proposed methods by the authors. In ETL, the focus shifts to "integration" and ETL "process(es)". Looking at the word cloud for BI Application Design's word cloud, the words "knowledge", "mining" and "analysis" appear more predominant, which is, again, consistent with the types of solutions presented by the authors. The word "model" is also emphasized since some solutions extract dimensional models into ontologies. Lastly, in the Technical and DW Architecture word cloud, the words "level", "information", "autonomic" and "service" are highlighted. Most of the works related to this task focus on service level agreements (quality of service) for the DW/BI systems and how to improve it using "autonomic" computing.

5.2.1. *Requirement analysis*

Ontologies proved to be valuable in formalizing the needs and requirements of users, with the added semantics being used to aid in requirements elicitation, reconcile users' semantics and resolve semantic ambiguity. In most cases, the knowledge from the requirement-filled ontology is used to create a dimensional model that fulfills user requirements. Dimensional modeling concepts, such as dimensions, facts, and hierarchies, are identified on the ontology and mapped into a dimensional model. S2RWC (Semantic Sources and Requirements driven tool for DW Conceptual design) (Khouri & Ladjel, 2010) and AMDO (Automating Multidimensional Design from Ontologies) Romero & Abelló (2010) are two illustrative methods that use ontologies to enable a semantic integration and unification of user requirements, and to support user requirements elicitation, respectively.

The materialization of data-driven requirements in ontologies can be used to integrate data from multiple data sources. Ontologies are used to capture the semantics of the involved data stores based on each user's decision needs. The alignment of these ontologies allows the integration of all concepts expressed by users in a single global ontology that can be used to build the dimensional data model (Aadil, Wakrime, Kzaz, & Sekkaki, 2016). Inference on a domain ontology, constructed following extracted terminology/semantics of the involved (source or target) data stores, can serve as a means for ETL requirements elicitation and design (Simitsis, Skoutas, & Castellanos, 2010). Zaharie, Pugna, & Radulescu (2011) propose the use of REA (Resource-Event-Agent) enterprise domain ontology to define user requirements at both operational (resources, events, and agents) and policy levels (use of hierarchies to typify and group entities to support description, targets and validation rules).

A goal-oriented DW requirement analysis method was used by Ren, Wang, & Lu (2018) to obtain an organizational and decision model. The organizational model captures high-level actors, their responsibilities, and relationships. In contrast, the decision model focuses on how the DW can support all decision-making necessities (associating facts and dimensions with the goal at different decision levels). In Bargui, Ben-Abdallah, & Feki (2011) ontologies are used to automate requirement elicitation also in goal-oriented DWs, by decomposing complex business goals into sub-goals, identifying indicators and generating analytical queries. Bellatreche, Khouri, Boukhari, & Bouchakri (2012) presented a solution where user requirements, represented by a goal-oriented model, are made persistent in the DW (through an ontology) to ensure traceability from the conceptual/ontological level to the physical level.

RAMEPs (Requirement Analysis Method for ETL Processes) is a goal-oriented method for ETL process design Taa, Abdullah, & Norwawi (2010); Taa, M.S, & Md Norwawi (2011); Ta'a & Abdullah (2011). DW requirements are collected and analyzed at the organizational, decisional, and developer (transformation needs) levels. User requirements semantics are obtained accordingly to an agreed-upon vocabulary of dimensional concepts (e.g., facts, dimensions), mitigating semantic heterogeneity problems.

5.2.2. Dimensional modeling

Ontologies are used to simplify dimensional design, discover business entities and their relationships, and find potential facts and dimensions from each data source. Thus, most works present the ontology as the primary source for the DW or as an intermediate layer between the source system and ETL. Some advantages include increased automation, flexibility, semantic information, and interoperability (between DWs). Ontologies are also used to solve heterogeneity problems. These advantages can impact subsequent phases, such as the ETL and exploration phases, especially when the DW is enriched with semantic information.

The dimensional model can be based on a requirement-driven ontology alone Bellatreche et al. (2012) or by comparing the requirements with a global/domain ontology (obtained by integrating ontologies or other heterogeneous data sources) (Chakiri, El Mohajir, & Assem, 2020; Khouri & Ladjel, 2010; Ren et al., 2018). Integration and data/semantic heterogeneity problems on traditional data sources (such as relational databases) can also be mitigated or resolved with the use of ontologies. One of the most commonly presented solutions is to obtain a global conceptual schema based on the source systems, along with the corresponding mapping for each data source (Aadil et al., 2016; Moreira, Cordeiro, Campos, & Borges, 2015; Tria, Lefons, & Tangorra, 2014). This domain ontology or vocabulary can then be used to find and uncover the facts, dimensions, and other dimensional entities (Romero & Abelló, 2010), including meaningful IDs (Abelló & Romero, 2010). Some works match multidimensional schemes and dimensions to ontological information to improve OLAP (Limongelli, Sciarrone, Starace, & Temperini, 2010) or data mining (Nimmagadda & Dreher, 2014; Nimmagadda, Nimmagadda, & Dreher, 2011) capabilities in the DW. Ontologies can also be used to facilitate DW schema evolution (Tanuska, Vlkovic, Vorstermans, & Verschelde, 2010).

Zaharie et al. (2011) present ontology-based dimensional design guidelines, where the REA ontology can be directly mapped to a star schema. He, Chen, Meng, & Liu (2011) introduce a conceptual modeling solution based on the BWW (Bunge-Wand-Weber) presentation model, including domain and property modeling, to better formalize users' needs and help solve heterogeneous problems. The quality, semantic expressiveness, and interoperability of conceptual models can be improved using ontological patterns (Amaral & Guizzardi, 2019). Automatic or semi-automatic methods that identify multidimensional concepts in OWL ontology sources are presented by Gulic (2013); Laadidi & Bahaj (2018); Liu & Iftikhar (2013). After finding these concepts, the multidimensional schema can be defined, together with the necessary mapping and transformations. Fernandes et al. (2012) present a similar solution, obtaining a fact table based on a concept map. Villanueva Chávez & Li (2011) extend this idea further and present an approach that generates a logical model, physical data models, and transformation rules based on extracted information from the ontology, obtaining a homogeneous solution.

5.2.3. ETL

Ontologies can enrich source data, provide mappings and increase ETL performance and efficiency. Data inconsistency, errors, and heterogeneity problems are also mentioned as motivation factors for integrating an ontology.

The design of the ETL process can be facilitated through the use of a domain ontology. Concepts, relationships are retrieved from the source schemas (Jiang, Cai, & Xu, 2010; Moreira et al., 2015; Villanueva Chávez & Li, 2011), making it possible for mappings to be automatically generated (since the target schema is based on the ontology, links between them are already in place). The RAMEPs method (Taa et al., 2010; Taa et al., 2011; Ta'a & Abdullah, 2011) automatically generates ETL processes by intersecting the goal-driven requirement ontology and data sources semantics, solving user requirements ambiguity and semantic heterogeneity problems. The representation of ETL requirements and process specifications in ontologies allows the creation of natural language reports, which can be used to communicate ETL process design choices, implementation, and maintenance (Simitsis et al., 2010).Furthermore, ontologies can also be used to enhance metadata from multimedia (Vanea & Potolea, 2011), or NoSQL Pticek & Vrdoljak (2018) databases, improving the integration process in these cases.

Ontologies are also used in ETL to support the configuration and instantiation of ETL patterns. By providing these regular and reusable patterns, Oliveira & Belo (2016) defend that data inconsistencies and errors can be mitigated. Ontologies can also be used to conceptualize data transformation processes and logical descriptions (Nimmagadda, Nimmagadda, & Dreher, 2010).

5.2.4. BI application design

The exploration phase (BI application design) can also take advantage of ontologies and their semantics. Ontologies, representing multidimensional models as OWL ontologies or RDF Data Cubes, are used as an intermediate layer between the user and the DW. This helps users semantically formalize queries and explore data, improving inference capabilities, knowledge extraction, and interoperability between DW/BI systems. Data mining/knowledge discovery processes are also facilitated and enhanced through the use of semantic OLAP frameworks.

Formal reasoning provided by ontologies, such as OWL-DL, can be used to validate multidimensional models and their summarizability (Prat, Akoka, & Comyn-Wattiau, 2012a; Prat, Megdiche, & Akoka, 2012b). Furthermore, ontologies allow multidimensional data to be distributed in the SW, improving interoperability with other systems. RDF Data Cube Vocabulary prepares multidimensional data to be published using RDF. The QB4OLAP extends the RDF Data Cube by introducing several OLAP functions (such as roll-up, slice, and dice). Matei, Chao, & Godwin (2015) propose the IGOLAP vocabulary to provide missing OLAP capabilities from QB4OLAP. In addition, relational implementations of data cubes were translated to RDF using an extended QB4OLAP vocabulary at both schema and instance level (Etcheverry, Vaisman, & Zimányi, 2014). Quamar et al. (2020) feature a "conversational interface" to support business analysis, exploiting typical BI analytical patterns and using natural language to translate input requests.

Ontologies have also proven to be very useful in supporting data mining and visualization. An ontology-based system can guide users in the mining process ("intelligent assistance"), helping in the selection and grouping of data, giving recommendations, and providing a way to detect semantic errors in the mining process. The efficiency and effectiveness of the mining process are improved, allowing users to find and extract useful knowledge in their data (Wu, Lin, Jiang, & Wu, 2011; Wu, Lin, & Wu, 2010). Ontologies can also be used to facilitate data interpretation and knowledge extraction, with ontologies supporting visual analysis, interactive explanation of data and enabling collaboration and knowledge sharing (chaining the "visual thinking") (Brahmi, 2019; Nimmagadda & Dreher, 2014; Nimmagadda et al., 2010; 2011). A semantic OLAP framework is presented by Neumayr, Anderlik, & Schrefl (2012); Neumayr, Schrefl, & Linner (2011), where ontologies are used as a conceptual layer between users and data, allowing ontology's multidimensional concepts to be mapped into SQL queries.

Limongelli et al. (2010) present an ontology-driven OLAP System where teachers use an ontology to find suitable Learning Objects from the Web. A similar framework was developed by Namnual, Nilsook, & Wannapiroon (2019), with the domain's concepts ontology being linked with existing DW concepts to support data visualization and analysis. Semantically enhanced metadata can help users to formulate queries and understand their results, helping with unforeseen queries Vanea & Potolea (2011). Aymoré Martins., C. Lustosa da Costa., & de Sousa Júnior. (2012) present a collaborative BI framework, where a global ontology is obtained by aligning and merging ontologies from different BI systems. Once this global ontology is obtained, heterogeneous concepts can be analyzed in a decentralized way, increasing interoperability and communications between DW/BI systems. Kurze, Gluchowski, & Bohringer (2010) also integrate different BI systems, using an extension of the BWW ontology to define core concepts of data warehousing.

5.2.5. Technical and DW architecture results

Other interesting works are related to the Technical Architecture design or Physical Design phases. Works include a DW reference model, with an ontology being used to describe DW architectures (Szwed, Komnata, & Dymek, 2015), support to Technical Architecture Design to improve shared resources allocation (Nicolicin-Georgescu, Benatier, Lehn, & Briand, 2010; Nicolicin-Georgescu, Benatier, Lehn, & Briand, 2010), and dimensional table partitions automation (Liu & Iftikhar, 2013). Villanueva Chávez & Li (2011) present an end-to-end process where logical and physical data models are automatically generated. ETL mappings between data sources and the models are defined based on the data meaning (using an ontology-based data model).

6. Discussion

This section analyzes the results and discusses the main challenges and outcomes of this review, then presenting its implications for practice and for the research agenda.

6.1. Synthesis of literature

As stated before, the main goal of this review is to understand how, where, and why ontologies are being used with DW/BI systems. Regarding the incorporation and integration of ontologies into DW/BI systems (RQ1), a large percentage of works use ontologies as intermediary support, either for data integration (or semantic integration of source data) or for exploration (exploratory OLAP). However, some researchers keep ontological data within the DW, usually in cases where the dimensional model was based on the ontology, to integrate semantics and increase DW interoperability and reusability.

In the literature, ontologies are used to support or improve DW/BI lifecycle tasks (RQ2). The primary use of ontologies in DW/BI systems is related to the task of dimensional modeling. Ontologies, due to their semantic interoperability and shared concepts, are used to streamline dimensional design, helping uncover business entities and their relations and finding potential facts and dimensions from each data source. After aligning each local ontology, knowledge from a domain ontology is extracted and transposed into a star schema or dimensional cube, with works such as Amaral & Guizzardi (2019); Gulic (2013); Romero & Abelló (2010) presenting similar methods. Requirement analysis is another task that can be largely influenced by the use of ontologies, supporting requirements elicitation, reconciliation of users' semantics and hopefully resolving requirements ambiguity. This knowledge is then used to create dimensional models that fulfill user requirements.

Ontologies are also used in ETL for supporting configuration and instantiation of ETL patterns (Oliveira & Belo, 2016). The ETL process is also facilitated when the model is designed via a domain ontology since mappings between source data, local ontology or schema, domain ontology and the dimensional domain are already in-place. This linkage allows ETL processes to be easily specified (Taa et al., 2011; Ta'a & Abdullah, 2011). Ontologies can also be used to enhance metadata from multimedia (Vanea & Potolea, 2011) or NoSQL (Pticek & Vrdoljak, 2018) databases, improving the integration process.

Exploration of the models (BI Applications Design) can also take advantage of ontologies and their semantics. Transforming dimensional models into OWL ontologies (Prat et al., 2012a) or RDF Data Cubes (Matei et al., 2015), creating a semantic OLAP framework, enables inference capability, knowledge extraction and, most importantly, interoperability. Ontologies can also improve data mining processes, facilitating knowledge discovery and improving data analysis (Wu et al., 2011). Other works include a DW reference model, with an ontology being used to describe DW architectures (Szwed et al., 2015), support to Technical Architecture Design to improve shared resources allocation (Nicolicin-Georgescu et al., 2010; Nicolicin-Georgescu et al., 2010), and dimensional table partitions automation (Liu & Iftikhar, 2013).

The main reasons given in the available literature for using SW techniques in DW/BI systems (RQ3) are diverse and generally take advantage of the semantics and inference provided by ontologies. Eliminating or solving the data/semantic heterogeneity problem, increasing interoperability, facilitating integration, and providing semantic content to both requirement and data analysis (better formalization) are some of the most indicated motivations.

6.2. Implications for practice

This SLR analyses the impact of ontologies on the design, development, and exploitation of DW/BI systems. Ontologies are mainly used in Requirement Analysis, Dimensional Modelling, ETL, and BI Application Design in various application fields, such as Natural Gas Distribution, Sales, and Education. OWL and its subtypes are the most popular languages for formalising ontologies, and in most of the analysed works the authors proposed the use of domain ontologies. Ontologies are used to eliminate problems of heterogeneity, facilitate data integration and provide semantics to requirements and data.

In practice, due to their semantics, reasoning, and interoperability, ontologies represent a new resource that traditional DW/BI systems should consider to facilitate the integration and analysis of structured data in the new IS paradigm. Dealing with web data and other unstructured or semi-structured data in a structured architecture represents a challenge in terms of volume, variety, and velocity, as well as how to connect and understand the meaning of different types of data. The impact of ontologies here is evident as it enables the formalisation of knowledge, meaning that decisions, and organisational or practical knowledge related to the system can be materialised and shared within or outside the organisation, providing a connection point between business users, data scientists, and different IS.

In short, ontologies support, simplify and help automate design and development tasks and processes in DW/BI systems. Ontologies are, however, not typically used for data enrichment purposes, such as adding attributes to existing dimensions. Dimensional models are created based on ontologies to take advantage of OLAP-style analysis, with all dimensions and facts being extracted from an ontology, or exported to an ontology to enable inference and interoperability (e.g., RDF Cube). System interoperability between different DW/BI systems was demonstrated. The integration of unstructured data in DW/BI systems was not within the scope of this review but could have been found as part of ontology-based solutions. However, authors did not present this as a motivation for their works.

The use of ontologies in DW/BI systems enables the elicitation of higher quality requirements, as DW/BI developers are able to improve communication and reduce misunderstandings between customers or stakeholders. Using these techniques also helps to reduce costs and time for schema designers and data engineers, particularly in cases where ontologies are used to integrate different sources, since mappings between source and target are easier to obtain.

From an application perspective, the decision-making process can benefit from the added semantics and inference. The representation of business knowledge and its reasoning allows the business user to be guided during data analysis. Knowledge bases can assist in query formulation, give additional context to data analysis, or ensure the novelty of new relationships (e.g., ensuring that data mining results are relevant to decision-making). Industries or domains already taking advantage of DW/BI systems can also benefit from ontology integration, especially industries within highly complex domains such as healthcare (Jiang et al., 2010; Neumayr et al., 2012; Neumayr et al., 2011; Nimmagadda et al., 2011; Quamar et al., 2020; Ren et al., 2018) or academic/education (Amaral & Guizzardi, 2019; Limongelli et al., 2010; Namnual et al., 2019; Taa et al., 2010; Tanuska et al., 2010).

Finally, while data warehousing as an integrated repository is still a focus of research, the relationship between structured data and the semantic web is being neglected by researchers, as shown in Fig. 6. However, the increasing complexity of (big) data, relationships and business domains will lead to increasingly complex business analysis and data mining. Structured data can be enriched, through a semantic layer, to cope with this change and enable new types of analysis over complex domains.

6.3. Limitations

The main limitation of this paper is related to the availability of academic research regarding the integration of SW techniques into traditional DW/BI systems, as discussed earlier. Most of the peer-reviewed research found in this SLR was published in domain-related conferences rather than academic journals.

Similar (or identical) keywords are simultaneously used in research related to knowledge-based DSS and DW/BI systems, which can lead to confusion when searching for articles related to a single type of system. Apart from the different main components, DW/BI and knowledge-based DSS systems are similar in terms of tasks and usage. Dimensional modeling, ETL processes, and exploration techniques (e.g. OLAP cubes) are addressed in both DW/BI and knowledge-based IS research. While this did not represent a problem *per se*, a substantial number of papers were rejected due to this overlap and, if not fully made explicit by the authors, may create confusion when analysing the original research. This misunderstanding usually results from a lack of clarification about the use of ontologies. Although most authors properly explain their work, some definitions can lead to misunderstandings. For example, 'ontology-based DW" can mean either that the design of the DW was based on an ontology (but the information is stored in the traditional relational star schema or multidimensional cube) or that the knowledge of the system is stored in an ontology (knowledge-based IS). Ontology information (such as the ontology language) was also not available or explicit in all papers.

On the other hand, some works misemploy key terms or denominations. For example, the term ontology is used to describe a Unified Modeling Language (UML) class diagram. While sometimes UML can be used to illustrate an ontology, a class diagram with no semantic relations should not be defined as an ontology. Another example is an overlap between development phases, with some authors intertwining the phases of requirement analysis and dimensional modeling (when in fact, business requirements should be an input to the dimensional modeling task).

6.4. Research agenda

This section presents some possible research paths not fully explored by the literature in this SLR, which could lead to new interesting research questions (see Table 6).

Different approaches can be used during requirement analysis and DW design. It has been shown that ontologies support data-driven approaches, in which source and operational systems are analyzed to derive analytical models, and goal-driven methods, which develop the DW to directly answer business queries and monitor goals (usually translated into SQL or SPARQL queries in the ontology). However, there was a noticeable lack of research on process-driven approaches, which focus on identifying and analyzing the business processes within the organization (Kimball's approach). Ontologies could also be used to support or validate existing process-driven DW design methodologies, such as BEAM - Business Event Analysis & Modelling (Corr & Stagnitto, 2011) (e.g., validate data stories, which are made-up examples of business events, in terms of detail and completeness).

Another possible unexplored opportunity is the use of ontologies for data enrichment in DW. Most of the reviewed works provide methods for designing dimensional models or analyze the dimensional models through an ontology. The works that use ontologies for the enrichment of dimensional models are rare. The idea here is to use an ontology as an external source to generate new attributes related to an existing business entity, e.g., to relate domain information otherwise not available in the DW source systems.

Furthermore, research regarding ontology-supported exploration usually uses a semantic representation of entities already existing in the DW and other dimensional data. Both for exploration through an RDF cube and ontology-supported BI applications, semantic representation allows for extended rules or new conceptual relationships, such as new types of hierarchies. However, most of the exploration-related works analyzed in this SLR use a domain ontology containing the same or slightly enriched information as that available in the DW. BI applications could use ontologies containing knowledge about different domains to enrich and support the exploration phase, taking advantage of the ontologies' interoperability. As DSS, DW/BI systems can be used to measure, monitor and evaluate business performance and strategy. Strategic information is not typically stored in the DW/BI system, especially in data- and process-driven DWs. Ontologies may be a useful tool for modelling the strategy and strategic information. This knowledge could then be used in a BI application to guide and support information retrieval and analysis. This integration between operational data and strategy is of utmost importance to ensure proper business performance management (Kaplan & Norton, 2008; Turban et al., 2010).

From a DSS perspective, there is a clear interest on creating an integrated ecosystem that enables the analysis of both structured and unstructured data (Inmon et al., 2021). As (Ravat & Zhao, 2019) state, whether DW coexists or is part of a DL architecture is still a matter of debate. However, information should always flow between the two, and metadata management systems should be in place to allow users to find the relevant data and cross-reference information as transparently and directly as possible. Ontologies could provide a missing connection point between DW data and other data types that are inside or outside the system/architecture. This interoperability could, for example, be ensured through the metadata representation of each repository. Mishra et al. (2021a) presented a predictive analysis based on structured data, which, although not in a dimensional model, clearly represented context (country, continent, year) and facts (number of arrivals). In another paper, the same authors presented an unstructured data analysis to obtain a sentiment analysis in the same tourism domain (Mishra, Urolagin, Jothi, Neogi, & Nawaz, 2021b). These works analyse, following different solutions, the impact of the COVID-19 pandemic on tourism and tourists. However, they are analysed separately. More research should be done to allow the data, results and findings to be combined and analysed as a whole, assuming that the context provided is the same,

Table 6
Research Agenda Summary.

Research Topic	Research Question Proposal
Process-Driven Semantic-Aware Requirements	How might ontologies aid in the elicitation and analysis of process-driven requirements?
Dimensional Enrichment	Could existing dimensional entities be enriched using ontologies as a source?
Semantic-Supported Business Analysis	How can ontologies support the analysis and exploration of an existing DW/BI system?
Semantic-Integrated DSS	Can structured and unstructured data analysis and exploration be linked using SW techniques?
Project Planning and Management	How can ontologies be used to support DW/BI Program/Project planning and management tasks?

i.e., the data or information has the same meaning in both systems/ repositories.

Finally, it would be interesting to apply ontologies in the other tasks of DW/BI lifecycle, such as Program/Project planning and management. These tasks were not analysed in this SLR due to the lack of research on the subject, however, ontologies and their semantics might be used to support stakeholder communication or validate project planning.

7. Conclusion

The SLR described in this paper aims to obtain an overview of the use of ontologies in DW/BI systems. The existing literature is surveyed regarding how, where and why ontologies are being used to improve the analytical capabilities of DW/BI systems or to simplify processes within the DW/BI lifecycle.

Despite the importance and emphasis given to the analysis of unstructured data in IS, researchers and organizations understand the business value that structured data offers for performance management. For this reason, DW/BI systems and their associated techniques are still relevant to obtain KPIs and other metrics quickly and easily, as business decision-makers expect. With the emergence of the Semantic Web, the use of ontologies has become increasingly common in IS due to their semantic, formalization, and inference qualities. The primary motivation of this work is to study if and how these ontologies can be used to enrich DW/BI, improve interoperability between IS or facilitate the design, development, and exploration of the DW/BI system.

For this purpose, research papers were collected from four search engines, with keywords related to DW/BI systems and ontologies. These works were classified to obtain information about the field of each case study, the motivation of its authors, as well as the SW techniques and DW/BI development tasks where they are used. Ontologies (usually domain- and task-specific) are mainly defined using the SW standard OWL, to support multiple DW/BI tasks, such as Dimensional Modeling, Requirement Analysis, ETL, and BI Application Design. Several reviewed papers use ontologies as an intermediary support for data integration and exploration. Authors present a variety of motivations for ontology-driven solutions in DW/BI, such as eliminating or solving data heterogeneity/semantics problems, increasing interoperability, facilitating integration, or providing semantic content for requirement and data analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aadil, B., Wakrime, A., Kzaz, L., & Sekkaki, A. (2016). Automating data warehouse design using ontology. In Proceedings of the international conference on electrical and information technologies (ICEIT) (pp. 42–48). 10.1109/EITech.2016.7519618

Abelló, A., & Romero, O. (2010). Using ontologies to discover fact IDs. In *Proceedings of the ACM 13th international workshop on data warehousing and olap* (pp. 3–10). New York, USA: Association for Computing Machinery. 10.1145/1871940.1871944.

Abelló, A., Romero, O., Pedersen, T. B., Berlanga, R., Nebot, V., Aramburu, M. J., & Simitsis, A. (2014). Using semantic web technologies for exploratory OLAP: A survey. *IEEE Transactions on Knowledge and Data Engineering, 27*(2), 571–588. 10.1109/TKDE.2014.2330822.

Adamson, C. (2010). *Star schema the complete reference*. McGraw Hill Professional.

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights, 1*(1), 100004. 10.1016/j.jjimei.2020.100004.

Altuntas, S., Selim, S., & Altuntas, F. (2022). A hierarchical clustering based panel data approach: A case study of regional incentives. *International Journal of Information Management Data Insights, 2*(2), 100098. 10.1016/j.jjimei.2022.100098.

Amaral, G., & Guizzardi, G. (2019). On the application of ontological patterns for conceptual modeling in multidimensional models. In *Advances in databases and information systems* (pp. 215–231). Cham: Springer International Publishing. 10.1007/978-3-030-28730-6_14.

Arjun, R., Kuanr, A., & Suprabha, K. (2021). Developing banking intelligence in emerging markets: Systematic review and agenda. *International Journal of Information Management Data Insights, 1*(2), 100026. 10.1016/j.jjimei.2021.100026.

Aymoré Martins, V., C. Lustosa da Costa, J. P., & de Sousa Júnior, R. T. (2012). Architecture of a collaborative business intelligence environment based on an ontology repository and distributed data services. In *Proceedings of the international conference on knowledge management and information sharing - kMIS, (ic3k 2012)* (pp. 99–106). SciTePress. 10.5220/0004107000990106. INSTICC

Bargui, F., Ben-Abdallah, H., & Feki, J. (2011). A decision-making ontology for analytical requirements elicitation. In *Computer and information sciences ii* (pp. 495–501). Springer. 10.1007/978-1-4471-2155-8_63.

Bellatreche, L., Khouri, S., Boukhari, I., & Bouchakri, R. (2012). Using ontologies and requirements for constructing and optimizing data warehouses. In Mipro 2012 - 35th international convention on information and communication technology, electronics and microelectronics - proceedings (pp. 1568–1573).

Brahmi, H. (2019). Ontology enhanced mining of multidimensional association rules from data cubes. In *Proceedings of the international conference on information networking (ICOIN)* (pp. 159–164). 10.1109/ICOIN.2019.8718172.

Budgen, D., & Brereton, P. (2006). Performing systematic literature reviews in software engineering. In Proceedings of the international conference on software engineering (p. 1051–1052). New York, NY, USA: Association for Computing Machinery. 10.1145/1134285.1134500

Burton, B., Geishecker, L., Schelegel, K., Hostmann, B., Austin, T., Herschel, G., Soejarto, A., & Rayner, N. (2006). Business intelligence focus shifts from tactical to strategic. Retrieved from Gartner database (G00139352),.

Cavalheiro, J., & Carreira, P. (2016). A multidimensional data model design for building energy management. *Advanced Engineering Informatics, 30*(4), 619–632. 10.1016/j.aei.2016.08.001.

Chakiri, H., El Mohajir, M., & Assem, N. (2020). A data warehouse hybrid design framework using domain ontologies for local good-governance assessment. *Transforming Government: People, Process and Policy*. 10.1108/TG-04-2019-0025.

Corr, L., & Stagnitto, J. (2011). *Agile data warehouse design: Collaborative dimensional modeling, from whiteboard to star schema*. DecisionOne Consulting.

Ereth, J., & Eckerson, W. (2018). *Ai: The new Bi - how algorithms are transforming business intelligence and analytics*. Data Strategy Insider. https://www.ibm.com/downloads/cas/M7VMLOPY

Etcheverry, L., Vaisman, A., & Zimányi, E. (2014). Modeling and querying data warehouses on the semantic web using QB4OLAP. *Lecture Notes in Computer Science, 8646 LNCS*, 45–56. 10.1007/978-3-319-10160-6_5.

Fernandes, A. A., Amaro, L. C., Da Costa, J. P. C. L., Serrano, A. M. R., Martins, V. A., & Júnior, R. T. D. (2012). Construction of ontologies by using concept maps: A study case of business intelligence for the federal property department. In Proceedings of the fifth international conference on business intelligence and financial engineering (pp. 84–88). 10.1109/BIFE.2012.26

Gandon, F. (2018). A survey of the first 20 years of research on semantic web and linked data. *Ingénierie Des Systèmes d'Information, 23*, 11–38. 10.3166/isi.23.3-4.11-38.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220. 10.1006/knac.1993.1008.

Gulic, M. (2013). Transformation of owl ontology sources into data warehouse. Proceedings of the 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), (pp. 1143–1148).

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management, 42*, 78–89. 10.1016/j.ijinfomgt.2018.06.005.

He, L., Chen, Y., Meng, N., & Liu, L. (2011). An ontology-based conceptual modeling method for data warehouse. In *Icm 2011- proceedings: vol. 4* (pp. 130–133). 10.1109/ICM.2011.171.

Hitzler, P. (2021). A review of the semantic web field. *Communications of the ACM, 64*(2), 76–83. 10.1145/3397512.

Hughes, R. (2012). *Agile data warehousing project management: Business intelligence systems using scrum*. Newnes.

Hussain, A. A., Al-Turjman, F., & Sah, M. (2020). Semantic web and business intelligence in big-data and cloud computing era. In *Proceedings of the third international conference on smart city applications* (pp. 1418–1432). Springer. 10.1007/978-3-030-66840-2_107.

Inmon, B. (2016). *Data lake architecture: Designing the data lake and avoiding the garbage dump*. Technics publications.

Inmon, B., Levins, M., & Srivastava, R. (2021). *Building the data lakehouse*. Technics Publications.

Jiang, L., Cai, H., & Xu, B. (2010). A domain ontology approach in the ETL process of data warehousing. In *Proceedings of the IEEE 7th international conference on e-business engineering* (pp. 30–35). 10.1109/ICEBE.2010.36.

Kaplan, R. S., & Norton, D. P. (2008). *The execution premium: Linking strategy to operations for competitive advantage*. Harvard Business Press.

Khouri, S., & Ladjel, B. (2010). A methodology and tool for conceptual designing a data warehouse from ontology-based sources. In Proceedings of the ACM 13th international workshop on data warehousing and olap (p. 19–24). New York, NY, USA: Association for Computing Machinery. 10.1145/1871940.1871946

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons.

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). *The data warehouse lifecycle toolkit*. John Wiley & Sons.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), 100008. 10.1016/j.jjimei.2021.100008.

Kurze, C., Gluchowski, P., & Bohringer, M. (2010). Towards an ontology of multidimensional data structures for analytical purposes. In Proceedings of the annual Hawaii international conference on system sciences (pp. 1–10). 10.1109/HICSS.2010.485

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), 100017. 10.1016/j.jjimei.2021.100017.

Laadidi, Y., & Bahaj, M. (2018). Simplification of owl ontology sources for data warehousing. In *Proceedings of the ACM international conference proceeding series* (pp. 77–81). 10.1145/3178461.3178483.

Laborie, S., Ravat, F., Song, J., & Teste, O. (2015). Combining business intelligence with semantic web: Overview and challenges. IFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID),.

Limongelli, C., Sciarrone, F., Starace, P., & Temperini, M. (2010). An ontology-driven OLAP system to help teachers in the analysis of web learning object repositories. *Information Systems Management, 27*(3), 198–206. 10.1080/10580530.2010.493810.

Liu, X., & Iftikhar, N. (2013). Ontology-based big dimension modeling in data warehouse schema design. *Lecture Notes in Business Information Processing, 157*, 75–87. 10.1007/978-3-642-38366-3_7.

Lukasiewicz, T. (2008). Expressive probabilistic description logics. *Artificial Intelligence, 172*(6), 852–883. 10.1016/j.artint.2007.10.017.

Lukić, J., Radenković, M., Despotović-Zrakić, M., Labus, A., & Bogdanović, Z. (2016). A hybrid approach to building a multi-dimensional business intelligence system for electricity grid operators. *Utilities Policy, 41*, 95–106. 10.1016/j.jup.2016.06.010.

Matei, A., Chao, K.-M., & Godwin, N. (2015). OLAP For multidimensional semantic web databases. *Lecture Notes in Business Information Processing, 206*, 81–96. 10.1007/978-3-662-46839-5_6.

Mishra, R. K., Urolagin, S., Jothi, J., Nawaz, N., & Haywantee, R. (2021a). Machine learning based forecasting systems for worldwide international tourists arrival,. 10.14569/IJACSA.2021.0121107

Mishra, R. K., Urolagin, S., Jothi, J., Neogi, A., & Nawaz, N. (2021b). Deep learning-based sentiment analysis and topic modeling on tourism during COVID-19 pandemic. *Frontiers in Computer Science, 3*(10.3389).

Mishra, R. K., Urolagin, S., & Jothi, J. A. A. (2020). Sentiment analysis for poi recommender systems. In *Proceedings of the seventh international conference on information technology trends (itt)* (pp. 174–179). IEEE. 10.1109/ITT51279.2020.9320885.

Moreira, J., Cordeiro, K., Campos, M. L. M., & Borges, M. (2015). Hybrid multidimensional design for heterogeneous data supported by ontological analysis: An application case in the Brazilian electric system operation. In Ceur workshop proceedings (pp. 72–77). (vol. 1330).

Namnual, T., Nilsook, P., & Wannapiroon, P. (2019). System architecture of data warehousing with ontologies to enhance digital entrepreneurs' competencies for higher education. *IJIET, 9*(6), 414–418. 10.18178/ijiet.2019.9.6.1237.

Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of indian farmers' protest using twitter data. *International Journal of Information Management Data Insights, 1*(2), 100019. 10.1016/j.jjimei.2021.100019.

Neumayr, B., Anderlik, S., & Schrefl, M. (2012). Towards ontology-based OLAP: Datalog-based reasoning over multidimensional ontologies. In *Proceedings of the international conference on information and knowledge management, proceedings* (pp. 41–48). 10.1145/2390045.2390053.

Neumayr, B., Schrefl, M., & Linner, K. (2011). Semantic cockpit: An ontology-driven, interactive business intelligence tool for comparative data analysis. In O. De Troyer, C. Bauzer Medeiros, R. Billen, P. Hallot, A. Simitsis, & H. Van Mingroot (Eds.), *Advances in conceptual modeling. Recent developments and new directions* (pp. 55–64). Berlin, Heidelberg: Springer Berlin Heidelberg. 10.1007/978-3-642-24574-9_9.

Nicolicin-Georgescu, V., Benatier, V., Lehn, R., & Briand, H. (2010). Ontology-based autonomic computing for decision support systems management: Shared ressources allocation between groups of data warehouses. In *Proceedings of the CTRQ 2010* (pp. 233–236). 10.1109/CTRQ.2010.46.

Nicolicin-Georgescu, V., Benatier, V., Lehn, R., & Briand, H. (2010). Ontology-based autonomic computing for resource sharing between data warehouses in decision support systems. In Proceedings of the ICEIS 2010 (pp. 199–206). 10.1109/CTRQ.2010.46

Nimmagadda, S. L., & Dreher, H. V. (2014). Multidimensional ontology modelling – A robust methodology for managing complex and heterogeneous petroleum digital ecosystems. In *Proceedings of the 12th IEEE international conference on industrial informatics (INDIN)* (pp. 740–747). 10.1109/INDIN.2014.6945605.

Nimmagadda, S. L., Nimmagadda, S. K., & Dreher, H. (2010). Multidimensional ontology modeling of human digital ecosystems affected by social behavioural data patterns. In *Proceedings of the 4th IEEE international conference on digital ecosystems and technologies* (pp. 498–503). 10.1109/DEST.2010.5610601.

Nimmagadda, S. L., Nimmagadda, S. K., & Dreher, H. (2011). Multidimensional data warehousing and mining of diabetes and food-domain ontologies for e-health. In *Proceedings of the 9th IEEE international conference on industrial informatics* (pp. 682–687). 10.1109/INDIN.2011.6034973.

Oliveira, B., & Belo, O. (2016). An ontology for describing etl patterns behavior. In Data 2016 - proceedings (pp. 102–109). 10.5220/0005974001020109

Pan, J. (2009). Resource description framework. In *Handbook on ontologies* (pp. 71–90). 10.1007/978-3-540-92673-3_3.

Power, D. J. (2009). *Decision support basics*. Business Expert Press.

Prat, N., Akoka, J., & Comyn-Wattiau, I. (2012a). Transforming multidimensional models into OWL-DL ontologies. In Proceedings of the international conference on research challenges in information science. 10.1109/RCIS.2012.6240451

Prat, N., Megdiche, I., & Akoka, J. (2012b). Multidimensional models meet the semantic web: Defining and reasoning on OWL-DL ontologies for OLAP. In Proceedings of the international conference on information and knowledge management (pp. 17–24). 10.1145/2390045.2390049

Pticek, M., & Vrdoljak, B. (2018). Semantic web technologies and big data warehousing. In *Proceedings of the 41st international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1214–1219). 10.23919/MIPRO.2018.8400220.

Quamar, A., Ozcan, F., Miller, D., Moore, R. J., Niehus, R., & Kreulen, J. (2020). Conversational BI: An ontology-Driven conversation system for business intelligence applications. *Proceedings of the VLDB Endowment, 13*(12), 3369–3381. 10.14778/3415478.3415557.

Ravat, F., & Zhao, Y. (2019). Data lakes: Trends and perspectives. In *Database and expert systems applications* (pp. 304–313). Springer International Publishing. 10.1007/978-3-030-27615-7_23.

Rawat, S., Rawat, A., Kumar, D., & Sabitha, A. S. (2021). Application of machine learning and data visualization techniques for decision support in the insurance sector. *International Journal of Information Management Data Insights, 1*(2), 100012. 10.1016/j.jjimei.2021.100012.

Ren, S., Wang, T., & Lu, X. (2018). Dimensional modeling of medical data warehouse based on ontology. In *Proceedings of the IEEE 3rd international conference on big data analysis (ICBDA)* (pp. 144–149). 10.1109/ICBDA.2018.8367666.

Ristoski, P., & Paulheim, H. (2016). Semantic web in data mining and knowledge discovery: A comprehensive survey. *Journal of Web Semantics, 36*, 1–22. 10.1016/j.websem.2016.01.001.

Romero, O., & Abelló, A. (2010). A framework for multidimensional design of data warehouses from ontologies. *Data & Knowledge Engineering, 69*(11), 1138–1157. 10.1016/j.datak.2010.07.007.

Roussey, C., Pinet, F., Kang, M. A., & Corcho, O. (2011). An introduction to ontologies and ontology engineering. In *Ontologies in urban development projects* (pp. 9–38). London: Springer London. 10.1007/978-0-85729-724-2_2.

Sawadogo, P., & Darmont, J. (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems, 56*(1), 97–120. 10.1007/s10844-020-00608-7.

Sharda, R., Delen, D., Turban, E., Aronson, J., & Liang, T. (2015). *Business Intelligence and Analytics* (10th). Pearson Edition Limited.

Siau, K., Woo, C., Story, V. C., Chiang, R. H., Chua, C. E., & Beard, J. W. (2022). Information systems analysis and design: Past revolutions, present challenges, and future research directions. *Communications of the Association for Information Systems, 50*(1), 33. 10.17705/1CAIS.05037.

Silva, R., & Neiva, F. (2016). Systematic literature review in computer science - a practical guide,. 10.13140/RG.2.2.35453.87524

Simitsis, A., Skoutas, D., & Castellanos, M. (2010). Representation of conceptual etl designs in natural language using semantic web technology. *Data and Knowledge Engineering, 69*(1), 96–115. 10.1016/j.datak.2009.08.009.

Singh, K. N., Devi, S. D., Devi, H. M., & Mahanta, A. K. (2022). A novel approach for dimension reduction using word embedding: An enhanced text classification approach. *International Journal of Information Management Data Insights, 2*(1), 100061. 10.1016/j.jjimei.2022.100061.

Smith, B. (2003). Ontology. In *Blackwell guide to the philosophy of computing and information* (pp. 155–166).

Sommerville, I. (2011). *Software engineering, 9th edition*. Pearson Education India.

Stephan, G. s., Pascal, H. s., & Andreas, A. s. (2007). Knowledge representation and ontologies. In *Semantic web services: Concepts, technologies, and applications* (pp. 51–105). 10.1007/3-540-70894-4_3.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering, 25*(1–2), 161–197. 10.1016/S0169-023X(97)00056-6.

Szwed, P., Komnata, W., & Dymek, D. (2015). Dwarm: An ontology of data warehouse architecture reference model. *Communications in Computer and Information Science, 521*, 222–232. 10.1007/978-3-319-18422-7_20.

Taa, A., Abdullah, M., & Norwawi, N. (2010). Rameps: A goal-ontology approach to analyse the requirements for data warehouse systems. *WSEAS Transactions on Information Science and Applications, 7*(2), 295–309.

Taa, A., M. S, A. A., & Md Norwawi, N. (2011). Goal-ontology ETL processes specification. *Journal of Information and Communication Technology, 10*, 15–43. 10.32890/jict.10.2011.8107.

Tanuska, P., Vlkovic, O., Vorstermans, A., & Verschelde, W. (2010). The proposal of ontology as a part of university data warehouse. In *2010 2nd international conference on education technology and computer: vol. 3* (pp. 21–24). 10.1109/ICETC.2010.5529608.

Ta'a, A., & Abdullah, M. S. (2011). Goal-ontology approach for modeling and designing ETL processes. *Procedia Computer Science, 3*, 942–948. 10.1016/j.procs.2010.12.154.

Tria, F., Lefons, E., & Tangorra, F. (2014). Ontological approach to data warehouse source integration. *Lecture Notes in Electrical Engineering, 264 LNEE*, 251–259. 10.1007/978-3-319-01604-7_25.

Turban, E., Sharda, R., & Delen, D. (2010). *Decision support and business intelligence systems* (9th). USA: Pearson Education, Inc..

Vanea, A., & Potolea, R. (2011). Semantically enhancing multimedia data warehouses - using ontologies as part of the metadata. In *Proceedings of the 13th international conference on enterprise information systems - volume 1: ICEIS* (pp. 163–168). SciTePress. 10.5220/0003434701630168. INSTICC

Villanueva Chávez, J., & Li, X. (2011). Ontology based ETL process for creation of ontological data warehouse. In *Proceedings of the 8th international conference on electrical engineering, computing science and automatic control* (pp. 1–6). 10.1109/ICEEE.2011.6106642.

Watson, H., & Wixom, B. (2007). The current state of business intelligence. *Computer, 40*, 96–99. 10.1109/MC.2007.331.

Wisnubhadra, I., Baharin, S. S. K., & Herman, N. S. (2019). Modeling and querying spatiotemporal multidimensional data on semantic web: A survey. *Journal of Theoretical and Applied Information Technology, 97*(23), 3608–3633. 10.1007/978-3-319-31676-5_1.

Wu, C.-A., Lin, W.-Y., Jiang, C.-L., & Wu, C.-C. (2011). Toward intelligent data warehouse mining: An ontology-integrated approach for multi-dimensional association mining. *Expert Systems with Applications, 38*(9), 11011–11023. 10.1016/j.eswa.2011.02.144.

Wu, C.-A., Lin, W.-Y., & Wu, C.-C. (2010). An active multidimensional association mining framework with user preference ontology. *International Journal of Fuzzy Systems, 12*, 125–135. 10.30000/IJFS.201006.0004.

Young, Z., & Steele, R. (2022). Empirical evaluation of performance degradation of machine learning-based predictive models–a case study in healthcare information systems. *International Journal of Information Management Data Insights, 2*(1), 100070. 10.1016/j.jjimei.2022.100070.

Zaharie, D., Pugna, I., & Radulescu, C. (2011). An ontology-based conceptual design of a data warehouse. *Economic Computation and Economic Cybernetics Studies and Research, 2*, 57–66.

CHAPTER 3

Journal Article 2

This chapter details the design and development of artifact Road Structures Ontology (artifact #2), denominated Engineering Structures Ontology. This artifact was developed to encode the shared conceptualization provided by the CoDEC Data Dictionary. The ontology is evaluated, validated, and demonstrated as a foundation for data exchange between BIM and AMS systems, using data from three different pilot projects.

As shown in Figure 3.1, this publication (JA2) also evaluates the artifacts designed and developed during the DSRM's Iter.2, namely the Integration Framework (version 2 of artifact #1) and API Services (version 2 of artifact #4). These artifacts, together with the domain-specific ontology, are used within the context of the CoDEC project (see Section 1.3.1) to link operational data with BIM environments, facilitating the decision-making process.

Article details:

- **Title**: Ontology-based BIM-AMS Integration in European Highways;
- **DOI**: https://doi.org/10.1016/j.iswa.2024.200366;
- **Date**: 2024;
- **Journal**: Intelligent Systems with Applications;
- **Publisher**: Elsevier.

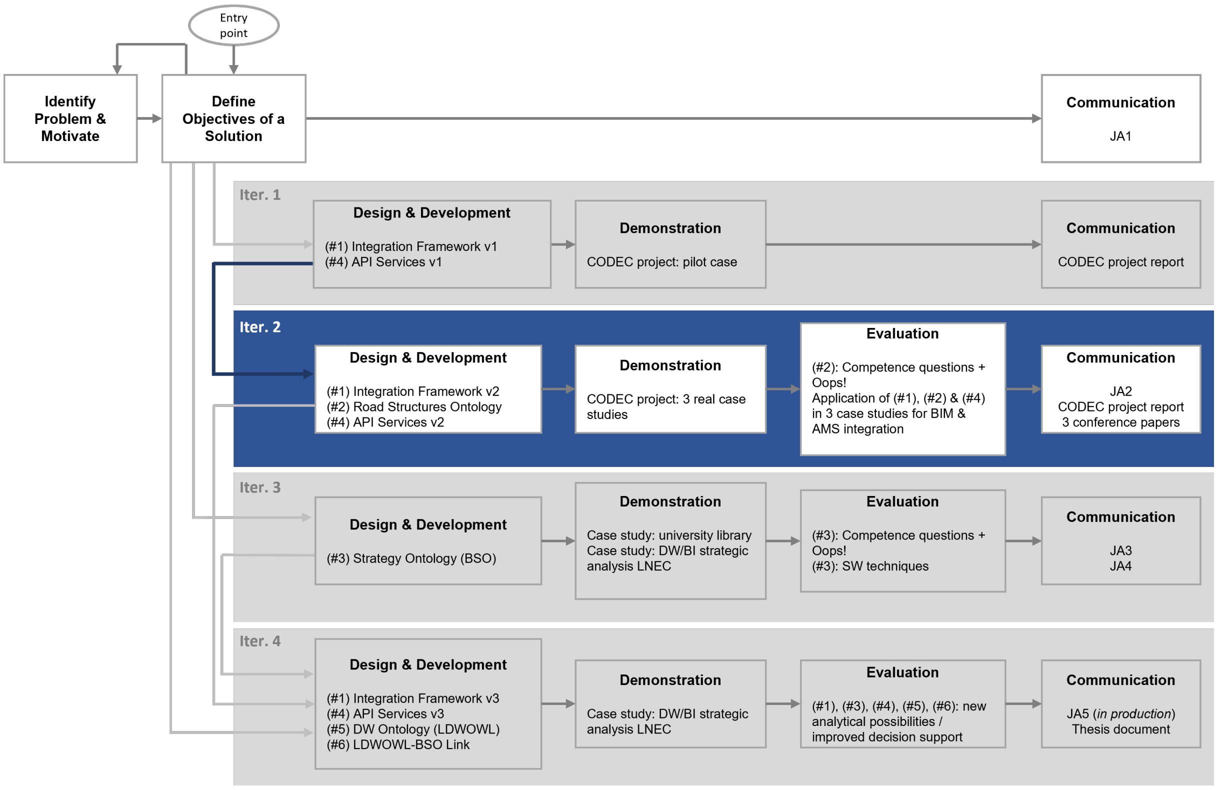


FIGURE 3.1. DSRM's JA2 Communication.

Intelligent Systems with Applications 22 (2024) 200366

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Ontology-based BIM-AMS integration in European Highways

António Lorvão Antunes [a,b,*], José Barateiro [e], Vânia Marecos [a], Jelena Petrović [c], Elsa Cardoso [b,d]

[a] LNEC - National Laboratory for Civil Engineering, Av. do Brasil 101, Lisbon, 1700-075, Portugal
[b] Department of Information Sciences and Technology, ISCTE - Instituto Universitário de Lisboa, Av. Forças Armadas, Lisbon, 1649-026, Portugal
[c] BEXEL Consulting, Ljubljana, Slovenia
[d] CIES-Iscte, Av. Forças Armadas, Lisbon, 1649-026, Portugal
[e] Faculdade de Ciência e Tecnologia, Universidade do Algarve, Campus de Gambelas, Faro, 8005-139, Portugal

ARTICLE INFO

Keywords:
Building Information Modeling (BIM)
Decision support
Risk and condition data
Ontology development
Ontology validation

ABSTRACT

BIM tools enable decision-making during the lifecycle of engineering structures, such as bridges, tunnels, and roads. National Road Authorities use Asset Management Systems (AMS) to manage and monitor operational information of assets from European Highways, including access to sensor and inspection data. Interoperability between BIM and AMS systems is vital for a timely and effective decision-making process during the operational phase of these assets. The European project Connected Data for Effective Collaboration (CoDEC) designed a framework to support the connections between AMS and BIM platforms, using linked data principles. The CoDEC Data Dictionary was developed to provide standard data formats for AMS used by European NRA. This paper presents the design and development of an Engineering Structures ontology used to encode the shared conceptualization provided by the CoDEC Data Dictionary. The ontology is evaluated, validated, and demonstrated as a base for data exchange between BIM and AMS.

1. Introduction

Building Information Modeling (BIM) is defined in ISO 19650-1:2018 (2018) as the "use of a shared digital representation of a built asset to facilitate design, construction and operation processes to form a reliable basis for decisions". Physical infrastructures, such as buildings, bridges and roads, can be modeled and managed across the whole asset lifecycle using BIM, together with necessary functional characteristics needed for decision making. The 3D visualization provided by BIM tools allows stakeholders to collaborate, share and exchange information, which is especially useful for decision support during the design, planning and construction phases. However, the use of BIM in transport infrastructures is still far from its wide application in the building industry, mainly due to the fact that vertical structures (buildings) have different operations, components and techniques in comparison to horizontal constructions (e.g., bridges, roads, tunnels) (Costin et al., 2018). Recent works show that the application of BIM in the transportation industry is slowly increasing and can be helpful along the lifecycle of the structures from the most common to the more complex activities (Biancardo et al., 2020, D'Amico et al., 2020, Koch et al., 2014). Furthermore, despite its potential application in all phases of the infrastructure life cycle, BIM use during the operational and management (O&M) phase is currently limited (Wijeratne et al., 2024).

National Road Authorities (NRA) in Europe have invested in Asset Management Systems (AMS) to ensure management, maintenance and structural safety during the operational phase of Engineering Structures in European Highways. These systems contain asset operational information such as sensor data and inspection results, usually stored in various formats. Ideally, information should be shared between BIM models and AMS so that more efficient and informed decisions can be taken during the operational phase of these engineering structures (in either system). While there are standards for BIM data, such as the Industry Foundation Classes – IFC (ISO 16739-1:2018 (2018)), these are not focused on operation phases or AMS integration. Due to the increasing number of solutions for asset monitoring (sensor technology and Internet-of-Things), the interoperability between these systems is vital for timely decision-making in an integrated environment. Linking 3D model data with asset management data allows access to an integrated view of information, reduces errors, and saves time and costs, while also enhancing compliance, customer satisfaction, safety, and risk mitigation during the operational phase (Wijeratne et al., 2024).

* Corresponding author.
E-mail address: alfas@iscte-iul.pt (A. Lorvão Antunes).

https://doi.org/10.1016/j.iswa.2024.200366
Received 23 November 2023; Received in revised form 15 March 2024; Accepted 1 April 2024
Available online 5 April 2024

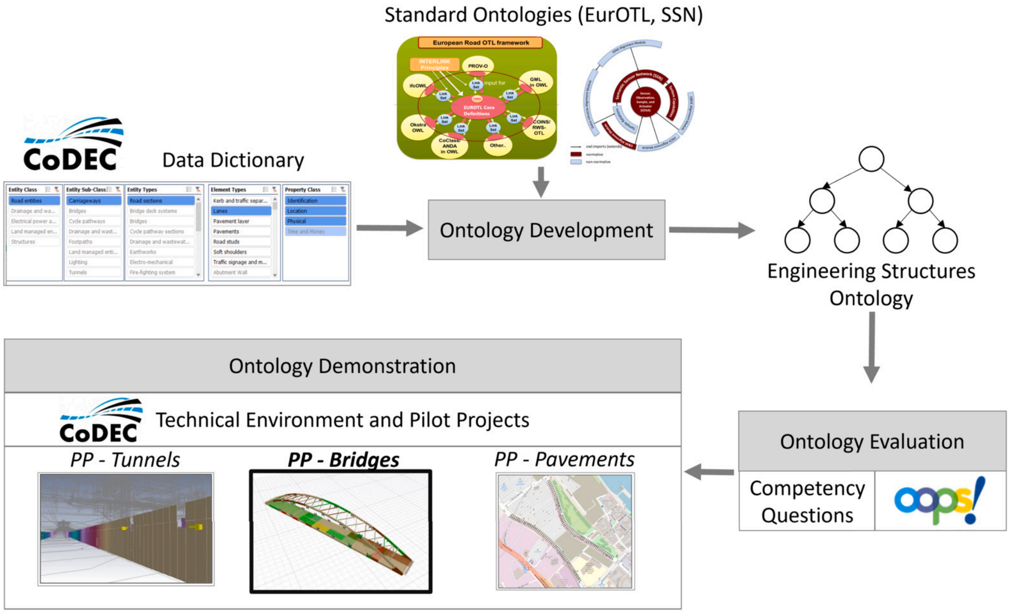


Fig. 1. Research methodology.

Highway infrastructures asset data stored in BIM models can provide AMS with a more accurate description of these structures and enable better decision-making in maintenance and repair activities. Interoperability between these systems can also improve stakeholder collaboration and coordination by ensuring that asset data remains available and consistent in both systems, regardless of the current stage of their lifecycle. Asset management can be improved by enriching BIM with semantic information through AMS, such as geographic information systems (GIS), or linked data integration. For example, Zhao et al. (2019) integrated BIM with GIS to improve the effectiveness of highway alignment and reduce planning risks, such as design errors and miscommunication, and avoid environmental hazards. Similarly, Meschini, Daniele, et al. (2022), Meschini et al. (2023), Meschini, Pellegrini, et al. (2022) integrated BIM information into GIS using a relational database to facilitate information management and improve the decision-making process through business intelligence reports in university buildings. Al-Kasasbeh et al. (2021) also proposed a relational approach to integrating asset management data with data extracted from BIM models, developing an integrated decision support system based on a work breakdown structure for all life cycle phases. Furthermore, Ait-Lamallam et al. (2021a, 2021b, 2021c) extended the IFC standard concepts and presented an ontological approach called IFCInfra4OM (Industry Foundation Classes for Operation and Maintenance of Infrastructures) to provide support to the O&M of transport infrastructures.

Semantically enriched solutions allow information to be presented to stakeholders more intuitively, enhancing the usability of BIM and improving the management of complex engineering structures (Jiang et al., 2023). A complete and detailed view of the structure, structural elements, and recent behavior-related dynamic data can enable structural engineers to plan, budget, and act more effectively, leading to cost savings, reduced downtime, and improved safety for road users. Nonetheless, different technologies, data formats, requirements, and standards used in AMS and BIM systems can hinder this interoperability (Kivits et al., 2013, Gao & Pishdad-Bozorgi, 2019, Garramone et al., 2020, Jiang et al., 2023).

In Europe, there have been efforts to standardize data formats for AMS, such as AM4INFRA (Marcovaldi & Biccellari, 2018, Kokot, 2019), but data management practices are largely tailored to the individual AMS within each NRA. Highways England (2020) and the Lithuanian NRA (Ratkevičiūtė, 2010) have made attempts to develop standardized "Data Dictionaries" for some asset types, but few other publicly-available data dictionaries were found in other countries. While some countries, like the Netherlands, Belgium, and Finland, have developed Object Type Libraries (OTL), there is a noticeable gap in the availability, extent and content of data dictionaries for highway assets, which hinders the effective use of data, especially within a BIM environment (Biswas et al., 2021a, 2021b)

The Connected Data for Effective Collaboration (CoDEC) project[1] aimed to implement BIM principles in the European Highways Industry, focusing on data exchange between BIM and AMS to manage asset data during the operational phase. The project was funded by the Conference of European Directors of Roads (CEDR). A "Master Data Dictionary" was developed during the project with legacy (AMS-based data) and sensor/scanner data concerning specific key infrastructures and assets, creating a base data structure for integrating different data management systems. This shared conceptualization was key to provide standard data formats that can be used between Europe's NRA and their systems.

In the CoDEC project, a framework was designed to support the connections between AMS and BIM platforms, allowing information to flow and be enriched between these systems. CoDEC linked operational data to BIM environments using semantic web and linked data principles. Using an ontology to model and represent structures, structural elements, and operational data, such as sensor information or legacy data, allows for the development of a single format for information exchange (Hartmann & Trappey, 2020) and enhances decision-making during the operational phase of these assets.

This paper presents the design of engineering structures ontology used in CoDEC and the main challenges faced during its development, validation and evaluation. The objectives of this research are as follows: a) Develop an ontology to encode the shared conceptualization provided by the CoDEC Data Dictionary in a machine-readable way to allow for system interoperability; b) Formally evaluate and validate the ontology; and c) Demonstrate its use as a base for data exchange between BIM and AMS.

This research follows the methodology presented in Fig. 1. First, the Engineering Structures ontology is developed following a standard ontology development methodology (NeOn Methodology Suárez-Figueroa et al. (2015)), based on existing standard ontologies and the CoDEC

[1] https://www.codec-project.eu/.

Data Dictionary. Afterwards ontology capabilities are evaluated using competency questions (defined by different stakeholders in the context of three pilot projects) and its formalization is validated using an online tool (OOPS! - Ontology Pitfall Scanner! Poveda-Villalón et al. (2014)). Finally, the ontology is demonstrated as a base for integration between BIM and AMS in three different pilot projects (tunnels, bridges and pavements).

The remainder of this paper is structured as follows: Section 2 introduces background concepts related to ontologies. Section 3 presents current literature related to BIM and ontology-based approaches. Section 4 presents the case study environment, detailing the CoDEC Data Dictionary and the real pilot projects. The Engineering Structures ontology specification and development is presented in Section 5, followed by the ontology evaluation and validation (in Section 6). Section 7 showcases the ontology demonstration, with a specific focus on the bridges' pilot project. Discussions and limitations are presented in Section 8. Finally, conclusions are found in Section 9.

2. Ontologies

Ontologies are used in Linked Data and Semantic Web to structure and share data between different users and systems. These "formal, explicit specifications of shared conceptualizations" Studer et al. (1998) allow sharing, reuse and analysis of knowledge concerning a domain of interest (Noy et al., 2001, Stephan et al., 2007). Ontologies encode domain concepts, properties, constraints, and relationships in a formal, explicit, and machine-readable way.

The World Wide Web Consortium (W3C)[2] defines the Resource Description Framework (RDF), RDF Schema (RDFS), SPARQL and the Ontology Web Language (OWL) as standards for the Semantic Web. RDF is the recommendation for the "creation, exchange and use of annotations on the Web" (Guarino et al., 2009, p.72). The resources are described in the form of triples (*subject property object*) (Pan, 2009), for example, "Professor" "rdfs:subClassOf" "Faculty Staff". The property used in the previous example (rdfs:subClassOf) is from the RDFS vocabulary, which added class and hierarchy concepts on top of RDF, together with the necessary inference rules. SPARQL is a W3C query language for accessing and manipulating data stored in RDF format, commonly used for querying semantic web data and knowledge graphs. Lastly, OWL provides additional vocabulary and expressiveness, such as disjointedness, symmetry, and cardinality. OWL also defines properties as either object (relationships between classes) or data (attributes) properties. The three OWL types, Lite, DL and Full, have different levels of expressiveness, with the choice of language coming down to a trade-off between expressiveness and inference capabilities, i.e., the more expressive a language is, the less inference it is capable of Su and Ilebrekke (2002).

2.1. Ontology engineering methodology

Mora et al. (2022) analyzed Ontology-Based Knowledge Management Systems (OKMS) implementation methodologies in real-world settings. 26 methodologies were identified in the literature review, from which the authors selected, through a set of criteria, analyzed and evaluated the following five methodologies: CommonKADS (Schreiber et al., 1994), Methontology (Fernández-López et al., 1997), On-to-Knowledge (Staab et al., 2001), NeOn (Suárez-Figueroa et al., 2015) and XDM, a agile methodology which was initially proposed as part of NeON (Blomqvist et al., 2016). CommonKADS and NeON were the most comprehensive and systematic for project management and technical systems development processes. The authors found that there are no standards or preferences for any of these methodologies in the literature and recommend using CommonKADS or NeON for medium or large OKMS projects, with agile methodologies, such as XDM, being preferred for smaller projects.

CommonKADS (Knowledge Acquisition and Documentation Structuring) (Schreiber et al., 1994) is a knowledge engineering methodology focused on knowledge management, analysis and knowledge system development. The construction of the system is based on a set of models: Organization, Task, Agent, Knowledge, Communication and Design. Templates are provided for these models, which can be completed or altered in parallel during the project (Schreiber et al., 2000).

Methontology was proposed in 1997 by Fernández-López et al. (1997) as an ontology engineering methodology. The authors present a set of activities and states, starting with planification. Specification, conceptualization, formalization, integration, implementation, and maintenance are the main activities identified by the authors for the development process. The evolving development lifecycle allows software or knowledge engineers to change between states during the development. Knowledge acquisition, evaluation and documentation are support activities that occur throughout the lifecycle.

Staab et al. (2001) proposed the On-To-Knowledge methodology for developing ontology-based Knowledge Management (KM) systems. The On-To-Knowledge methodology comprises six activities: feasibility study, kickoff, refinement, evolution and maintenance. It ranges from the early stage of starting a KM project to the final version of the ontology-based KM application.

The NeOn Methodology (Suárez-Figueroa et al., 2015) was developed during the Neon Project[3] to provide a framework for building ontology networks. It identifies and defines processes and activities for the construction process and introduces a set of nine scenarios that consider different Knowledge Resources inputs. According to Gómez-Pérez and Suárez-Figueroa (2009), the Ontology Requirement Specification Document (ORSD) is the main output of the ontology requirement specification activity. The authors propose that the conceptualization, formalization and implementation activities in NeOn should follow the Methontology or the On-To-Knowledge methodologies.

3. BIM and ontology-based approaches

The road infrastructure asset management field is rapidly becoming digital, leading to increasing data accessibility, integration, and collaboration challenges. Current processes lack full integration and face compatibility issues between systems, including BIM (Biswas et al., 2021a). To address this, ontology-based approaches have been proposed by several authors to integrate BIM data with other information (Farghaly et al., 2019, Zhong et al., 2019, Lei et al., 2021), such as sensor-based environmental information (Zhong et al., 2018) or regulatory data (Wang, 2021). Ontologies enable semantic representation for this information, trying to bridge the existing gaps in data management and automation. When compared to relational approaches, ontologies provide machine-readable and standardized models that allow accessibility and interoperability of knowledge related to an entity, which can be used to semantically enrich BIM data (Cursi et al., 2022). Jiang et al. (2023) state the integration of BIM with new technologies such as Linked Data as a future direction of BIM semantic enchantment to promote collaboration and improve efficiency of engineering projects.

Ontologies can be used in safety management, improving personal and structural safety during the construction stage (Chen & Bria, 2022, Li et al., 2022, Fang et al., 2020, Lee & Yu, 2023), but can also be used during the design and O&M stage (Jiang et al., 2023). By encoding product features information with an ontology, reasoning and validation rules can be used to ensure that manufacturing rules are followed during the design phase enabling real-time feedback to designers, regarding the product manufacturability (Cao et al., 2022). During the O&M stage, ontologies can be used together with BIM to improve

[2] https://www.w3.org/.

[3] http://neon-project.org/.

several processes, such as energy performance assessment and management (Wu et al., 2023), monitor building environment variables (e.g., temperature, light, CO2) (Zhong et al., 2018), and provide a base for sharing construction defects information (Lee et al., 2016). Ontologies can also be used in project management during the infrastructure's lifecycle (Wu et al., 2021).

Wang (2021) presents a domain ontology to support O&M of underground utility called Utility Operation and Maintenance Ontology (UOMO) to integrate and encode standards, regulations, and expert knowledge, utility and environment data, and, inspection and maintenance reports. Based on this ontology, the author proposes a framework that supports O&M activities and decision-making, taking advantage of ontology querying, inference and rules. The integration with other systems (GIS) is presented as future work.

Ding et al. (2016) present an ontology-based methodology for risk knowledge management in construction, and integrate this knowledge within a BIM-environment for risk analysis. However, the authors present the lack of compatibility with IFC and other standards as one of their limitations. Furthermore, Zhou et al. (2023) introduce a dam safety monitoring systems domain ontology (OntoDSMS) to address the analysis of heterogeneous data and sources needed for evaluating dam safety. The authors reuse existing ontologies for sensor data and IFC and find that SPARQL is more efficient and allows for improved logical reasoning than traditional methods.

Hagedorn et al. (2023) present a solution for enhancing BIM-enabled infrastructure asset management for road owners using Information Containers for Linked Document Delivery (ICDDs) to meet the diverse requirements of stakeholders during the operational phase. The authors present the development of a web-based platform for asset management, utilizing the ICDDs, Semantic Web technologies (like RDF and SPARQL), and domain-related ontologies as schemas. Two use cases demonstrate the practical application, showing how ICDDs can be used in tasks such as visual inspection of bridges and decision-making for road pavement maintenance activities. Future research directions include aligning existing ontologies, automating geometric representation updates, and integrating sensor data for infrastructure digital twins.

In summary, three main limitations were found in this related work analysis. Firstly, most authors develop and use their domain- and task-specific ontologies. However, most works fail to use standard or higher-level ontologies, which undermines their interoperability efforts and hinder the use of the respective knowledge by other systems or potential users. Secondly, BIM data is usually imported to the ontologies, leading to ontology-based analyses most of the times, and creating an uni-directional flow of information. While not necessarily a problem *per se*, a bi-directional flow, where ontology knowledge can be integrated into the BIM model, can allow for BIM-based systems to display ontological information managed by external systems, such AMS (Ding et al., 2016). Lastly, the ontology-based analysis most of the times is presented using the development system (e.g., Protégé) or through SPARQL analyses. While effective, these solutions do not take into consideration user-friendliness, and better ontology visualizations should be provided (Lee et al., 2016, Lei et al., 2021).

4. Case study: the CoDEC project

This section introduces the research context for this manuscript, namely, the CoDEC Data Dictionary that details the main concepts and vocabulary for highway infrastructures, and three real-case pilot projects across European countries, focused on different types of assets.

4.1. Data dictionary

Although European NRA have started to use BIM during the design and building phase of projects and have well-defined processes and AMS, little has been done to use BIM for long-term asset maintenance management (Biswas et al., 2021b). An AMS holds information about a specific asset and allows users to analyze the data, but each NRA has their own AMS to suit their needs and often each asset type has its own AMS and there is no interaction of data across different AMS. On the other hand, BIM is a system to digitally model an asset, which makes it easier to create and share information during asset design, construction, and operating phases.

For the purpose of standardizing the data connectivity, the CoDEC Data Dictionary was developed to provide a shared vocabulary to enable the integration and sharing of data between different systems with a common data definition and an hierarchical system (Biswas et al., 2021a).

To obtain information for the data dictionary, engineering structures and highways' stakeholders were inquired, which include NRA from 14 different European countries (Austria, Belgium, Denmark, Finland, France, Germany, Lithuania, the Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, and the United Kingdom), and implementations partners, such as BIM and AMS software companies. Also, several works such as AM4INFRA (Kokot, 2019), the Highways England UK-ADMM data dictionary (Highways England, 2020), the Data Standard for Road Management and Investment in Australia and New Zealand (Austroads, 2019) and ifcRoad (buildingSMART, 2020) were used to support the Data Dictionary development.

CoDEC had also a specific goal to handle sensors and their data, as these are increasingly used to support infrastructure asset management. Sensors were considered as separate objects, and not as an asset, and various property sets were created for both mobile and fixed-location sensors. This explains the variations in how fixed and mobile sensors are located and referenced covering different criteria (e.g. skid resistance, longitudinal evenness, rutting, cracking, raveling, potholes), different technical parameters for same criterion (e.g. IRI, WLP, NBO, EC for longitudinal evenness) and different combinations in indicators (e.g. safety indicator with different components).

The Data Dictionary was formalized in Excel and contains asset data, its metadata and attributes, the logical and hierarchical connections, and the list of data types for creating an Object Type Library (OTL). The last version of this resource is available on the project's website.[4]

4.2. CoDEC pilot projects

The Data Dictionary followed the requirements of the three pilot projects, with a focus on three key highway civil assets (tunnels, bridges and pavements), as well as preliminary concepts and relationships for supporting systems and assets (e.g., lighting, fire-fighting, and drainage).

Focused on tunnel structures, the first pilot project (*PP - Tunnels*) case study aims to demonstrate sensor data integration into the BIM exploitation environment. It was necessary to encode information about the sensors and their data to provide the BIM environment with the necessary operational data for decision support. This information is used to colorize the sensors in the 3D model, using a color pallet related to the air quality in the tunnel.

The bridges pilot project, *PP - Bridges*, aims to provide data about the risk and condition of bridge structural elements. The information from each assessment campaign about the structure's condition is loaded into the ontology and then used to apply a color encoding to the model elements according to a given scale, with respect to the risk level.

The last case study, *PP - Pavements*, focuses on road networks and highways. While the previous two pilot projects have the objective of delivering operational data into a BIM environment, this pilot project aims to enrich their GIS with information from BIM (requiring accurate spatial mapping between the two). GIS-based AMS are used for decision support in these types of structures.

4 https://www.codec-project.eu/Resources/projectreports.

5. Engineering structures ontology

This section presents the main contribution of this paper, namely the development of Engineering Structures ontology. The NeOn methodology (Suárez-Figueroa et al., 2015) was used to define the required activities for this development process due to its focus on knowledge resources inputs, in addition to being the most recent and complete methodology (see Section 2.1). Ontology requirements are presented in the Ontology Requirements Specification Document. Afterward, the development and conceptualization process is reported, discussing the main challenges and decisions.

5.1. Ontology requirements specification document

5.1.1. Domain and scope

The Engineering Structures ontology was developed to describe and store knowledge related to the European highways industry. Specifically, the ontology should represent concepts related to bridges, tunnels, and pavements, their structural elements, and the dynamic data associated with these assets.

5.1.2. Goals

The ontology should represent structures, such as bridges and tunnels, and their structural elements, such as pylons and cables, providing asset information in a formal, comprehensible, and explicit way. The concepts and relationships described in the ontology are based on the CoDEC's Data Dictionary. The ontology should also store information about sensor and inspection data (Risk and Condition Data) and ensure the connection between the BIM model and these entities. For interoperability purposes, the ontology should extend Interlink project's EurOTL.[5]

5.1.3. Users, use cases and applications

The Engineering Structures ontology should provide information about its domain to the users, i.e., structural owners, managers, and operators. The ontology should allow users to analyze structures and structural elements (information related to location, activities, size and other physical attributes) and how they are related, i.e., which elements are part of a particular structure. Furthermore, the ontology can provide sensor and sensor data information to the user, such as how many observations a sensor made and where they are located. The same should be valid for inspections and risk and condition analysis. Lastly, users can obtain information concerning pavement sections, layers and their geometric representation.

The ontology will be used as part of the CoDEC Technical Environment to provide the necessary information and knowledge for the execution of the three pilot projects and allow information exchange between AMS and BIM environments.

5.1.4. Knowledge sources and reusable ontologies (inputs)

The following Knowledge Resources were identified:

a) CoDEC Data Dictionary (see Section 4.1) is a Non-Ontological Resource that provides a shared vocabulary for knowledge acquisition and elicitation from the different stakeholders. This resource provides the main body of knowledge that will be formalized and encoded by the ontology;
b) EUROTL Framework Ontologies are ontological resources extended by the ontology. By extending these concepts, Engineering Structures ontology can be used by any EurOTL interface or application. The European Road OTL (EurOTL) was developed during the Interlink project and contains ontologies and Linking Rule Sets related to European roads. The core ontology is available at "http://www.roadotl.eu/def/". The linksets in Table 1 were also used, providing machine-readable mapping descriptions between the framework's domain ontologies and the EUROTL core ontology.
c) Sensor Network Ontology is a W3C recommendation for describing sensors, sensor networks and their observations. This ontology provides a starting for encoding the necessary dynamic data ad is available at "http://www.w3.org/ns/ssn/".

Table 1
EurOTL domain ontologies.

Domain Ontologies	Linkset Location
AM4INFRA	http://www.roadotl.eu/AM4Infra--eurotl/def/
IFC4x1_Final	http://www.roadotl.eu/IFC4x1_Final--eurotl/def/
GeoSPARQL	http://www.roadotl.eu/geosparql--eurotl/def/
INSPIRE transport networks	http://www.roadotl.eu/inspire--eurotl/def/
ISO19148 transport networks	http://www.roadotl.eu/iso19148--eurotl/def/

5.1.5. Competency questions

The definition of an ontology's scope is a crucial step in ontology development. The use of competency questions (CQ) to determine an ontology' scope is a standard practice in ontology development (Noy et al., 2001). CQs have been used in several works related to construction to evaluate an ontology's capability. (e.g., Cao et al. (2022), Zheng et al. (2021), Kukkonen et al. (2022)). This process helps to ensure that the ontology is designed to capture the relevant knowledge and information within its intended domain.

The CQ were formalized in the context of the three pilot projects (see Section 4.2) based on inputs from the different stakeholders. Table 2 presents a sub-set of the above-mentioned CQ for which the ontology is required to provide answers. The CQs are divided into General Questions and PP-specific questions related to the pilot project requirements and corresponding use cases. Specifically, *PP - Tunnels* focuses on sensor data, *PP - Bridges* is related to risk and condition data of bridge structural elements, and *PP - Pavements* focuses on road network pavements.

5.2. Ontology development

The Engineering Structures ontology development followed an incremental lifecycle. The first conceptualization was based on the Data Dictionary, while the remaining lifecycles focused on each pilot project requirements. The Engineering Structures ontology was developed in OWL using Stanford's Protégé.[6]

5.2.1. Initial development

The Engineering Structures ontology initial development was done by mapping or aligning Data Dictionary concepts (classes or properties) to EurOTL concepts. If a given concept is already available in EurOTL, there is no need to create and extend the same concept in CoDEC. However, if this is not the case, the new CoDEC concept is created as a sub-class of an existing EurOTL entity, ensuring interoperability between the two ontologies. For example, the "Bridge" concept already exists in the EurOTL framework, specifically in the AM4Infra vocabulary, so it is not necessary to extend concepts. However, "Structural Elements", or equivalent, are not found in any of the vocabularies or ontologies from the EurOTL framework. In this case, the "PhysicalObject" class from EurOTL was extended in the Engineering Structures ontology with a new class used to represent structural elements. Fig. 2 shows an example of the mapping between the Data Dictionary and the Engineering Structures ontology.

5.2.2. Semantic sensor network

The main requirement from *PP - Tunnels* was the integration of sensor metadata and data in the ontology for operational safety man-

[5] https://www.roadotl.eu/.

[6] https://protege.stanford.edu/.

Table 2
Competency questions.

General Questions	
CQ1	When did a certain structure ended its construction phase?
CQ2	Where is a certain structure located?
CQ3	What are the measurements of a certain structure?
CQ4	Who is the owner of a certain structure?
CQ5	Which and how many elements are part of a structure?
***PP - Tunnels* Specific Questions**	
CQ6	Which sensors are hosted by a structure and how many observations did they make?
CQ7	What is the location of the sensor data related to an observation?
***PP - Bridges* Specific Questions**	
CQ8	What are the results of a certain inspection by element?
CQ9	What is the risk of a given structure according to an inspection?
CQ10	What is the last risk analysis result of a certain element?
***PP - Pavements* Specific Questions**	
CQ11	What is the total thickness of a given section and how many layers does it contain?
CQ12	How is a given section subdivided?
CQ13	What is the geometric representation of a given pavement subsection?

Data Dictionary			Ontology		
Property	**Description**	**Format**	**Domain**	**Object/Data Property**	**Range**
Bridge ID	The unique reference identifier for bridge	String	bridgeID	rdf:type	Bridge
Bridge name	The name of the bridge	String	bridgeID	rdfs:label	xsd:string
Environment	Classification of surrounding environment (e.g.: Rural/Urban)	String	bridgeID	inEnvironment	xsd:string
Region/District/ Area	Relevant geographical situation	String	bridgeID	prov:atLocation	eurotl:LocationBy Identifier
Owner	Owner of the asset	String	bridgeID	hasOwner	prov:Agent (Person or Org.)

Fig. 2. Data dictionary to engineering structures ontology. Mapping example for "bridge" entity. Adapted from CoDEC Project Report (2021).

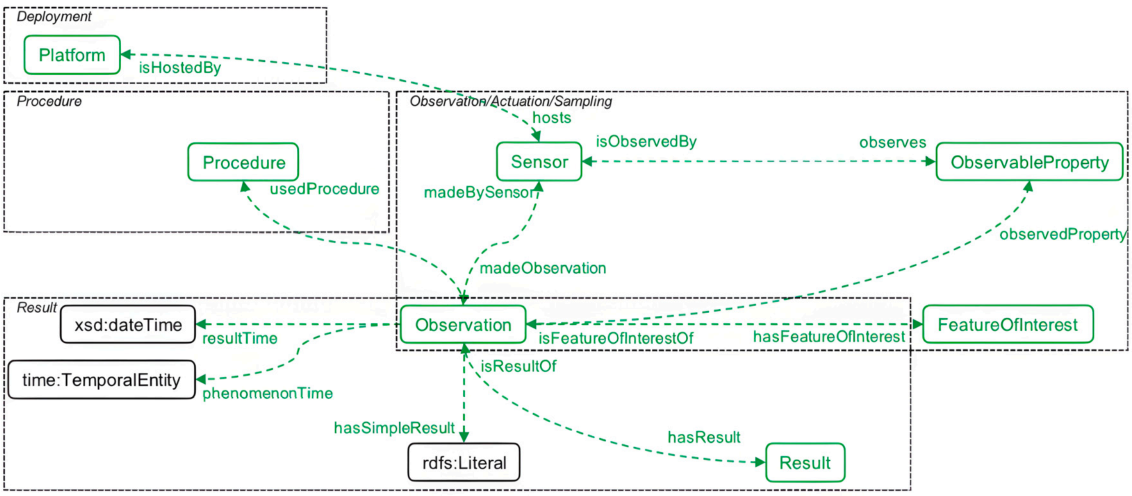


Fig. 3. Semantic sensor network ontology. Retrieved from Open geospatial consortium (2017).

agement and monitoring, specifically air quality analysis in tunnels. The EurOTL framework does not provide vocabulary or domain ontologies concerning dynamic data. The Semantic Sensor Network (SSN) Ontology,[7] a W3C standard, was used to encode this information (Fig. 3).

In the Engineering Structures ontology, structures are seen as platforms that host a set of Sensors. Each time-series concerning an "ObservableProperty" is encoded as an Observation. However, the sensor data itself is not stored within the ontology. Instead, each time-series is stored in a JSON file. The location of this file is obtained from any Observation using the data property "hasDocument", from eurOTL.

5.2.3. *Risk and condition data*

As stated before, *PP - Bridges* aims at analyzing Risk and Condition data. Contrary to the dynamic data automatically collected by sensors, Risk and Condition data is generated during assessment campaigns, rep-

[7] https://www.w3.org/TR/vocab-ssn/.

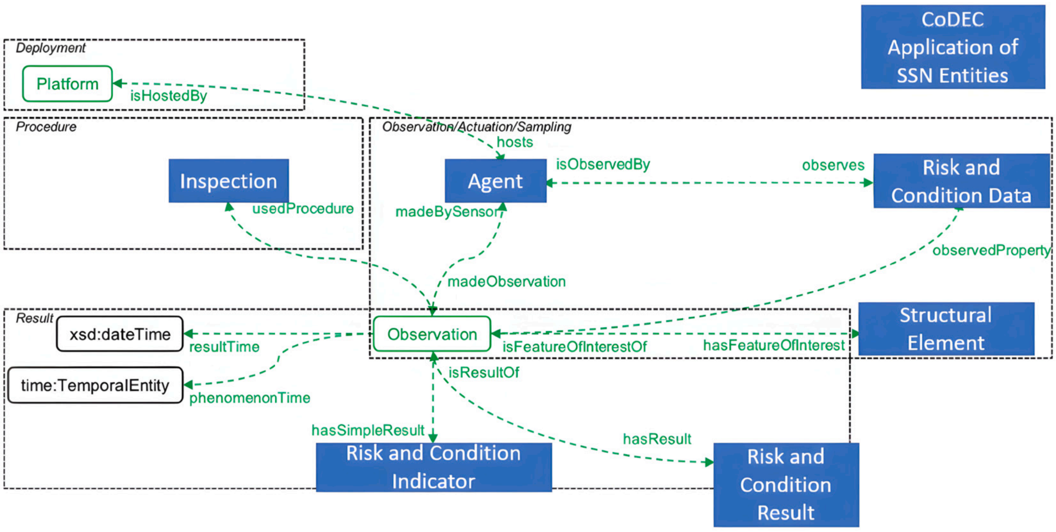


Fig. 4. CoDEC risk and condition data over SSN.

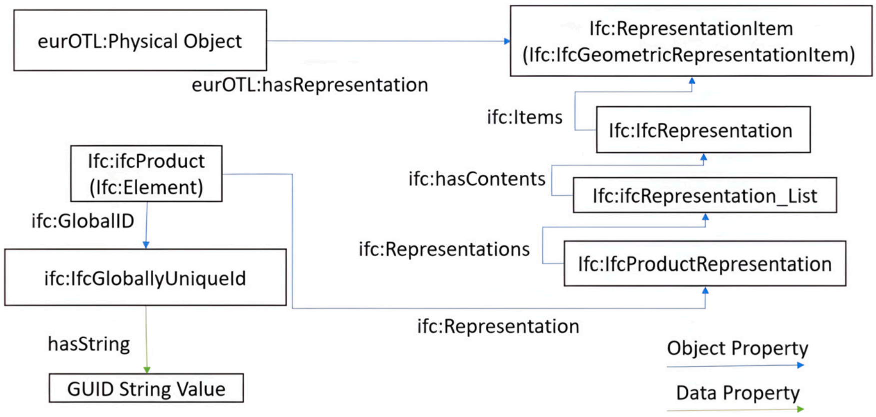


Fig. 5. Connection between PhysicalObject to a ifcElement GUID.

resented as Inspections in CoDEC. The SSN concepts were then extended to encode the necessary information about Observations obtained by a particular Procedure (in this case, the Inspection itself, as seen in Fig. 4). The Observation is done by an Agent, taking the Sensor role, and concerns a given Structural Element ("hasFeatureOfInterest").

Two properties can be used to obtain the results from an Observation: (1) "hasSimpleResult", returning a Risk and Condition Indicator with a numeric scale from 1 to 5; and (2) "hasResult", which points to a Risk and Condition Result, containing a descriptive state, the inspection due date and a URL for photos.

5.2.4. *ifcOWL*

A link needs to be established to be able to transfer any exchange any information between Engineering Structures ontology and the BIM model. The ifcOWL ontology, which is part of the eurOTL framework, provides a way to represent IFC models (a BIM data format) in OWL. Through a series of complex relationships (see Fig. 5), the ontology can relate any eurOTL Physical Object (from which structures and structural elements are extended in the Engineering Structures ontology) to an ifcElement global unique identifier. In *PP - Bridges*, this link is needed to relate Risk and Condition indicators to the element's BIM representation and colorize each Structural Element.

5.2.5. *Pavement sections and layers extensions*

The EurOTL's linear referencing method was used in *PP - Pavements* to identify and locate pavement sections in a given road network. Pavement sections, subsections and layers were added to the Engineering Structures ontology to ensure the needed representation detail, together with data properties such as layer thickness or vertical position.

5.3. *Ontology population*

Ontology population is the process of adding instances in the ontology (called A-Box statements). To validate and evaluate the Engineering Structures ontology, data related to each pilot project was added using a Protege plugin called Cellfie.[8] Cellfie was used to define a set of import rules and mappings (based on Manchester OWL Syntax[9]) from Excel spreadsheets into OWL axioms (see Fig. 6). This solution was used for *PP - Tunnels* and *PP - Bridges*, while *PP - Pavements* used a different method, based on Python scripts, to directly import and export data from the ontology.

6. Ontology evaluation

This section presents the ontology evaluation process. Following the NeON methodology, the ontology is evaluated regarding competency question answering and common pitfall detection. The evaluation should be done independently from the application scenario or technical environment that will take advantage of this ontology.

[8] https://github.com/protegeproject/cellfie-plugin.
[9] https://www.w3.org/TR/owl2-manchester-syntax/.

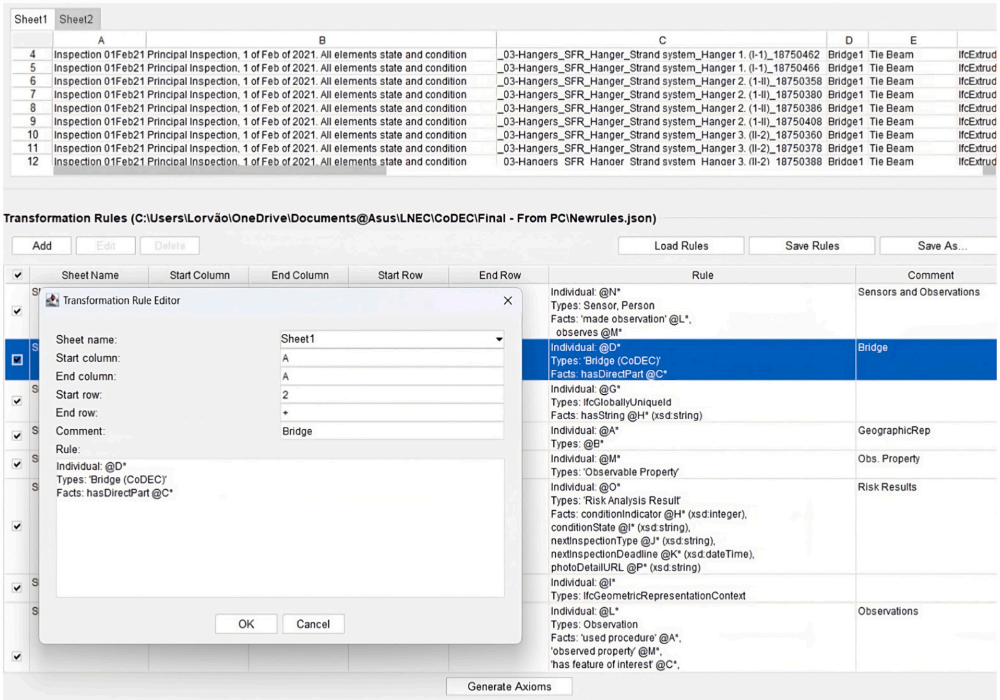


Fig. 6. Cellfie import rules and mappings.

6.1. Competency questions

In this section, the Engineering Structures ontology will be used to answer the Competency Questions defined in the Ontology Requirements Specifications Document.

The set of attributes from the Data Dictionary, now formalized by the Engineering Structures ontology, allows for a more detailed definition of structures and structural elements. Information related to time, location, physical properties, such as measurements and materials, and relationship with agents, such as the owner or commissioner, can now be asserted in the ontology. Competency questions were defined to illustrate how the ontology can currently answer these questions (CQ1 to CQ4). The SPARQL query for CQ1 is shown in Listing 1, which returns the end date (xsd:date) of a given structure's construction phase.

Listing 1. CQ1 - when did a certain structure ended its construction phase?

```
SELECT ?date
WHERE {
    <Structure> codec:hasConstructionDate ?date
}
```

Another competency question concerns structural elements and their relation to a structure (CQ5). The "cmo-simple-cdt:hasDirectPart" object property from eurOTL was used to encode this relationship. Although the relationship itself is not transitive, SPARQL can be used to make inferences as if this were the case. This inference allows information to be obtained about elements that are directly part of a structure or all elements related to a structure, as shown in Listing 2.

The query in Listing 3 can be used to obtain the risk associated with a structure according to a given inspection. All elements related to the structure are obtained, as well as the observations of these elements in a given inspection. The Structure risk (on a scale from 1 to 5) is obtained by calculating the minimum Risk and Condition Indicator from all observations.

However, the ontology can also be used without specifying an inspection. The query for CQ10, presented in Listing 4, collects all in-

Listing 2. CQ5 - which and how many elements are part of a structure?

```
SELECT (?type as ?ELEMENTTYPE) (COUNT(?element) AS ?
    ELEMENTCOUNT)
WHERE {
    <Structure> cmo-simple-cdt:hasDirectPart+ ?element.
    ?element rdf:type ?type.
    ?type rdfs:subClassOf codec:Structural_Element.
} GROUP BY ?type
```

Listing 3. CQ9. What is the risk of a given structure according to an inspection?

```
SELECT (MIN(?result) as ?minResult)
WHERE {
    <Structure> cmo-simple-cdt:hasDirectPart+ ?element.
    ?o sosa:usedProcedure <Inspection>;
        sosa:hasFeatureOfInterest ?element;
        sosa:hasSimpleResult ?result.
}
```

Listing 4. CQ10. What is the last risk analysis result of a certain element?

```
SELECT ?predicate ?object
WHERE{
        ?inspectionID rdf:type sosa:Procedure;
                rdf:type eurotl:InspectionActivity.
    {SELECT (MAX(?time) as ?mostRecent) WHERE{ ?
        inspectionID prov:atTime ?time}}
    ?inspectionID prov:atTime ?mostRecent.
    ?o sosa:usedProcedure ?inspectionID;
        sosa:hasFeatureOfInterest <ELEMENT>;
        sosa:hasResult ?result.
    ?result ?predicate ?object.
```

spection activities that are also procedures and selects the most recent. Then, given a structural element, the query returns all the information related to the Risk and Condition Results, including the condition indi-

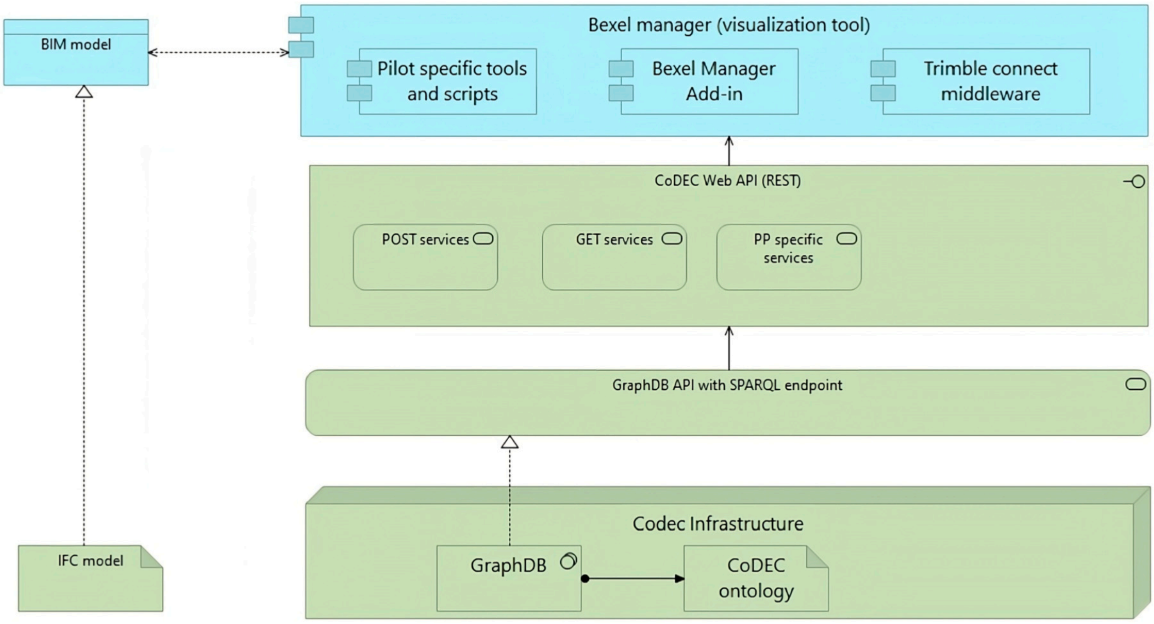


Fig. 7. CoDEC technical architecture. Retrieved from CoDEC Project Report (2021).

cator and state, the next inspection deadline and type, and a URL of the photo detail.

Listing 5. CQ8 - what are the results of a certain inspection by element?

```
SELECT ?element ?connectedTo ?result ?GUID where {
    ?o sosa:usedProcedure <Inspection>;
        sosa:hasFeatureOfInterest ?element;
        sosa:hasSimpleResult ?result.
    ?connectedTo cmo-simple-cdt:hasDirectPart ?element.
    ?element eurotl:hasRepresentation ?representation.
    ?shape ifc:items_IfcRepresentation ?representation.
    ?repList <https://w3id.org/list#hasContents> ?shape.
    ?prodDef ifc:
        representations_IfcProductRepresentation ?
        repList.
    ?ifcElement ifc:representation_IfcProduct ?prodDef;
                ifc:globalId_IfcRoot ?GUIDObj.
    ?GUIDObj <https://w3id.org/express#hasString> ?GUID.
}
```

Lastly, Listing 5 presents the SPARQL query that answers CQ8 and showcases the intricate connection between elements and their BIM representation. This complex set of relationships utilizes ifcOWL to connect elements' risk and condition indicators collected during a particular inspection to their IFC's global unique identifier.

6.2. OOPS!

The ontology was validated using OOPS! (Poveda-Villalón et al., 2014). OOPS! (OntOlogy Pitfall Scanner!) detects common mistakes and pitfalls made during ontology development. When analyzing the Engineering Structures ontology, the tool did not detect any critical pitfalls, which "could affect the ontology consistency, reasoning, applicability, among others" (Poveda-Villalón et al., 2014, p.15). However, 55 important pitfalls are reported, although only one is directly related to the ontology, with the remainder being related to the imported ontologies (i.e., EurOTL, SSN). The tool also detected nine minor pitfalls, with three being related to the Engineering Structures ontology.

However, some of the detected pitfalls do not represent a problem or error. For example, the important pitfall identified related to the Engineering Structures ontology relates to equivalent classes not being explicitly declared. The tool warns that "Span" and "Bridge" classes might be equivalent (Span is a synonym for bridge outside the civil engineering context), which is not the case. The remainder pitfalls are minor and related to different concepts in the same class and the different naming conventions in the ontology. For example, the "Drainage and wastewater collection" class was identified as the terminology for a type of "Structural Element" in the data dictionary, which the tool identifies as a possible pitfall.

7. Ontology demonstration

The CoDEC pilot projects were used to demonstrate the use and usefulness of Engineering Structures ontology. *PP - Tunnels* takes advantage of the integration of the SNN ontology with the eurOTL concepts and Engineering Structures ontology to present air quality analysis in a BIM environment. The Engineering Structures ontology's extension of the SNN ontology, which allows risk and condition analysis of structural elements, is demonstrated in *PP - Bridges*, together with the connection of these elements with their BIM representation (IFC model). Lastly, *PP - Pavements* utilizes Linear Referencing concepts (provided by eurOTL) and Engineering Structures ontology's section and layers pavement extensions to correctly map GIS and BIM elements. Due to the focus on the integration and extension of SNN ontology for Risk and Condition data and the use of ifcOWL to connect structural and operational data with BIM elements, this section is focused on *PP - Bridges*, showcasing the use of the Engineering Structures ontology as an enabler for data exchange between BIM and AMS systems.

7.1. Technical architecture

Fig. 7 presents the Technical Architecture used for the pilot projects. The figure uses ArchiMate 3.0 notation[10] to define the components and layers.

The CoDEC infrastructure (bottom layer) stores ontology instances according to pilot project requirements, allowing knowledge to be accessible and manipulated. The Engineering Structures ontology (presented as CoDEC ontology in Fig. 7) and its details are present in Sec-

[10] https://pubs.opengroup.org/architecture/archimate3-doc/.

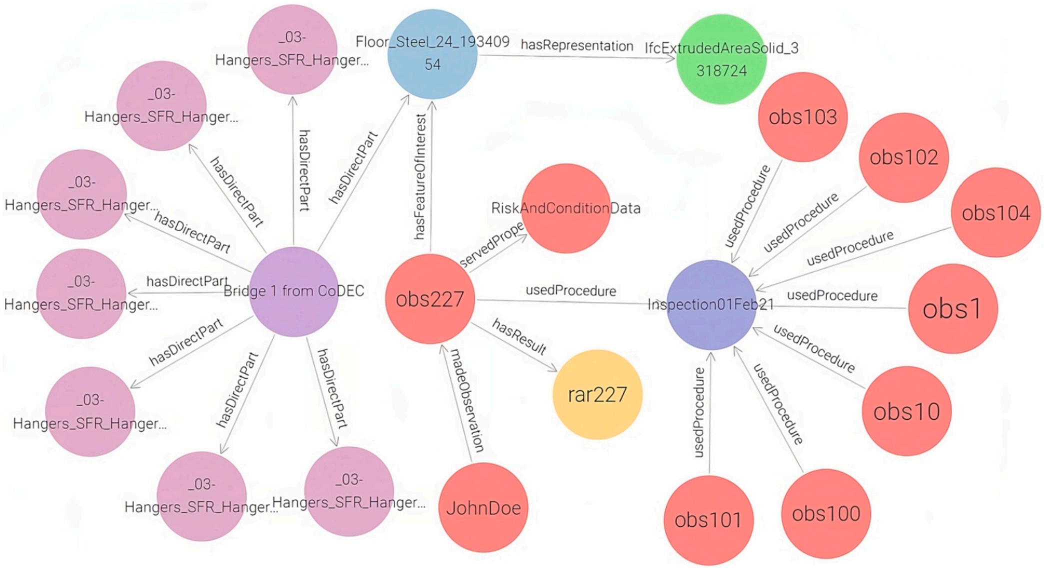


Fig. 8. Populated ontology in GraphDB.

tion 5. To access this environment, the CoDEC Web API was developed. The Application Programming Interface (API) services are critical for this solution because they create an abstraction layer between the ontology (data) and its users or applications (logical levels). This abstraction layer allows any solution to access the linked data environment without any technical dependencies and without needing to know and follow the ontology's structure or its evolution (i.e., there is no need to develop standalone queries for each application or scenario).

Finally, applications or tools for data management and visualizations are created, such as Bexel Manager Add-In. These tools allow access to the API to retrieve and present the information, hiding environment and ontology complexity from the end user.

7.2. Accessing the ontology

CoDEC uses GraphDB[11] as a linked data environment to store the populated ontology (see Fig. 8). GraphDB is a graph database compliant with W3C standards (i.e., RDF, OWL, SPARQL). Once stored, the ontology can be queried or updated using SPARQL endpoints. Inside the CoDEC environment (see Section 7.1), an API was developed to create an abstraction layer between application and (ontological) data layers. A set of REST services are exposed by CoDEC's API, which allows users and applications to easily communicate with complex linked data environments stored in GraphDB.

One of the services provided by CoDEC's API, "GetInspectionResult", uses a simplification of the query presented for CQ8 (see Listing 5) to return, given an inspection, pairs of risk and condition indicators and element's IFC global unique identifier. Using the inspection "codec:Inspection01Feb21" as the request parameter, the API returns a response as demonstrated in Fig. 9, encoded in JSON.

result	GUID
"2"^^xsd:integer	"23wH3G9sL3GgDaiKQFq9tl"
"1"^^xsd:integer	"23wH3G9sL3GgDaiKQFq9s5"
"2"^^xsd:integer	"3xfYuu1AL2Sh2mXvoR5PKz"
"2"^^xsd:integer	"3xfYuu1AL2Sh2mXvoR5PL3"
"2"^^xsd:integer	"3xfYuu1AL2Sh2mXvoR5PL1"
"1"^^xsd:integer	"3xfYuu1AL2Sh2mXvoR5PL7"

Fig. 9. Response example.

7.3. PP - Bridges demonstration

For *PP - Bridges*, an existing bridge IFC model was imported using Bexel Manager,[12] and an add-in was created from the application side. This add-in communicates with the CoDEC API to retrieve existing inspections related to the bridge. Afterwards, the user can select an inspection and risk indicators related to each element are retrieved ("GetInspectionResult" service) from the ontology and used to colorize the bridge (see Fig. 10). Furthermore, the completed risk and condition assessment of a single or a set of elements can also be obtained, with additional detail such as photo URLs and information related to the following inspection schedule for each element.

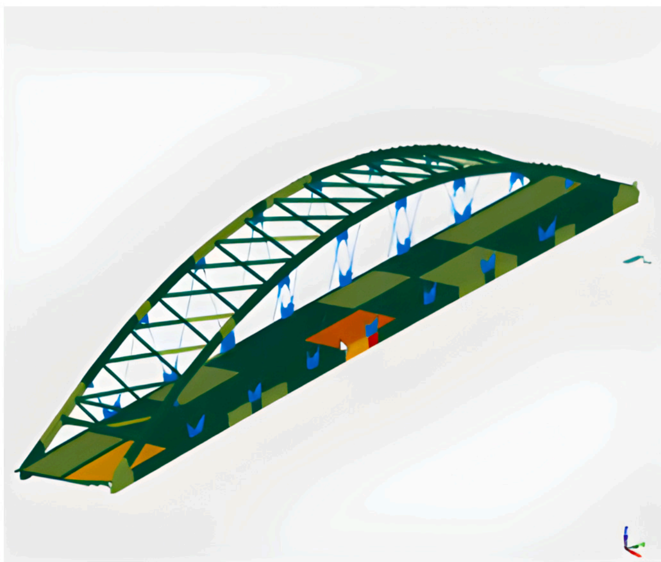


Fig. 10. Bridge elements colored according to risk indicator from an inspection. Retrieved from CoDEC Project Report (2021).

[11] https://graphdb.ontotext.com/.
[12] https://bexelmanager.com/.

8. Discussion

This article presents the design and development of the Engineering Structures ontology. The Engineering Structures ontology was developed to address: a) the interoperability challenge within NRA systems, e.g., the integration of operational and sensor data with BIM models; and b) the sharing of relevant information between NRA, based on a shared and formal conceptualization.

The shared conceptualization provided by the CoDEC Data Dictionary was validated by experts from several European NRA in the CEDR project's scope. The proposed ontology, based on the Data Dictionary, allows NRA to encode their data in a homogeneous way, enabling semantic and data interoperability between them. The use of the ontology for representing operational asset information was validated by experts and is now formally evaluated using competency questions and validated with OOPS! (see Section 6). The ontological approach offers a flexible, scalable, and interoperable integration framework for integrating AMS data into the BIM environment, ensuring semantic clarity and facilitating efficient data management and analysis.

8.1. Contributions to the theory

This research makes several contributions to the theory regarding the integration of BIM with AMS, particularly focusing on European highways. Physical infrastructure elements like buildings, bridges, and roads can be modeled and managed with BIM, enabling stakeholders to collaborate, share, and exchange information needed for decision support throughout their entire lifecycle, namely during the design, planning, and construction phases. However, the application of BIM in transport infrastructures is not yet widespread (Costin et al., 2018), particularly in the O&M phase (Talebi, 2014, Wijeratne et al., 2024). NRAs use AMS to manage, maintain, and ensure structural safety during their O&M phase.

The seamless integration of BIM with AMS presents an opportunity to optimize decision-making processes in the O&M phase of engineering structures. As asset monitoring solutions continue to evolve, managing critical information, interoperability between these systems becomes increasingly vital for timely decision-making in an integrated environment. By linking 3D model data with asset management information, stakeholders can access crucial insights more readily, leading to reduced errors, improved cost-effectiveness, and heightened safety and compliance measures (Wijeratne et al., 2024). However, achieving seamless interoperability is a complex challenge due to the different technologies, data formats, and standards utilized across AMS and BIM platforms (Kivits et al., 2013, Gao & Pishdad-Bozorgi, 2019, Garramone et al., 2020). The adoption of semantically enriched approaches offers a way of improving the usability of BIM environments, contributing to the management of complex engineering structures. By presenting information in a more intuitive way, these solutions promote better decision-making, which can ultimately lead to the optimization of the performance and longevity of infrastructure assets (Jiang et al., 2023).

This work addresses the gaps identified in the literature review regarding the use of semantic techniques with BIM (see Section 3). Firstly, this study addresses the identified limitations in existing literature regarding the development and use of domain-specific ontologies. While previous works often develop and utilize their own task-specific ontologies (e.g., Wang (2021), Ding et al. (2016), Zhou et al. (2023)), authors frequently neglect to incorporate standard or higher-level ontologies. By taking advantage of standard ontological frameworks, such as the EurOTL framework and the Sensor Network Ontology, the Engineering Structures ontology proposed in this study contributes to ameliorate interoperability challenges. The proposed ontology serves as a formal representation of the shared conceptualization provided by the CoDEC Data Dictionary, ensuring that NRAs can encode their data in a homogeneous way, addressing the identified need for standardized data formats and facilitating semantic and data interoperability between European NRA and their systems.

Secondly, this research contributes to the literature by acknowledging and addressing the uni-directional flow of information often observed in ontology-based analyses. In current literature, BIM data is typically imported into ontologies (e.g., Zhao et al. (2019), Al-Kasasbeh et al. (2021), Meschini, Pellegrini, et al. (2022), Zhou et al. (2023)), creating an uni-directional flow of information. Although existing solutions, such as the use of Protege or SPARQL analyses, are effective, they often lack an intuitive and familiar interface (i.e. similar to the interfaces of BIM environments for managing the assets of road structures). This study contributes to the literature by highlighting the importance of user-friendliness in ontology-based analyses. By integrating operational asset information managed by AMS with BIM models, as shown in Section 7, the ontology-based approach enables BIM-based systems to display relevant ontological knowledge, thereby enhancing interoperability and decision-making processes during the O&M phase (Lee et al., 2016, Lei et al., 2021).

8.2. Contributions to the practice

Being machine-readable, the ontology can be used as a base for data exchange between BIM and AMS, or other systems. As demonstrated in Section 7, when used within the CoDEC framework, the ontology enables BIM-based analysis of operational data, such as risk and condition data, from AMS. While the Engineering Structures ontology can be used in isolation to provide a similar analysis, users (i.e., structural owners, managers, and operators) can now access data in a familiar decision environment by integrating operation information in a BIM environment. Furthermore, ontology-based analysis can be enabled in the decision-support environment, allowing users to take advantage of the knowledge representation and inference provided by this encoded shared conceptualization and other semantic web technologies, such as SPARQL, Shapes Constraint Language (SHACL)[13] or SWRL: A Semantic Web Rule Language.[14] These technologies can be used to validate information stored in the ontology (e.g., ensure that each structural element is represented by a BIM model entity), or automatically infer new knowledge, allowing, for example, risk evaluation to be performed based on previous inspections or sensor data analysis.

Semantic and data interoperability is key for enabling data and information exchange between NRAs and their systems, namely in the European highway industry. CEDR has helped European NRAs to obtain and manage asset information, ensuring timely and informed decision-making and reducing risks. In this industry, asset data can change ownership or be exchanged throughout its lifecycle, often being shared between different entities within and outside NRAs (which are often interdependent) (Biswas et al., 2021a, 2021b). Consequently, the integrated management of asset data from various NRAs necessarily has to solve typical integration and data quality challenges, such as lack of interconnectivity (silo view), inaccuracies, incompleteness and semantic inconsistencies. These challenges significantly increase the risk of making wrong decisions and the costs associated with information management. This is why European CEDR projects, such as Interlink and CoDEC, have sought to find open and scalable information management standards that European NRAs can use to find, manage, use, analyze and share AMS data, while allowing them to connect to other industry standards, such as IFC.

The presented approach or any future applications designed with this ontology can be used by any NRA, provided their data is mapped to the ontology. Furthermore, the use of standards and higher-level ontologies, such as eurOTL and SSN, also ensures the interoperability of the Engineering Structures ontology with these ontologies or applica-

[13] https://www.w3.org/TR/shacl/.

[14] https://www.w3.org/submissions/SWRL/.

Table 3
Cost-benefit dimensions.

Benefits	Costs
B1. Ontology-driven analysis (including knowledge validation and inference) in a familiar decision-support environment (BIM)	C1. Multidisciplinary Team Required
B2. Industry standards adoption (use of eurOTL and SSN ensures interoperability)	C2. Data Integration (including ontology mapping)
B3. Semantic and Data interoperability improve communication, data exchange, and decision-making across European NRAs	C3. Software Development
B4. Provide a base for several asset management processes across the whole asset lifecycle	C4. User Training and Adoption
	C5. Infrastructure and upkeep

tions based on them. The integration and reuse of industrial standards in a flexible way minimizes the risk of obsolescence of such solutions, extending their adoption lifetime and, consequentially, reducing operational costs in the future.

In addition, semantic web technologies can play a crucial role in facilitating compliance and governance efforts by improving external communication and alignment. Semantic web technologies streamline documentation and reporting, enabling users to verify adherence to regulatory requirements and compliance with relevant standards, such as those set by industry policymakers. Thus, the European Commission, as a policymaker, can establish a set of formalized rules, which can be translated into SWRL and SHACL, and automatically validated against the ontological representation of each asset. For example, this can be used to validate whether or not a certain type of structural element is being monitored according to a set of observable variables.

AMS support industrial practitioners in key areas, including condition assessment, forecasting future deterioration, and identifying maintenance and repair needs and strategies. Structural operators typically use visual inspections and sensor networks to monitor structural response and environmental variables (such as displacements, temperatures or energy consumption) and collect essential data, enabling effective and efficient management during the O&M phase. The Engineering Structures ontology allows users to access dynamic data, normally stored in AMS, providing a basis for the analysis of European road sector assets, namely highway structures and their structural elements. By providing reliable historical information about the assets, the ontology can be used for various asset management processes, including maintenance, condition assessment and performance prediction.

CoDEC's three pilot projects demonstrated the applicability of this ontological artifact: *PP - Tunnels* showcased the integration of sensor data in a BIM environment, *PP - Bridges* presented a risk analysis based on structural inspections, and, finally, *PP - Pavements* demonstrated the possibility of exchange data between BIM and GIS. There are no discernible barriers preventing the application of this semantic-based approach throughout the entire lifecycle of infrastructure assets. By facilitating data exchange between systems and NRAs, this approach enables data-driven decision-making across the various lifecycle processes of each structure, simplifying workflows and reducing data duplication and errors. This semantic interoperability also improves collaboration and communication among stakeholders involved in facility management, maintenance planning, and asset lifecycle management. Moreover, it reduces discrepancies and improves data reliability, leading to a more efficient and reliable decision-making, optimizing asset performance throughout the lifecycle.

However, integrating AMS data into the BIM environment through an ontology-based approach entails several important cost considerations. Firstly, the application of essential components such as the Engineering Structures ontology and the CoDEC framework requires a multidisciplinary team including domain/industry experts, stakeholders, knowledge engineers and software developers to ensure alignment with user needs.

Secondly, the costs associated with conceptual integration and data migration include efforts to map concepts between AMS and BIM (which usually requires human intervention), to transform and populate data into the ontology. In addition, although there are tools to automatically obtain an ifcOWL representation from a BIM model, it is still necessary to manually map the relationships between the entities in the Engineering Structures ontology and their BIM representation.

Thirdly, additional costs may arise from the need to develop or customize software for seamless integration between AMS and BIM systems, such as creating additional API services or customizing existing software to visualize ontological knowledge within the BIM environment. Training and adoption costs should also be considered, as users must learn to access, interpret, and use AMS data in this decision-support environment. Infrastructure and upkeep costs, as well as those related to the maintenance and evolution of the ontology (including version control), are also critical considerations.

It is worth noting that estimating project costs depends on variables such as the size of the project, the tools/software used, data requirements and forward planning. For example, recognizing the need to use BIM throughout the asset lifecycle (including O&M) can lead to the design of AMS with BIM identifiers for each asset, simplifying future ontology mappings. Although the assessment of costs and benefits will inevitably vary depending on the specific case, Table 3 provides an overview of the main dimensions for benefits and costs of using the Engineering Structures Ontology as the basis for data exchange between BIM and AMS. This table provides future stakeholders with a starting point to derive cost-benefit factors for their project-specific analysis.

8.3. Limitations and future work

The Engineering Structures ontology and its use still have limitations that require further research and development. Currently, knowledge of the ontology is limited to the primary structures associated with the CoDEC pilot projects (i.e. tunnels, bridges and pavements). Likewise, the use cases presented in this research focus on these pilot projects. Although formalized in a W3C standard, the full potential of the ontology remains to be explored. Standards such as SHACL or SWRL, capable of validating and inferring knowledge, offering new analytical possibilities, have not been addressed in this research. Furthermore, the process of populating the ontology still depends on manual intervention, including the export of operational data and the mapping between the Engineering Structures ontology and the ifcOWL instances

Future research should focus on resolving these limitations and expanding the usefulness of the Engineering Structures ontology. In particular, future research should broaden the scope of the ontology to cover a wider range of structures and elements beyond those addressed in the CoDEC pilot projects. The independent use of the Engineering Structures ontology to validate, analyze and infer knowledge in its domain is another avenue of research that could be explored. The use of W3C standards and other reasoning tools can fully exploit the ontological representation of this knowledge. In addition, it is imperative to develop automated tools or algorithms to simplify the process of populating operational data into the ontology. These tools should also facilitate the seamless mapping between the entities of engineering structures and the instances of ifcOWL, enabling real-time decision making in BIM environments.

9. Conclusion

BIM environments are used for decision support during the design, planning and construction phases. However, decision support in the operational phase is usually ensured by each NRA's AMS, supported by information related to monitoring, maintenance, and sensor data. Integrating operation data from AMS with BIM models is key to providing a real-time and continuous decision-making process throughout the complete structure lifecycle in the same (integrated) environment. The CoDEC research project proposed using semantic web and linked data principles to link operational data with the BIM environment, increasing system interoperability. A Data Dictionary was developed to provide a shared conceptualization that can be used as a base for a standard data format to enable interoperability between Europe's NRA and their systems.

This paper presents the development and evaluation of an Engineering Structures ontology used in the CoDEC project to encode this shared conceptualization. The Engineering Structures ontology represents structures, their structural elements and relationships in a formal, comprehensible and explicit way. Furthermore, the ontology can also describe sensor, and risk and condition data. The Ontology Requirements Specification Document (ORSD, see Section 5.1) is presented, following the NeON methodology, with information regarding the ontology a) goals, domain and scope, b) users, use cases and applications, c) knowledge inputs and d) competency questions.

The ontology design and development process is described, focusing on the requirements of each pilot project. The ontology was validated and evaluated by answering the competency questions defined in the ORSD and using the OOPS! tool. Lastly, the integration of this knowledge in a decision-support environment was demonstrated using a pilot-project related to risk and condition data in bridge elements, showcasing the ontology as a base for data exchange between BIM and AMS systems.

Theoretical and practical implications are presented, including an analysis on cost and benefit dimensions and applicability of this approach. Limitations are also presented in the discussion section. By addressing these limitations in future research, the Engineering Structures ontology can evolve into a more comprehensive and versatile tool for facilitating data exchange and decision-making in the context of BIM and AMS integration.

CRediT authorship contribution statement

António Lorvão Antunes: Conceptualization, Methodology, Software, Validation, Writing – original draft. **José Barateiro:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Vânia Marecos:** Investigation, Writing – original draft. **Jelena Petrović:** Resources, Software, Visualization. **Elsa Cardoso:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Funding: This work was partially supported by the Portuguese Foundation for Science and Technology [grant numbers 2021.07134.BD, UIDB/03126/2020].

The CoDEC project was carried out as part of the CEDR Transnational Road Research Programme Call 2018: BIM. The funding for the research was provided by the national road administrations of Austria, Belgium - Flanders, Denmark, Finland, Germany, Netherlands, Norway and Sweden.

References

Ait-Lamallam, S., Sebari, I., Yaagoubi, R., & Doukari, O. (2021a). Ifcinfra4om: An ontology to integrate operation and maintenance information in highway information modelling. *ISPRS International Journal of Geo-Information*, *10*(5), 305.

Ait-Lamallam, S., Sebari, I., Yaagoubi, R., & Doukari, O. (2021c). Towards an ontological approach for the integration of information on operation and maintenance in bim for road infrastructure. In *Proceedings of sixth international congress on information and communication technology: ICICT 2021, vol. 1* (pp. 701–712). Springer.

Ait-Lamallam, S., Yaagoubi, R., Sebari, I., & Doukari, O. (2021b). Extending the ifc standard to enable road operation and maintenance management through openbim. *ISPRS International Journal of Geo-Information*, *10*(8), 496.

Al-Kasasbeh, M., Abudayyeh, O., & Liu, H. (2021). An integrated decision support system for building asset management based on bim and work breakdown structure. *Journal of Building Engineering*, *34*, Article 101959.

Austroads (2019). Data standard for road management and investment in Australia and New Zealand (dsrmi). Retriebed from https://austroads.com.au/publications/asset-management/ap-r597-19. version 3.0.

Biancardo, S. A., Viscione, N., Cerbone, A., & DessìJr, E. (2020). Bim-based design for road infrastructure: A critical focus on modeling guardrails and retaining walls. *Infrastructures*, *5*(7), 59.

Biswas, S., Proust, J., Andriejauskas, T., Wright, A., van Geem, C., Kokot, D., Antunes, A., Marecos, V., Barateiro, J., Bhusari, S., et al. (2021a). *Codec: Connected data for road infrastructure asset management*. *IOP conference series: Materials science and engineering: Vol. 1202*. IOP Publishing (p. 012002).

Biswas, S., Proust, J., Andriejauskas, T., Wright, A., Van Geem, C., Kokot, D., Antunes, A., Marecos, V., Barateiro, J., Bhusari, S., et al. (2021b). Demonstrating connectivity and exchange of data between bim and asset management systems in road infrastructure asset management. In *International road federation world meeting & exhibition* (pp. 379–392). Springer.

Blomqvist, E., Hammar, K., & Presutti, V. (2016). Engineering ontologies with patterns-the extreme design methodology. *Ontology Engineering with Ontology Design Patterns*, *25*, 23–50.

buildingSMART (2020). Industry foundation classes (ifc), ifcroad. Retriebed from http://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/.

Cao, J., Vakaj, E., Soman, R. K., & Hall, D. M. (2022). Ontology-based manufacturability analysis automation for industrialized construction. *Automation in Construction*, *139*, Article 104277.

Chen, W. T., & Bria, T. A. (2022). A review of ontology-based safety management in construction. *Sustainability*, *15*(1), 413.

CoDEC project report - deliverable D3A: Pilot projects report and consolidated implementation resources. Retriebed from https://www.codec-project.eu/Resources/projectreports (2021).

Costin, A., Adibfar, A., Hu, H., & Chen, S. S. (2018). Building information modeling (bim) for transportation infrastructure–literature review, applications, challenges, and recommendations. *Automation in Construction*, *94*, 257–281.

Cursi, S., Martinelli, L., Paraciani, N., Calcerano, F., & Gigliarelli, E. (2022). Linking external knowledge to heritage bim. *Automation in Construction*, *141*, Article 104444.

D'Amico, F., Calvi, A., Schiattarella, E., Di Prete, M., & Veraldi, V. (2020). Bim and gis data integration: A novel approach of technical/environmental decision-making process in transport infrastructure design. *Transportation Research Procedia*, *45*, 803–810.

Ding, L., Zhong, B., Wu, S., & Luo, H. (2016). Construction risk knowledge management in bim using ontology and semantic web technology. *Safety Science*, *87*, 202–213.

Fang, W., Ma, L., Love, P. E., Luo, H., Ding, L., & Zhou, A. (2020). Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Automation in Construction*, *119*, Article 103310.

Farghaly, K., Abanda, F. H., Vidalakis, C., & Wood, G. (2019). Bim-linked data integration for asset management. *Built Environment Project and Asset Management*, *9*(4), 489–502.

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: From ontological art towards ontological engineering.

Gao, X., & Pishdad-Bozorgi, P. (2019). Bim-enabled facilities operation and maintenance: A review. *Advanced Engineering Informatics*, *39*, 227–247. https://doi.org/10.1016/j.aei.2019.01.005. Retriebed from https://www.sciencedirect.com/science/article/pii/S1474034618303987.

Garramone, M., Moretti, N., Scaioni, M., Ellul, C., Re Cecconi, F., & Dejaco, M. C. (2020). Bim and gis integration for infrastructure asset management: A bibliometric analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information*

Sciences, *VI-4/W1-2020*, 77–84. https://doi.org/10.5194/isprs-annals-VI-4-W1-2020-77-2020. Retriebed from https://isprs-annals.copernicus.org/articles/VI-4-W1-2020/77/2020/.

Gómez-Pérez, A., & Suárez-Figueroa, M.C. (2009). Neon methodology for building ontology networks: A scenario-based methodology.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (pp. 1–17). Springer.

Hagedorn, P., Liu, L., König, M., Hajdin, R., Blumenfeld, T., Stöckner, M., Billmaier, M., Grossauer, K., & Gavin, K. (2023). Bim-enabled infrastructure asset management using information containers and semantic web. *Journal of Computing in Civil Engineering*, *37*(1), Article 04022041.

Hartmann, T., & Trappey, A. (2020). Advanced engineering informatics - philosophical and methodological foundations with examples from civil and construction engineering. *Developments in the Built Environment*, *4*, Article 100020. https://doi.org/10.1016/j.dibe.2020.100020. Retriebed from https://www.sciencedirect.com/science/article/pii/S2666165920300168.

Highways England (2020). Asset data management manual (ADMM). Retriebed from https://www.standardsforhighways.co.uk/ha/standards/admm/index.htm. version 12.0.

ISO 19650-1:2018 (2018). Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 1: Concepts and principles. Standard, International Organization for Standardization. Geneva, CH.

ISO 16739-1:2018 (2018). *Industry foundation classes (IFC) for data sharing in the construction and facility management industries — part 1: Data schema*. Geneva, CH: Standard, International Organization for Standardization.

Jiang, S., Feng, X., Zhang, B., & Shi, J. (2023). Semantic enrichment for bim: Enabling technologies and applications. *Advanced Engineering Informatics*, *56*, Article 101961.

Kivits, R. A., Furneaux, C., et al. (2013). Bim: Enabling sustainability and asset management through knowledge management. *The Scientific World Journal*, 2013.

Koch, C., Paal, S. G., Rashidi, A., Zhu, Z., König, M., & Brilakis, I. (2014). Achievements and challenges in machine vision-based inspection of large concrete structures. *Advances in Structural Engineering*, *17*(3), 303–318.

Kokot, D. (2019). Asset management approach for transport infrastructure networks: The am4infra project. In *Airfield and highway pavements 2019: Design, construction, condition evaluation, and management of pavements* (pp. 374–381). Reston, VA: American Society of Civil Engineers.

Kukkonen, V., Kücükavci, A., Seidenschnur, M., Rasmussen, M. H., Smith, K. M., & Hviid, C. A. (2022). An ontology to support flow system descriptions from design to operation of buildings. *Automation in Construction*, *134*, Article 104067.

Lee, D.-Y., Chi, H-l., Wang, J., Wang, X., & Park, C.-S. (2016). A linked data system framework for sharing construction defect information using ontologies and bim environments. *Automation in Construction*, *68*, 102–113.

Lee, S.-K., & Yu, J.-H. (2023). Ontological inference process using ai-based object recognition for hazard awareness in construction sites. *Automation in Construction*, *153*, Article 104961.

Lei, X., Wu, P., Zhu, J., & Wang, J. (2021). Ontology-based information integration: A state-of-the-art review in road asset management. *Archives of Computational Methods in Engineering*, 1–19.

Li, X., Yang, D., Yuan, J., Donkers, A., & Liu, X. (2022). Bim-enabled semantic web for automated safety checks in subway construction. *Automation in Construction*, *141*, Article 104454.

Marcovaldi, E., & Biccellari, M. (2018). Asset data dictionary. *Deliverable D*, *3*(1).

Meschini, S., Daniele, A., Marco, A., Seghezzi, E., Tagliabue, L. C., Di Giuda, G., et al. (2022). Data integration through a bim-gis web platform for the management of diffused university assets. In *Proceedings of the 2022 European conference on computing in construction* (pp. 237–244).

Meschini, S., Pellegrini, L., Locatelli, M., Accardo, D., Tagliabue, L. C., Di Giuda, G. M., & Avena, M. (2022). Toward cognitive digital twins using a bim-gis asset management system for a diffused university. *Frontiers in Built Environment*, *8*, Article 959475.

Meschini, S., Accardo, D., Locatelli, M., Pellegrini, L., Tagliabue, L. C., Di Giuda, G. M., et al. (2023). Bim-gis integration and crowd simulation for fire emergency management in a large diffused university. In *Proceedings of the international symposium on automation and robotics in construction* (pp. 357–364).

Mora, M., Wang, F., Gómez, J. M., & Phillips-Wren, G. (2022). Development methodologies for ontology-based knowledge management systems: A review. *Expert Systems*, *39*(2), Article e12851.

Noy, N.F., McGuinness, D.L. et al. (2001). Ontology development 101: A guide to creating your first ontology.

Open geospatial consortium, semantic sensor network ontology. w3c recommendation. Retrieved from https://www.w3.org/TR/vocab-ssn/ (2017).

Pan, J. (2009). Resource description framework. In *Handbook on ontologies* (pp. 71–90).

Poveda-Villalón, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2014). Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *10*(2), 7–34.

Ratkevičiūtė, K. (2010). Model for the substantiation of road safety improvement measures on the roads of Lithuania. *The Baltic Journal of Road and Bridge Engineering*.

Schreiber, A. T., Schreiber, G., Akkermans, H., Anjewierden, A., Shadbolt, N., de Hoog, R., Van de Velde, W., Wielinga, B., Nigel, R., et al. (2000). *Knowledge engineering and management: The CommonKADS methodology*. MIT Press.

Schreiber, G., Wielinga, B., de Hoog, R., Akkermans, H., & Van de Velde, W. (1994). Commonkads: A comprehensive methodology for kbs development. *IEEE Expert*, *9*(6), 28–37.

Staab, S., Studer, R., Schnurr, H.-P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, *16*(1), 26–34.

Stephan, G. s., Pascal, H. s., & Andreas, A. s. (2007). Knowledge representation and ontologies. In *Semantic web services: Concepts, technologies, and applications* (pp. 51–105).

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1–2), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6.

Su, X., & Ilebrekke, L. (2002). A comparative study of ontology languages and tools. In *International conference on advanced information systems engineering* (pp. 761–765). Springer.

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernandez-Lopez, M. (2015). The neon methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, *10*(2), 107–145.

Talebi, S. (2014). Exploring advantages and challenges of adaptation and implementation of bim in project life cycle.

Wang, M. (2021). Ontology-based modelling of lifecycle underground utility information to support operation and maintenance. *Automation in Construction*, *132*, Article 103933.

Wijeratne, P. U., Gunarathna, C., Yang, R. J., Wu, P., Hampson, K., & Shemery, A. (2024). Bim enabler for facilities management: A review of 33 cases. *International Journal of Construction Management*, *24*(3), 251–260.

Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M., & Wang, X. (2021). Ontological knowledge base for concrete bridge rehabilitation project management. *Automation in Construction*, *121*, Article 103428.

Wu, Z., Cheng, J. C., Wang, Z., & Kwok, H. H. (2023). An ontology-based framework for automatic building energy modeling with thermal zoning. *Energy and Buildings*, Article 113267.

Zhao, L., Liu, Z., & Mbachu, J. (2019). Highway alignment optimization: An integrated bim and gis approach. *ISPRS International Journal of Geo-Information*, *8*(4), 172.

Zheng, Y., Törmä, S., & Seppänen, O. (2021). A shared ontology suite for digital construction workflow. *Automation in Construction*, *132*, Article 103930.

Zhong, B., Gan, C., Luo, H., & Xing, X. (2018). Ontology-based framework for building environmental monitoring and compliance checking under bim environment. *Building and Environment*, *141*, 127–142.

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, *101*, 17–31.

Zhou, Y., Bao, T., Shu, X., Li, Y., & Li, Y. (2023). Bim and ontology-based knowledge management for dam safety monitoring. *Automation in Construction*, *145*, Article 104649.

CHAPTER 4

Journal Article 3

This chapter describes the design and development of the Strategy Ontology (artifact #3), denominated Balanced Scorecard Ontology (BSO). This artifact was developed to bridge the gap between strategy management and data within the Balanced Scorecard framework. The ontology is demonstrated using an existant BSC of a public university library, evaluated using competency questions, and further validated by an online validation tool. As shown in Figure 4.1, this publication (JA3) reports on contributions related to the DSRM's third iteration (Iter. 3).

Article details:

- **Title**: The Balanced Scorecard Ontology: A Semantic Approach to Enhance Strategy Management;
- **Status**: Awaiting approval, in review process (minor reviews submitted/answered in January 2024);
- **Journal**: IEEE Engineering Management Review;
- **Publisher**: IEEE.

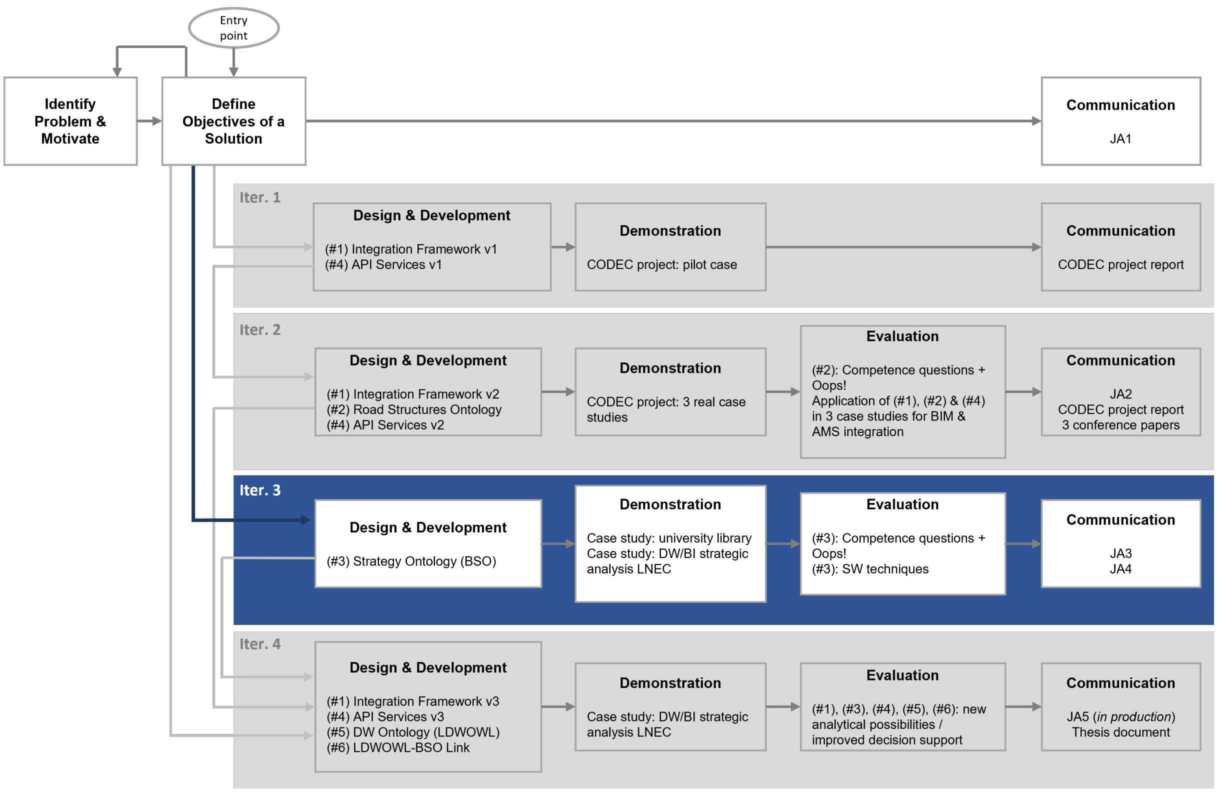


FIGURE 4.1. DSRM's JA3 Communication.

The Balanced Scorecard Ontology: A Semantic Approach to Enhance Strategy Management

Abstract—**The Balanced Scorecard, developed in 1992 by Kaplan and Norton, has evolved into a communication and strategy execution system widely adopted by organizations across various industries. This article explores the use of an ontology to bridge the gap between strategy management and data within the Balanced Scorecard framework. The Balanced Scorecard Ontology is introduced to store, validate, and analyze knowledge, containing information about the Strategy Map and Quantification Frameworks, essential for evaluating the strategy execution. The proposed ontology is designed, developed, and evaluated using competency questions, and further validated by an online tool. Specifically, the proposed formalization of the Balanced Scorecard framework provides a semantic layer aimed at facilitating an effective Balanced Scorecard implementation, enabling accurate, traceable, and continuous monitoring and improvement of the strategy execution, based on a data-driven approach. The formalization of this knowledge through an ontology encompasses several advantages, such as improved interoperability and validation of the framework's elements, inference of new knowledge, and enhanced communication between different stakeholders. Additionally, managerial implications include ensuring alignment between the Balanced Scorecard and organizational goals, supporting compliance and governance efforts, improving communication and knowledge transfer, enhancing the strategic decision-making process, and facilitating the integration of data into the Balanced Scorecard.**

Index Terms—**Balanced Scorecard, Ontology, Strategy, Strategy Map, Quantification Framework**

I. INTRODUCTION

The Balanced Scorecard (BSC) was developed in 1992 by Robert S. Kaplan and David P. Norton as a performance management system to support problem-solving and decision-making [1]. Initially, the BSC divided measures into four perspectives: Financial, Customer, Internal Processes, and Learning & Growth. This complementary set of measures was presented to business users as "dials and indicators in an airplane cockpit," allowing for a comprehensive view of past results, current operational performance, and, at the same time, monitoring future drivers.

The BSC has evolved significantly since its creation in the early 1990s, with many organizations adopting and adapting it to fit their specific needs and objectives [2]–[4]. Today, the BSC is seen as a communication and strategy execution system [5], [6]. It has been shown to improve organizational performance, enhance strategic alignment, and facilitate communication and coordination across different departments and levels of an organization. The BSC has been successfully applied in many industries, including Higher Education [7], [8], Healthcare [9]–[11], and Tourism [12]. Recent research has also explored the potential of the BSC to promote sustainability and corporate social responsibility by incorporating environmental and social measures [13], [14].

Combining the BSC with other systems and tools can lead to a more effective implementation [5]. Supino, Barnabè, Giorgino, *et al.* [15] enhanced the application of a BSC by integrating System Dynamics to improve decision-making and help in strategy formulation and implementation. Tawse and Tabesh [6] state that "the BSC has the potential to improve organizational performance, but to realize that potential, it must be effectively implemented." The authors provide three recommendations: (1) The development of a strategy map to ensure that BSC elements are causally linked; (2) Ensure Top Management Team commitment and support; and (3) Improve key stakeholder engagement through participation and frequent communication. Knowledge formalization techniques, such as ontologies, can be used to represent and make knowledge machine-readable and support the decision-making process [16], [17]. By formalizing BSC knowledge, interoperability between systems and the BSC could be improved, BSC elements and their relationships can be validated, new knowledge can be inferred, and lastly, ontology semantics can be used to enhance communication and reduce misunderstandings.

By an effective implementation of a BSC we mean that the BSC must enable an accurate, traceable, and continuous monitoring and improvement of the strategy execution, based on a data-driven approach. Since the early 2000's, authors have defended the importance of a quantitative and financial calculus when validating the BSC's strategic assumptions or hypotheses modeled using the cause-and-effect relationships [18]. However, to our knowledge, the BSC model has not evolved conceptually to incorporate these 'technical' validations, remaining primarily a 'business'-oriented strategic management approach. Organizations already use different management systems to retrieve, store, and analyze data. The technical-side implementation of data-driven decision-making has evolved in the last decades. Business Intelligence (BI) and Analytics systems have been used for data-driven decision support since the 1990s [19], [20], and there are currently industry guidelines or best practices that can be used to implement these systems (e.g., Data Warehouse and BI Systems [21] or Data Mining [22]).

The Execution Premium Process (see Figure 1) was presented in Kaplan and Norton [4], outlining key steps for effectively implementing a BSC, clearly stating the use of BI to facilitate the data optimization phase ("Monitor and Learn" and "Test and Adapt"). This article proposes a technological and data-driven approach that formalizes the BSC model, bridging the gap between strategy definition and data-driven decision-making through a comprehensive Business Intelligence implementation. Particularly, the proposed semantic layer aims to support the integration, alignment, and traceability between strategic models and the organizational

Fig. 1. Execution Premium Process. Retrieved from Khakbaz and Hajiheydari [24]

information systems necessary for providing data to the BSC's performance indicators. In today's fast-paced business environment, organizations are often forced to continuously adapt to changes, which may lead to a misalignment between the planned and executed strategies. This reinforces the need and relevance of establishing traceability and monitoring capabilities between strategic models and organizational information systems [23].

To this end, this article presents an ontology to store and analyze knowledge related to the BSC. The Balanced Scorecard Ontology (BSO) is introduced, containing information about the BSC's Strategy Map and Quantification Frameworks used to evaluate the strategy execution. The ontology is validated and evaluated using competency questions and an online tool designed to identify pitfalls in ontology development. The remainder of the article is structured as follows: Section 2 presents background research concerning ontologies, strategic models, and balanced scorecards; Section 3 introduces other existing BSC ontologies; Section 4 formalizes the BSC framework for this research's scope; The design and development of the BSO is presented in Section 5, and the ontology is validated and evaluated in the following section (Section 6); Lastly, conclusion and future work is presented in Section 7.

II. Background

This section describes the background concepts necessary for this research: balanced scorecards, strategic models, and ontologies.

A. Balanced Scorecard

The Balanced Scorecard was first introduced by Robert S. Kaplan and David P. Norton in a 1992 Harvard Business Review article [1]. In this article, Kaplan and Norton argued that traditional financial measures did not provide a complete picture of an organization's performance. They proposed using a more balanced set of measures, including financial and non-financial metrics, to better reflect an organization's performance.

Over the years, the Balanced Scorecard has evolved from a performance measurement tool to a strategic management system. In 1996, Kaplan and Norton published another article [2] that emphasized the importance of using the Balanced Scorecard to align an organization's strategy with its performance measures and to drive continuous improvement. The authors further expanded on the strategic management aspects of the Balanced Scorecard. They introduced the concept of strategy maps, a visual representation of an organization's strategic objectives and the cause-and-effect relationships between them [25]. Strategy maps help organizations to better understand how their objectives are interconnected and how they can best allocate resources to achieve their goals. The authors argue that the BSC is "agnostic to the formulation model used," [3] meaning that any business strategy formulation may be executed and communicated utilizing the BSC and its elements.

The BSC should be cascaded to align all levels of the organization to its strategy. This means that the organizational or corporative level BSC is translated to lower tiers of the organization (such as departments, teams, or individuals), with objectives and indicators becoming more specific or detailed as the BSCs are cascaded down. This vertical alignment creates an outlook between the employees and the high-level strategy, clarifying how each strategy level contributes to achieve organizational success and how they help in realizing the organization's vision [4].

B. Strategic Models

The definition of a Business Strategy is essential for any entity to achieve its goals and vision, guiding the decisions to obtain a competitive advantage against the competition. Porter's Five Forces, Blue Ocean Strategy, and the Business Model Canvas are some of the models that can be used to formulate a strategic approach, clarify the business model and help to define a BSC.

Porter states that the "essence" of strategy formulation is to define how to adapt and stay competitive against your competition [26]. Porter presents five fundamental forces that can change an industry's competition state, from which companies must defend or influence to achieve long-run profitability. Possible entrants to the industry, the power of suppliers and buyers, the arrival of substitute products, and the existing competition within the industry must be analysed and monitored to ensure that the company's advantage is achieved and defended.

The Blue Ocean Strategy [27] looks for an unknown market space where competition is non-existing. To do so, it is necessary to create a new value curve, where we look to eliminate, reduce or raise some factors in an existing industry or create something new to the industry. This leads to cost reduction and added (or new) value for the customers, allowing the business to keep existing customers and attract new ones.

The Business Model Canvas (BMC) [28] simplifies the business concept, by clarifying the organization methods and functions and developing an agile strategy definition framework. The BMC design includes the identification of customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships and cost structure as the main building blocks for the "rationale of how an organization creates, delivers, and captures value [29].

The customer value proposition defines how a company creates value for its customers to increase customer acquisition, satisfaction, and retention. Treacy and Wiersema [30] studied how various industry leaders achieved a dominant market position, and discovered that this could be achieved by increasing the focus on customer intimacy, operational excellence, or product leadership. They then proposed that a company should strive to stand out by performing exceptionally in one of the three proposals, while maintaining the industry's minimum threshold on the other two. This model was used by Kaplan and Norton [31] to structure the strategic objectives definition in the BSC customer perspective, in terms of three very different strategies: Best Total solution (customer intimacy), Best Buy (operational excellence), and Best Product (product leadership).

Osterwalder, Pigneur, Bernarda, *et al.* [32] proposed another value proposition model, aligned with the BMC [28], called the Value Proposition Canvas (VPC). This model helps a company to design a product or service aligned with the customers' wants and needs. Given the Customer Profile (defined in terms of the the jobs customers are trying to get done, the gains they expect to achieve, and the negative impacts (or pains) they might suffer), the goal is to define an aligned Value Map. This component defines the main characteristics of the product/service offered to help the customer to complete its jobs, demonstrating how the company intends to create the expected gains, and relieve the pains. This value proposition model is not referenced in Kaplan and Norton's work. However, we have been using it for almost ten years in university-level business and information systems classes to design BSC, as shown in works such as Silva [33], Cardoso, Santos, Costa, *et al.* [34], and Sacoor, Arsenio, Cardoso, *et al.* [35]. We have found that the VPC enables a richer strategy definition for the customer perspective. Moreover, Treacy and Wiersema [30] focused on industry leaders, while the VPC can be applied to any company, even a startup, and to a strategy that does not aim simply to gain a dominant market position.

C. Ontologies

Ontologies are "formal, explicit specifications of shared conceptualizations" [16]. They are used to describe knowledge about a certain domain of interest, its concepts, properties, and relationships. Ontologies are used to share, reuse and analyze knowledge, facilitating interoperability and heterogeneity [36], which is why they are an integral part of the Semantic Web[1]. Knowledge Base refers to an ontology populated with individual instances [37].

Resource Description Framework (RDF) is a World Wide Web Consortium (W3C) recommendation to "create, exchange and use annotations on the Web". The resources are described in the form of triples (*subject property object*) [38]. RDF Schema (RDFS) provides a vocabulary for RDF, introducing class and hierarchy concepts. The Ontology Web Language (OWL) was developed on top of RDFS, adding disjointness, cardinality, object and data properties, and other additional vocabulary and expressiveness. There are three OWL sub-languages/types: Lite, DL, and Full, with different levels of expressiveness. The choice of a language depends on the problem domain and modeling requirements, with an identified trade-off between expressiveness and inference capabilities (reasoning) [39].

III. Balanced Scorecard Ontologies

A set of works was retrieved from the Web Of Science Core Collection[2] using the search query: **"Balanced Scorecard" (All Fields) AND (Ontology OR Ontologies OR "Semantic Web" OR "Knowledge Base" OR "Knowledge Representation" OR "Ontological Model") (All Fields)**. The filter Languages = (English) was the only additional filter used. The results were added to VosViewer[3] where an analysis of keywords co-occurrence was performed on the bibliographic data (see Figure 2). Note the importance of benchmarking, agency and ontology, and the connection between the strategy, performance, and the Balanced Scorecard.

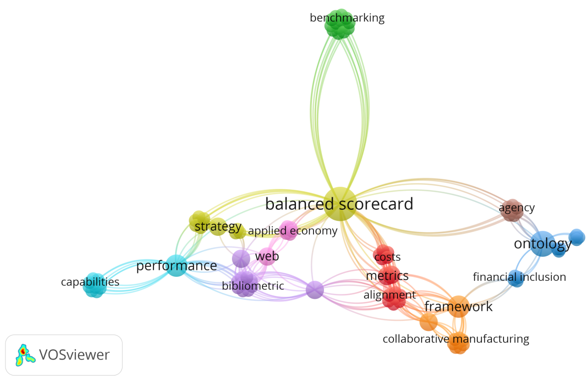


Fig. 2. VosViewer Network Visualization for BSC Ontologies Research

The 18 publications were published between 2002 and 2022, from which full text concerning three works were unavailable. From the 15 available works, only three presented original ontologies related to the BSC framework. Tables I and II present a summary of the ontologies found in these publications, analyzing which BSC elements were mapped into the ontology, the primary objective presented for the ontology development and information about other ontologies used (linked ontologies).

The Balanced Scorecard Ontology (BSCO) [42] was developed to "achieve a conceptualization of the business processes, aligned to the strategy of the organization, to be captured, represented, disseminated and processed by the people and

[1] https://www.w3.org/standards/semanticweb/

[2] www.webofscience.com

[3] https://www.vosviewer.com/

TABLE I
BSC ONTOLOGIES

Work	Ontology	Objective	Linked Ontologies
Hartanto, Sarno, and Ariyani [40]	Warning Criterion Ontology (WCO)	"Detect the wrong pattern and wrong resource in the organization"	BSCO, WCO-Master, Petri net
Bobillo, Delgado, Gómez-Romero, *et al.* [41]	Fuzzy Balanced ScoreCard Ontology (fBSCO)	Integrate fuzzy logic with BSC methodology	FKRO
Navarro-Hernandez, Perez-Soltero, Sanchez-Schmitz, *et al.* [42]	Balanced Scorecard Ontology (BSCO)	Link BSC to Business Models	eBMO[43]

software systems". The ontology allows the definition of Objectives, Initiatives, Perspectives and Measures (see Figure 3). No information is given regarding relationships between these entities.

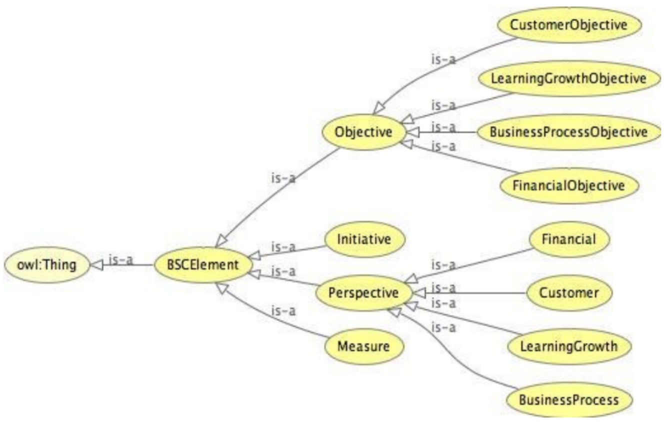


Fig. 3. Balanced Scorecard Ontology (BSCO). Retrieved from Navarro-Hernandez, Perez-Soltero, Sanchez-Schmitz, *et al.* [42]

Warning Criterion Ontology (WCO) [40] represents some BSC concepts, such as Cascading (on organizational units and employees) and defines relationships to some BSCO entities (e.g., Activities to BSCO indicators). Lastly, the Fuzzy Balanced Scorecard (fBSC) ontology [41] utilizes fuzzy logic to deal with uncertainty in BSC variables. However, the ontology is focused on Perspectives and indicators (Variables), with no information regarding the remainder of BSC elements.

None of the ontologies found during the literature review were able to represent all, or most, of the identified BSC elements, and available online. Therefore, a new ontological model can be developed to achieve the goals of this work.

TABLE II
BSC ONTOLOGIES ELEMENTS
(✓: FULLY MAPPED, P: PARTIALLY MAPPED AND -: THERE'S NO INFORMATION ABOUT THE MAPPING OF THIS ELEMENT)

Ontology	Mission, Vision, Values	Perspectives	Strategic Objectives	Strategic Theme	Strategy Map	Performance Indicators	Targets	Initiatives	Cascading
WCO	-	p	-	-	-	p	p	-	✓
fBSCO	-	✓	✓	-	-	p	-	-	-
BSCO	-	✓	✓	-	-	p	-	✓	-

IV. FORMALIZING THE BALANCED SCORECARD FRAMEWORK

Over the years, Kaplan & Norton have refined an adaptable tool that enables executives and managers to tailor and employ their BSC with the detail needed to define their strategy [1]–[4], [25]. According to Speckbacher, Bischof, and Pfeiffer [44], Niven [45], and Lawrie and Cobbold [46], a first-generation or type I BSC only needs to contain financial and non-financial indicators grouped by the four perspectives to support strategic performance management (see Table III). Authors also concur that in order to advance a type I BSC to a type II BSC, it is necessary to define a strategy map. However, various approaches are found in the literature for achieving a Type III BSC. Cardoso [47] expands on the definition of Speckbacher, Bischof, and Pfeiffer [44] and states that a Type III BSC involves the integration the different management systems already in use by the organization. This type of BSC requires the use of Business Intelligence techniques, providing analytical capabilities to monitor the strategy execution. Following this definition, and for this work's scope, a Type III BSC is defined as a system for communicating and implementing the strategy that is fully integrated with all other systems.

At the end of the last century, the original authors of the BSC recommended the use of cause-effect relationships, which is necessary to achieve a type II or second generation BSC. Nevertheless, recent studies [6] still feel the need to recommend using a strategy map (a BSC component that displays these relationships) to implement a BSC effectively. To formalize a BSC, it must be clear what components and elements are needed to maximize the benefits of the BSC framework as a strategic management system. This section defines the framework elements used in this work and how they relate, based on the work developed in Cardoso [47].

Two major components are needed to define a BSC at any strategic level: a Strategy Map and a Quantification framework (see Figure 4). A Strategy Map presents the long-term view of the strategy: the strategy statement, the main objectives, and how they are organized, while the Quantification Framework offers a shorter-term view containing the tangible indicators, goals, and initiatives needed to translate the strategy into operational terms.

A. Strategy Map

In a BSC, the Strategy Map provides a visual representation of the long-term strategy, which is a value-creation roadmap. This component contains the set of strategic objectives and displays the cause-effect relationships needed to clarify how each

TABLE III
BALANCED SCORECARD TYPES

Work	Speckbacher, Bischof, and Pfeiffer [44]	Niven [45]	Lawrie and Cobbold [46]
1st Generation / Type I	"A specific multidimensional framework for strategic performance management that combines financial and non-financial strategic indicators"	"Utilized almost exclusively to capture and analyze financial and non-financial measures across the four perspectives."	Combination of an integrated set of performance indicators and measures (financial and non-financial), grouped into four perspectives. Focused on performance evaluation.
2nd Generation / Type II	"A Type I BSC that additionally describes strategy by using cause-and-effect relationships"	Addition of strategic objectives to provide a context for selecting measures, resulting in the development of strategy maps. Furthermore, this generation introduced cause-and-effect modeling.	Identification of key business factors (key performance indicators) and their causal interrelations, materialized in the Strategy Map. Focused on performance management.
3rd Generation / Type III	"A Type II BSC that also implements strategy by defining objectives, action plans, results and connecting incentives with BSC"	The BSC requires a destination statement, with a quantitative detail, of what the future aspect of the organization should lool like at a certain date	Focused on Strategic Alignment and Change Management Support

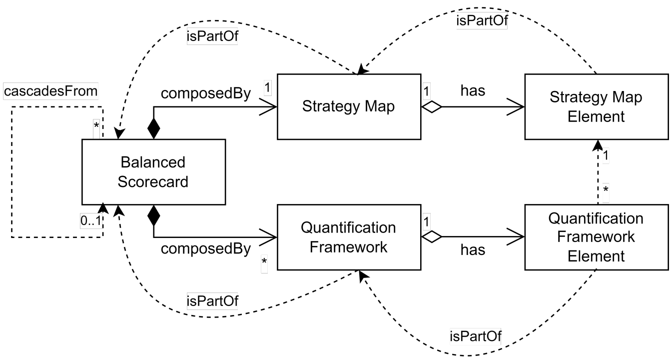


Fig. 4. Balanced Scorecard Components

strategic objective contributes to the execution of the strategy. Other elements present in a strategy map are perspectives, strategic themes (which group objectives in a set of cause-and-effect relationships, coherently showing how to achieve the strategic theme) that can be used to decompose the vision statement. This vertical use of strategic themes is aligned with the most recent contribution of Kaplan and Norton regarding this topic Kaplan and Norton [3], [4]. The main elements of the Strategy Map are presented in Table IV-A, and their primary relationships are shown in Figure 5.

As a long-term strategy tool, strategy statement elements, such as vision, mission, and values, should also be considered part of the strategy map (although not always part of the visual representation). While an organization's Mission typically remains unchanged over time, the Vision statement is normally a three to five-year concise, inspiring, and realistic (medium/long-term) goal. The BSC is intended to serve as a roadmap, guiding organizational endeavors towards attaining this desired position within the specified timeframe and niche. The Vision should include a well-defined stretch goal, establishing the performance indicator and a target value to assess the success of the vision's realization.

TABLE IV
STRATEGY MAP ELEMENTS DESCRIPTION

Element	**Element Description**
Perspective	Perspectives divide the BSC into different views. The standard perspectives are Financial, Customer, Internal Process, and Learning & Growth.
Strategic Objectives	Strategic objectives are used to break down strategy into actionable steps, operationalizing the strategy. They should be concise and quantifiable, mapping how the organization can achieve its Vision;
Strategic Themes	Major strategic forces or high-level areas of action, covering the different perspectives. The Vision is usually decomposed to obtain these themes;
Mission	The mission statement defines the purpose of an organization, i.e., the reason for its existence;
Vision	A concise, inspiring, visionary and realistic objective statement for the medium/long term goals. All organizational efforts should be made to achieve this desired position. A Vision must have a time period, a stretch goal, and a niche (aligned with the latest recommendation by Kaplan and Norton [4]);
Values	Organizational values define the guiding principles for the day-to-day employee behaviour, decisions and interactions;
Stretch Goal	Defines the target value related to a performance indicator with a clear timeframe to achieve it, enabling a clear quantification of the vision statement.

B. *Quantification Framework*

A Quantification Framework provides a short-term view of the strategy execution and concerns a defined time interval, usually a year, meaning that a set of Quantification Frameworks is expected to be defined for a strategy map in a BSC project. The main elements of this component are presented in Table V and their primary relationships are shown in Figure 6.

The central element of a Quantification Framework is the Performance Indicator. Performance indicators are used to monitor and evaluate a specific strategic objective and can be divided into Lead (drivers, enablers, predictive) or Lag (results) indicators. The relationship between objectives and indicators (see Figure 7) ensures the connection between a strategy map and its quantification frameworks inside a BSC.

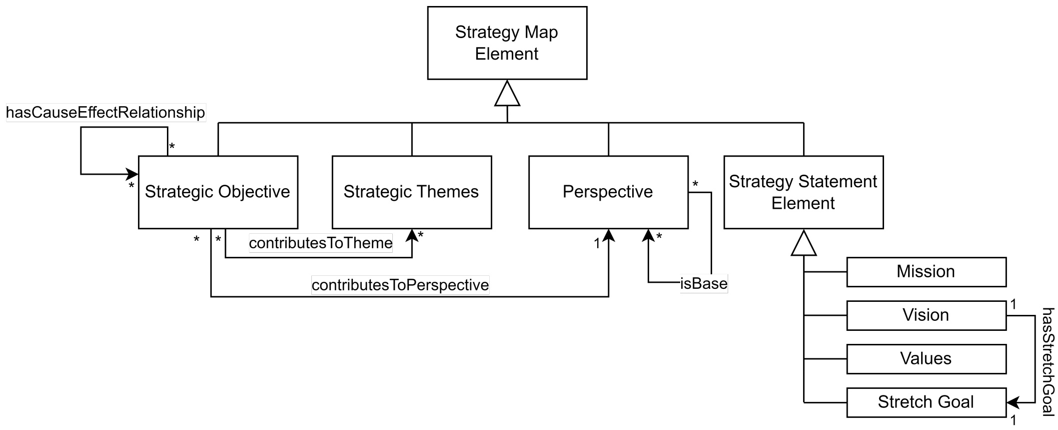


Fig. 5. Strategy Map Elements and their main relationships

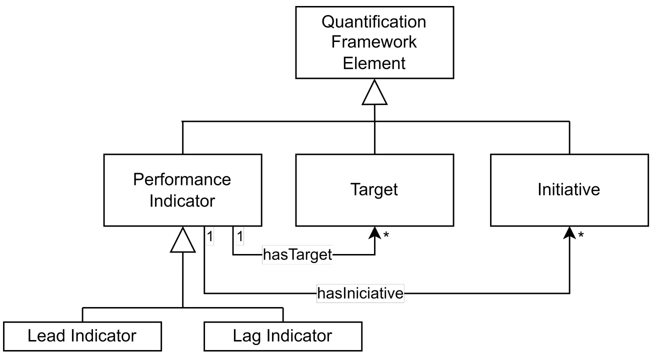


Fig. 6. Quantification Framework Elements and their main relationships

Each performance indicator must have a set of associated metadata attributes, for example: a frequency (e.g., quarterly), polarity, unit type (e.g., percentage), calculation formula, and other information related to the data origin (source, quality, collector). These attributes are generally associated with performance indicators templates [45], describing mandatory and optional attributes.

Each Performance Indicator should have a well-defined Target, indicating the desired future state to be achieved within a specific time interval. Additionally, a set of Initiatives must be identified to provide actionable plans directly impacting these indicators.

C. Cascading the BSC

Balanced Scorecards should be defined throughout the organizational levels, allowing the managers to define strategy at the corporate, department, team, or even at the individual level. Information needs are distinct, as is the level of detail (or summarization) of performance indicators and data.

As noted, a corporate or enterprise-level BSC should consist of a well-defined Strategy Map and Quantification Frameworks to effectively execute its strategy. However, the strategic elements within an corporate-level BSC, such as strategic objectives and performance indicators, are likely impacted by the corresponding elements at lower levels of detail, which should be defined by BSCs at lower hierarchical levels. Conceptually, a BSC is the sum of all the BSCs defined at different organizational levels, from the corporate level (if this is the highest level at which it has been defined) to the lowest level of cascading.

TABLE V
QUANTIFICATION FRAMEWORK ELEMENTS

Element	Element Description
Performance Indicators (KPIs)	Performance indicators are used to monitor and evaluate the strategic objectives' state or fulfillment. Key Performance Indicators (KPI) are highly aggregated metrics that assess critical organizational aspects. Performance Indicators can be divided into lead (enablers or predictive) and lag (results);
Targets	Targets establish objective goals for each indicator, by defining a "value and time" pair. These targets identify value gaps between the current reality of an organization and its desired future state;
Strategic Initiatives	Strategic initiatives are projects with a defined priority that have a direct impact on a set of indicators;

Regarding the BSC elements, two types of cascading have been identified (see Figure 8). An element within the framework may be the same as another element at a lower level of detail (for example, a corporate indicator or objective with a specific filter/focus on a singular department, represented by the "isDecompositionOf" relationship). Alternatively, it may be a distinct element but share a cause-effect relationship (such as an individual-set objective contributing to a department-level objective, represented by the "hasCauseEffectRelationship" relationship).

V. BALANCED SCORECARD ONTOLOGY

This section presents the main contribution of this work, which is the development of the Balanced Scorecard Ontology (BSO). The On-to-Knowledge methodology [48] was utilized to outline the necessary activities for the ontology's development process. Below, the Ontology Requirements Specification Document is presented. Subsequently, the development

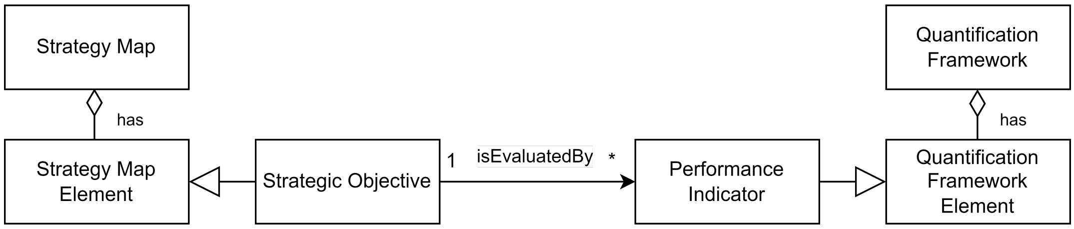


Fig. 7. Relationship between Strategic Objective and Performance Indicator

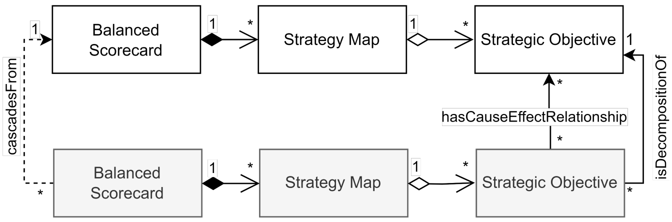


Fig. 8. Cascading at BSC and Strategic Objective level. The BSC represented in grey cascades from the white BSC.

process is discussed, highlighting the major decisions taken throughout the process.

A. *Ontology Requirements Specification Document*

*1) **Domain and Scope**:* The BSO was developed to describe and store knowledge related to the Balanced Scorecard framework, following the formalization presented in Section IV, which divides the BSC into a long-term view (Strategy Map) and a shorter-term view focused on strategy execution (Quantification Framework). The ontology must be able to describe at least a Type II BSC (see Section II-A).

*2) **Goals**:* The ontology should represent, provide information and allow inference on BSC components, specifically the Strategy Map and Quantification Framework, the BSC elements, such as Strategy Statement elements (Vision, Mission, and Values), Strategic Objectives, Perspectives, Themes, and Performance Indicators, and the relationships between these elements (e.g., cause-effect between objectives).

*3) **Users, Use Cases and Applications**:* The BSO should allow any organization and manager to formalize, translate, communicate, align, and execute its strategy. The ontology should also allow for strategy validation (e.g., ensure every strategic objective has a performance indicator) and improve interoperability between performance management systems and strategy.

*4) **Knowledge Sources and Reusable Ontologies (Inputs)**:* The BSO was based on Kaplan & Norton's work [1]–[4], [25] and the formalization presented in Cardoso, Trigueiros, *et al.* [7]. Descriptions were based on Niven [45]

*5) **Competency Questions**:* Table VI presents the main competency questions (CQ) for which the ontology must provide answers. However, it is essential to recognize that this set of CQ is not exhaustive. These questions aid in defining the ontology's scope, identifying core concepts and relationships, and ensuring completeness within the representation of domain

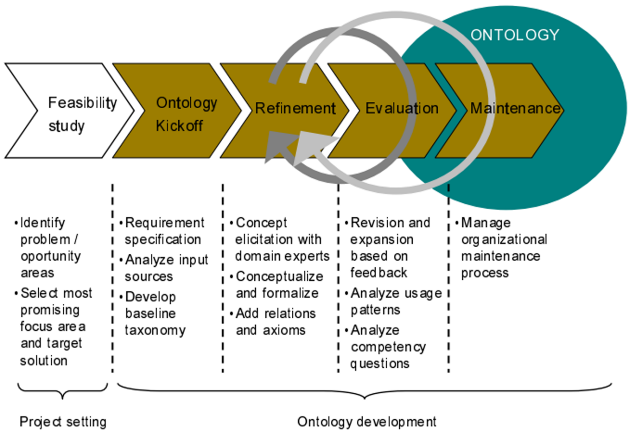


Fig. 9. On-to-Knowledge Methodology. Retrieved from Staab, Studer, Schnurr, *et al.* [48]

knowledge. While CQ are valuable guides for ontology development, they do not cover every possible scenario or nuance within a domain. In this case, CQ were defined to ensure that the ontology correctly represents a BSC, while some CQ, such as CQ8, were defined to exemplify the use of the BSO in new knowledge extraction and inference.

B. *Ontology Development*

The ontology was developed following the specification presented in Section IV. Figure 10 presents BSO's class hierarchy, including the Balanced Scorecard, its Components, and its Elements. Each class is annotated using a label (rdf:label) and/or a description (dc:description[4]).

Object and data relationships were also created. From a structural point of view, the Balanced Scorecard is composed by a set of Components ("**hasComponent**") which in turn have a set of BSC Elements ("**hasElement**"). The Strategy Statement Elements are related to each BSC using the relationships "**hasMission**", "**hasVision**" and "**hasValue**". The Vision class is defined as Strategy Statement Element with a defined deadline (represented as a xsd:dateTime[5] using the data property "**hasTimeFrameEnd**") and a Stretch Goal. This Stretch Goal is related to a Performance Indicator and must have a defined target (stated using the data property "**hasValue**").

[4]dc: Dublin Core Metadata - https://www.dublincore.org/
[5]xsd: XML Schema Definitions

TABLE VI
COMPETENCY QUESTIONS

Balanced Scorecard	
CQ1	What are the strategy statements associated with a certain BSC?
CQ2	What is the ”time horizon” associated with a certain BSC?
CQ3	What is the strategic level of a certain BSC?
Strategy Map	
CQ4	How many objectives are part of a Strategy Map of a certain BSC?
CQ5	Which Perspectives or Themes are used in a certain Strategy Map?
CQ6	How are Perspectives related in a certain Strategy Map?
Strategic Objectives	
CQ7	What are the Perspective and Themes of a certain Strategic Objective?
CQ8	Which objectives are directly or indirectly impacted by a certain Strategic Objective?
CQ9	Which Performance Indicators are used to evaluate a certain Strategic Objective?
Performance Indicators	
CQ10	Is a certain indicator a lag or lead indicator?
CQ11	What is the Unit type/Frequency/Polarity of a certain indicator?
CQ12	What is the Formula/Data Source/Data Quality of a certain indicator?
CQ13	Which targets are defined for a certain indicator?
Strategy Execution	
CQ14	Which initiatives are planned, and which performance indicators do they impact?
CQ15	What is the latest value for a certain performance indicator? And which is the next target?
Cascading	
CQ16	How is a certain BSC cascaded?
CQ17	Which are the Strategic Objectives within the cascaded Balanced Scorecards that impact a certain objective, either through decomposition or cause-effect relationships?

The Balanced Scorecard must also have a strategic level (”hasStrategicLevel”).

The focal point of a Strategy Map are Strategic Objectives and their contributions to other elements in the Strategy Maps, namely the Perspectives, Strategic Themes, and other Strategic Objectives. To formalize these relationships, the following object properties were created as a sub-property of ”**contributesTo**”:

1) **contributesToPerspective** - Direct contribution from a Strategic Objective to a Perspective (functional);
2) **contributesToTheme** - Direct contribution from a Strategic Objective to a Strategic Theme;
3) **hasCauseEffectRelationship** - Direct contribution from a Strategic Objective to a Strategic Objective in a Strategy Map;
4) **isDecompositionOf** - Contribution from a Strategic Objective to another in a higher level of detail.

Each Strategic Objective is evaluated by a set of Performance Indicators, which is formalized using the relationship ”**isEvaluatedBy**”. Each indicator can be characterized by a group of data properties related to data sources, quality, and formula, as well as indicator frequency, polarity, and unit type (e.g., percentage), among others. Using the ”hasTarget” and ”hasIniciative”, a Performance Indicator can be related to a Target or a Strategic Initiative, respectively. A Target must have a defined deadline. Lastly, an Actual Value related to the execution of a Performance Indicator is formalized using the ”hasActualValue” relationship. The Actual Value must have a certain value (using the data property ”hasValue”) related to a certain time window (”hasTimeFrame”).

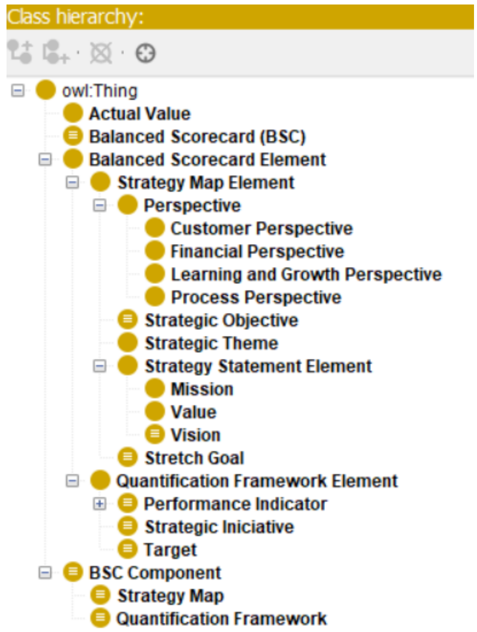


Fig. 10. BSO Class Hierarchy

VI. EVALUATION AND VALIDATION

This section presents the ontology evaluation process. Following the proposed methodology, the BSO is analyzed regarding the defined competency question. To achieve this, the ontology was previously populated, which is described below. Common pitfall detection is also realized using a well-known online tool.

A. Ontology Population Wand Case Study

Instance data was added to validate and evaluate the Balanced Scorecard Ontology. The process of adding instances to the ontology (A-box statements) is called ontology population, which was accomplished using a Protégé plugin called Cellfie[6]. Cellfie was used to define a set of import rules and mappings (based on Manchester OWL Syntax[7]) from Excel spreadsheets into OWL axioms (see Figure 11).

Strategy information was based on a public scorecard from a library repository of a higher education faculty [33]. Information related to Strategic Objectives, the cause-and-effect relationships between them, themes, and perspectives were available, as shown in A. Missing information was later supplemented, mainly information concerning indicators and

[6]https://github.com/protegeproject/cellfie-plugin
[7]https://www.w3.org/TR/owl2-manchester-syntax/

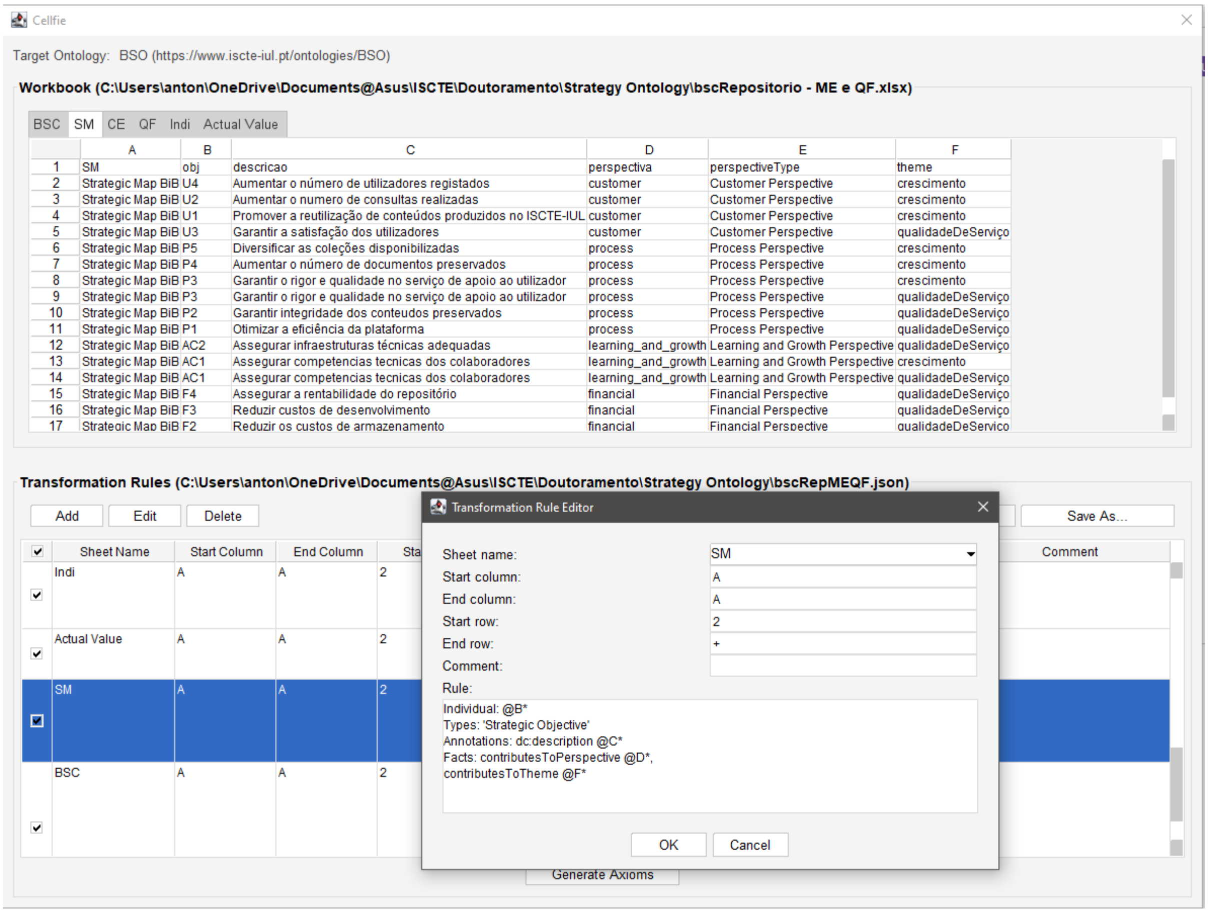


Fig. 11. Cellfie Rule Example

execution values. In the end, instance information is available for querying inside the Protégé tool.

B. *Ontology Evaluation: Competency Questions*

In this section, the BSO will be used to answer the Competency Questions defined in the Ontology Requirements Specifications Document (see Section V-A). Due to space limitations, the following CQ were selected to demonstrate the ontology:

TABLE VII
COMPETENCY QUESTIONS

CQ1	What are the strategy statements associated with a certain BSC?
CQ2	What is the "time horizon" associated with a certain BSC?
CQ5	Which Perspectives or Themes are used in a certain Strategy Map?
CQ8	Which objectives are directly or indirectly impacted by a certain Strategic Objective?
CQ13	Which targets are defined for a certain indicator?
CQ15	What is the latest value for a certain performance indicator? And which is the next target?
CQ17	Which are the Strategic Objectives within the cascaded Balanced Scorecards that impact a certain objective, either through decomposition or cause-effect relationships?

As stated before, a Balanced Scorecard is defined as a strategy management system to help companies to achieve a desired future state. To define this state, organizations state their mission, a vision (the desired future state) and values that will guide the organization for the following years, which are formalized in the BSO using sub-properties from the "hasStrategyStatementElement" object property. Competency questions CQ1 and CQ2 were defined to illustrate how the ontology can currently answer these questions. The SPARQL query for CQ1 is shown in Listing 1, while CQ2 is shown in Listing 2 which returns the date (xsd:dateTime) associated with the defined Vision of a certain BSC. The notation of a class name between angle brackets (<>), e.g., $< Balanced\_Scorecard >$ is used to define any instance of that class.

Competency Questions from CQ4 to CQ8 are related to the Strategy Map and its elements. The query presented in Figure 12 returns the number of strategic objectives from the Strategy Map of a certain BSC grouped by its Perspectives (the query can be adapted for Strategic Themes instead of Perspectives). The query results are also presented. The "hasCauseEffectRelationship" property is used to analyse the impact between Strategic Objectives, as showed in Listing 3. In SPARQL, the plus sign in front of a property evaluates the property as if it is transitive, meaning that, despite only the direct relationships between the objectives being asserted, the query can infer over this relationship to analyze the indirect impact between them.

In order to evaluate the execution of the strategy, the

Listing 1
CQ1 - WHAT ARE THE STRATEGY STATEMENTS ASSOCIATED WITH A CERTAIN BSC?

```
SELECT ?Element ?Statement WHERE {
    <Balanced_Scorecard> bso:hasStrategyStatementElement ?Element.
    ?Element bso:hasValue ?Statement}
```

Listing 2
CQ2 - WHAT IS THE "TIME HORIZON" ASSOCIATED WITH A CERTAIN BSC?

```
SELECT ?Vision ?timeFrame WHERE {
    <Balanced_Scorecard> bso:hasVision ?Vision.
    ?Vision bso:hasTimeFrame ?timeFrame}
```

Listing 3
CQ8 - WHICH OBJECTIVES ARE DIRECTLY OR INDIRECTLY IMPACTED BY A CERTAIN STRATEGIC OBJECTIVE?

```
SELECT ?contributesTo WHERE {
    <Strategic_Objective> bso:hasCauseEffectRelationship+ ?contributesTo}
```

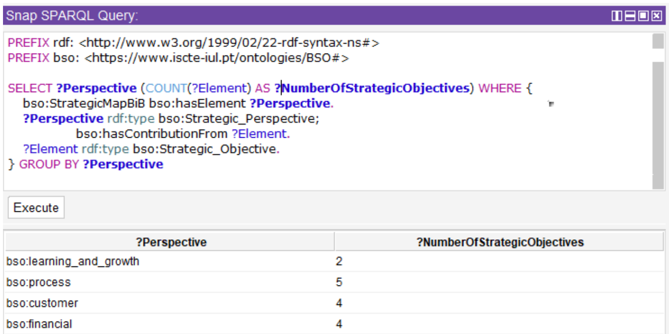


Fig. 12. SPARQL query and results from CQ5 - Which Perspectives or Themes are used in a certain Strategy Map?

ontology must provide information about the performance indicators that allow to evaluate each of the strategic objectives. The relationship between an objective and indicators is materialized through the object relation "isEvaluatedBy", which serves as a "link" between the Strategy Map and the Evaluation Framework. Thus, to answer questions such as the one on CQ9, it is enough to select the URI (Uniform Resource Identifier) of the objective and observe the range of this property (e.g., $< BSO : U1 >$:isEvaluatedBy ?PerformanceIndicator"). Information concerning each indicator can be obtained through the data property "hasIndicatorInformation", which has sub-properties on unit type ("hasUnitType"), frequency ("hasFrequency"), data source ("hasDataSource") and data quality ("hasDataQuality"), among others (CQ10/11). Targets and initiatives are related to the indicators through the object relations "isTargetFor" and "hasImpactOn", respectively.

Furthermore, the ontology should store and evaluate information regarding the actual values collected for each indicator. These values should be captured through the information systems of each organization. To store these values, the "Actual Value" class was created, which encompasses the value and the time frame to which it refers. This data enables the ontology to answer questions such as CQ15 (Listing 5), which allows for assessing the success of the defined targets by comparing the latest value of an indicator with their target values and respective deadlines.

Finally, one of the key benefits enabled by utilizing the BSO is the ability to validate alignment between BSCs. This can be achieved by either employing "isDecompositionOf" object property between Strategic Objectives, which is a sub-property of "isContribution", establishing a cause-effect relationship between Strategic Objectives in different Strategy Maps, or by defining the cascading at BSC level ("cascadesFrom"). By establishing this link, comprehensive alignment analysis between objectives, indicators, and other essential elements becomes feasible, as shown in Listing 6.

C. Ontology Validation

The ontology was validated using the OntOlogy Pitfall Scanner! (OOPS!) tool [49]. OOPS! detects common mistakes and pitfalls made during ontology development. When analysing the BSO, the tool did not detect any critical pitfalls, which "could affect the ontology consistency, reasoning, applicability, among others" [49, p.15]. Also, only one important pitfall was reported by the tool (P41: No license declared). OOPS! detected thirteen (13) minor pitfalls, however, these do not represent a problem or error.

The tool detected "Learning and Growth Perspective" as one case of "Merging different concepts in the same class", which is not applied since the this is a BSC perspective (and therefore a single element). Another minor pitfall detected related to the different naming convention used for the ontology elements (which followed a different pattern for Classes and relationships). Lastly, OOPS! found 11 cases of "inverse relationships

Listing 4
CQ13 - WHICH TARGETS ARE DEFINED FOR A CERTAIN INDICATOR?

```
SELECT ?Target ?Time ?Value WHERE {
    ?Target bso:isTargetFor <Performance_Indicator>;
       bso:hasTimeFrameEnd ?Time;
       bso:hasValue ?Value.}
```

Listing 5
CQ15 -WHAT IS THE LATEST VALUE FOR A CERTAIN PERFORMANCE INDICATOR? AND WHICH IS THE NEXT TARGET?

```
SELECT ?ac ?endDate ?value ?targetValue ?targetDate WHERE {
    <Performance_Indicator> bso:hasActualValue ?ac;
        bso:hasTarget ?target.
    ?target bso:hasValue ?targetValue;
        bso:hasTimeFrameEnd ?targetDate.
    ?ac bso:hasTimeFrameEnd ?endDate;
        bso:hasValue ?value.
    FILTER NOT EXISTS{
        <Performance_Indicator> bso:hasActualValue ?otherAC.
        ?otherAC bso:hasTimeFrameEnd ?date.
        FILTER (?date > ?endDate)}
    FILTER (?targetDate > ?endDate && NOT EXISTS{
        <Performance_Indicator> bso:hasTarget ?otherTarget.
        ?otherTarget bso:hasTimeFrameEnd ?date.
        FILTER (?date < ?targetDate)})}
```

Listing 6
CQ17 - WHICH ARE THE STRATEGIC OBJECTIVES WITHIN THE CASCADED BALANCED SCORECARDS THAT IMPACT A CERTAIN OBJECTIVE

```
SELECT DISTINCT ?StrategicObjective WHERE {
    BIND (<StrategicObjective> AS ?targetObjective).
    {?StrategicObjective bso:isDecompositionOf ?targetObjective}
    UNION{
     ?StrategicObjective bso:hasCauseEffectRelationship ?targetObjective;
           bso:isPartOf ?lowerBSC.
     ?higherBSC bso:hasPart ?targetObjective;
           bso:cascadesTo ?lowerBSC.}}
```

not explicitly declared (e.g., "hasMission", "contributesToPerspective").

OOPS! tool also suggests that some properties, such as "hasCauseEffectRelationship" or "hasPart", could be either transitive or symmetric since they have the same domain and range. These suggestions were not followed due to the reasons below:

1) OWL reasoners cannot infer over complex properties, such as transitive plus asymmetric and irreflexive property[50], which could be the case of the "hasPart" property;
2) Most of these properties are used to define direct relationships between classes. While a cause-effect relationship could be seen as transitive, without a different property to model the direct and indirect contributions, the materialization of this transitivity would lead to a loss of knowledge;
3) This type of transitive analysis can still be obtained using SPARQL queries, as previously shown (see Listing 3).

VII. DISCUSSION

This article introduces the BSO in an endeavor to bridge the gap between strategy management and data related to the BSC framework. The BSO provides a structured framework to store and analyze knowledge related to the BSC, incorporating information about the Strategy Map and Quantification Frameworks used for evaluating strategy execution. Specifically, the suggested formalization of the BSC framework provides a semantic layer to facilitate the integration, alignment, and traceability of strategic models with organizational information systems, which are essential for supplying data to evaluate the BSC's performance indicators. As previously discussed, performance indicators measure the organizational progress in relation to the strategic goals, supporting decision-makers in the evaluation of the effectiveness of current strategies. In a comprehensive BI implementation, the BSO facilitates the data optimization phase (see Figure 1) enabling an effective BSC implementation, using a data-driven approach.

The BSO provides a formal, structured, and semantically rich representation of the BSC framework, ensuring consistency in how strategic objectives, performance indicators, and their relationships are defined and interpreted, and providing decision-makers with a shared and unambiguous understanding of the BSC components. This knowledge representation can capture the complex inter-dependencies and cause-and-effect relationships between various components, providing a deeper understanding of how they impact one another. Ontologies also support automated reasoning, enabling logical inferences that can help identify implicit relationships or conflicts within the proposed BSC model. For example, it can provide rules that enable the detection of wrongfully defined strategic objectives, alert when certain indicators are irrelevant to the organization's strategy (i.e., are not being used to evaluate any objective or long-term goal), or facilitate the analysis and validation of transitive cause-and-effect relationships. By adopting an ontology-based approach, this solution offers a flexible and semantically enriched environment for representing the complex relationships inherent to strategic management and data-driven decision-making. When compared to traditional BSC implementations, the BSO provides improved clarity and interoperability to an organizational strategy, necessary for improved strategic decision support throughout the organization.

A. Contributions to the Literature

The present study contributes to the existing literature by addressing various identified gaps associated with the BSC implementation as a communication and strategy execution system. As shown in Table III, there is a consensus among authors concerning the definitions of first and second-generation BSCs. However, a shared definition for a Type III BSC was absent from the literature. Based on Cardoso [47] definition of a third-generation BSC, a comprehensive strategy communication and implementation system needs to integrate the BSC with the different systems already in use by the organizations. This integration is required to enable an accurate, traceable, and continuous monitoring and improvement of the strategy execution, based on a data-driven approach. Existing studies, such as those by Kumar, Prince, and Baker [5] and Tawse and Tabesh [6] also emphasize the importance of combining the BSC with other systems and tools for an effective implementation.

Moreover, recent studies [6], including ours, still find the need to recommend the use of strategy maps for an effective BSC implementation. The elements and relationships of this adaptable framework need to be formalized to ensure that the BSC implementation fully harnesses the benefits inherent to the BSC as a strategic management system. The formalization of knowledge through techniques such as ontologies offers several benefits [16], [17], including enhanced interoperability between systems, knowledge validation, inference of new knowledge, and the utilization of semantics to improve communication and minimize misunderstandings. However, none of the ontologies identified during our literature review could comprehensively or satisfactorily represent all BSC elements. Furthermore, none of these ontologies were available online.

Our work addresses these gaps by introducing and developing the Balanced Scorecard Ontology, which formalizes the BSC framework, explicitly defining its components, elements, and relationships. Additionally, this semantic layer facilitates the integration of the BSC implementation with other organizational information systems, due to the increased interoperability. The proposed BSO is an additional layer seamlessly integrated into the Business Intelligence part of the Execution Premium Process (as proposed by Kaplan and Norton [4]), enhancing the organizational strategic monitoring and improvement capabilities. This is a crucial contribution, given the growing importance of leveraging data in strategic decision-making processes in an evolving business environment [23].

B. Managerial Implications

The utilization of the BSO presents several advantages for managers. Firstly, it helps to ensure alignment between the BSC and the organization's overarching goals. This formal representation enables managers to assess whether BSC elements contribute to the organization's strategy, thereby preventing the allocation of resources towards nonessential or superfluous indicators and objectives. Additionally, the ontology can aid in compliance and governance efforts by allowing managers to verify that the organizational strategy adheres to regulatory requirements and facilitates the documentation, reporting, and tracking of compliance with pertinent standards, such as European Commission policies and performance evaluation in public administration.

The BSO provides a clear and unambiguous representation of the BSC framework, ensuring that all stakeholders have a common understanding of the strategy, strategic objectives, and indicators. This can improve communication and alignment throughout the organization, across organizational levels, or between departments. Furthermore, the BSO can be a valuable tool for facilitating the transfer of knowledge within the organization. By formalizing the cascading impact of each BSC element, the contribution of individual or departmental objectives to the overall organizational strategy can be made clear. This clarity facilitates a better understanding of the strategic framework among employees and stakeholders, potentially serving as a motivational factor.

Moreover, the incorporation of the BSO in the strategic decision-making process can help safeguard that all decisions align with the organization's mission, vision, and strategic objectives. This proactive approach helps to avoid decision-making that may not contribute to the long-term success of the organization. The ontology can enable scenario analysis, facilitating an understanding of how changes in specific indicators or objectives influence the overall strategy and making it easier to evaluate the potential consequences of different decisions. By encoding the relationships between strategic objectives, indicators, and other BSC elements, the BSO can help managers to understand the risks and benefits associated with each decision, make more informed choices, and adapt to changing circumstances.

The BSO can also facilitate the integration of data from various sources into the BSC model, streamlining the collection, analysis, and reporting of performance indicators, which can become key in supporting real-time or near-real-time monitoring of performance indicators and decision support.

Finally, the ontology can be integrated with decision support systems (e.g., BI systems) to improve decision-makers' perspectives on organizational strategy and performance and empower managers with user-friendly information and tools to make informed, data-driven strategic decisions.

VIII. Conclusions and Research Directions

This article presents the development and evaluation of the Balanced Scorecard Ontology (BSO). The BSO represents elements from the Balanced Scorecard framework and their relationships in a formal, comprehensible and explicit way. The Ontology Requirements Specification Document (ORSD, see Section V-A) is presented with information regarding the ontology a) goals, domain and scope, b) users, use cases and applications, c) knowledge inputs and d) competency questions. The main challenges found in the ontology design and development processes are described. The ontology was validated and evaluated by answering the competency questions defined in the ORSD, using a real-case study of a university library, and using the OOPS! tool. Through this process, it was proved that the BSO is able to formalize BSC knowledge, validate BSC elements and relationships, and infer new knowledge related to them.

With the design and development of the BSO concluded, future research directions include the introduction of rules that can validate ontological knowledge. Some validations are already in place. For example, the BSC class is defined as the equivalent of the class of things [(hasComponent some 'Quantification Framework') and (hasComponent exactly 1 'Strategy Map') and (hasStrategicLevel exactly 1 xsd:string)]. This will trigger an error on the ontology when a BSC has, for example, two "hasComponent" relationships to two Strategy Maps. However, due to the Open World Assumption used in OWL, if no "hasComponent" property is found to a Strategy Map, the ontology and the instances are still valid and no error is shown. Shapes Constraint Language (SHACL)[8] and Semantic Web Rule Language (SWRL)[9] can be used on top of RDF and OWL to constrain and validate ontological knowledge.

Furthermore, it is important to use the BSO in different applications and decision-support scenarios. The interoperability gained from the ontology could be used together with Enterprise Architecture (EA) models, such as ArchiMate, to ensure an alignment between strategy and other EA layers, such as business, application, and infrastructure. This alignment would ensure the integration between strategic business vision down to the IT infrastructure, allowing analysis between EA layers.

Lastly, and as stated before, analysis and evaluation of strategy execution should use real data managed by organizational information systems. However, the relationship between this data, i.e., the values collected for each indicator and the ontology representation of these values, is not trivial (different indicators, different detail levels, etc). Ideally, the values should be retrieved from information systems, such as BI systems, and loaded into the ontology using an automated or semi-automated process. This approach would enable an accurate and continuous evaluation of the strategy execution, leading to the realization of a Type III BSC, a comprehensive strategic management and execution system.

[8]https://www.w3.org/TR/shacl/
[9]https://www.w3.org/Submission/SWRL/

Appendix

The case study used in Section VI contains strategy information based on a public scorecard from a library repository of a higher education university developed and published in Silva [33]. Figure 13 presents the strategy map which includes:

- Four Perspectives: Financial, Learning & Growth, Internal Process, and Users. In public or non-profit organizations, the financial perspective is usually presented as the base of the strategy map;
- Two Strategic Themes: Quality of Service and Growth;
- Thirteen (13) Strategic Objectives, such as "Increase visibility" and "Increasing institutional reputation", from the Users perspective, and their cause-effect relationships;
- and, the Mission, on the top of the map.

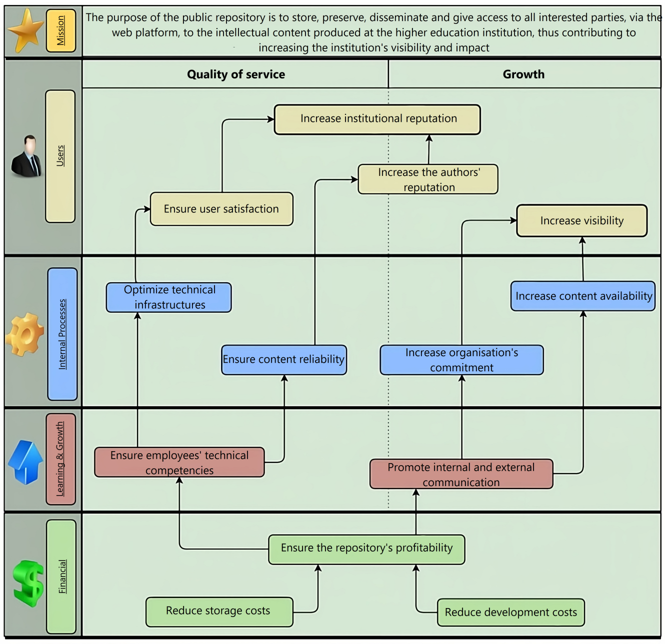


Fig. 13. Proposed strategy map for the library repository (translated from the original)

References

[1] R. S. Kaplan, D. P. Norton, *et al.*, *The balanced scorecard: measures that drive performance*. Harvard business review US, 1992, vol. 70.

[2] R. S. Kaplan and D. P. Norton, "Linking the balanced scorecard to strategy," *California management review*, vol. 39, no. 1, pp. 53–79, 1996.

[3] R. S. Kaplan and D. P. Norton, *Alignment: Using the balanced scorecard to create corporate synergies*. Harvard Business Press, 2006.

[4] R. S. Kaplan and D. P. Norton, *The execution premium: Linking strategy to operations for competitive advantage*. Harvard Business Press, 2008.

[5] J. Kumar, N. Prince, and H. K. Baker, "Balanced scorecard: A systematic literature review and future research issues," *FIIB Business Review*, vol. 11, no. 2, pp. 147–161, 2022.

[6] A. Tawse and P. Tabesh, "Thirty years with the balanced scorecard: What we have learned," *Business Horizons*, vol. 66, no. 1, pp. 123–132, 2023.
[7] E. Cardoso, M. J. Trigueiros, *et al.*, "Using the balanced scorecard as a tool for performance management of higher education degrees," in *Proceedings of the 13th International Conference of European University Information Systems (EUNIS 2007), Grenoble, France*, 2007, pp. 27–29.
[8] M. A. Camilleri, "Using the balanced scorecard as a performance management tool in higher education," *Management in Education*, vol. 35, no. 1, pp. 10–21, 2021.
[9] I. B. Soysa, N. P. Jayamaha, and N. P. Grigg, "Validating the balanced scorecard framework for nonprofit organisations: An empirical study involving australasian healthcare," *Total Quality Management & Business Excellence*, vol. 30, no. 9-10, pp. 1005–1025, 2019.
[10] S. Victor and A. Farooq, "Combining the use of data analytics and balanced scorecard to enhance healthcare delivery: A study," *Journal of Health Management*, vol. 24, no. 2, pp. 248–255, 2022.
[11] F. Betto, A. Sardi, P. Garengo, and E. Sorano, "The evolution of balanced scorecard in healthcare: A systematic review of its design, implementation, use, and review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 16, p. 10 291, 2022.
[12] T. Fatima and S. Elbanna, "Balanced scorecard in the hospitality and tourism industry: Past, present and future," *International Journal of Hospitality Management*, vol. 91, p. 102 656, 2020, ISSN: 0278-4319. DOI: https://doi.org/10.1016/j.ijhm.2020.102656. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278431920302085.
[13] E. G. Hansen and S. Schaltegger, "The sustainability balanced scorecard: A systematic review of architectures," *Journal of Business Ethics*, vol. 133, pp. 193–221, 2016.
[14] R. Fernandez-Gonzalez, F. Puime-Guillen, and J. E. Vila-Biglieri, "Environmental strategy and the petroleum industry: A sustainability balanced scorecard approach," *Journal of Petroleum Exploration and Production Technology*, vol. 13, no. 2, pp. 763–774, 2023.
[15] E. Supino, F. Barnabè, M. C. Giorgino, and C. Busco, "Strategic scenario analysis combining dynamic balanced scorecards and statistics," *International Journal of Productivity and Performance Management*, vol. 69, no. 9, pp. 1881–1902, 2019.
[16] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data & knowledge engineering*, vol. 25, no. 1-2, pp. 161–197, 1998. DOI: 10.1016/S0169-023X(97)00056-6.
[17] A. L. Antunes, E. Cardoso, and J. Barateiro, "Incorporation of ontologies in data warehouse/business intelligence systems-a systematic literature review," *International Journal of Information Management Data Insights*, vol. 2, no. 2, p. 100 131, 2022.
[18] H. Norreklit, "The balance on the balanced scorecard a critical analysis of some of its assumptions," *Management accounting research*, vol. 11, no. 1, pp. 65–88, 2000.
[19] R. Sharda, D. Delen, E. Turban, J. Aronson, and T. Liang, *Business intelligence and analytics*, 10th. Pearson Edition Limited, 2015.
[20] J. Ereth and W. Eckerson, "Ai: The new bi - how algorithms are transforming business intelligence and analytics," *Data Strategy Insider*, 2018. [Online]. Available: https://www.ibm.com/downloads/cas/M7VMLOPY.
[21] R. Kimball and M. Ross, *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons, 2013.
[22] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Manchester, vol. 1, 2000, pp. 29–39.
[23] R. M. Grant, *Contemporary strategy analysis*. John Wiley & Sons, 2021.
[24] S. B. Khakbaz and N. Hajiheydari, "Proposing a basic methodology for developing balanced scorecard by system dynamics approach," *Kybernetes*, vol. 44, pp. 1049–1066, Jun. 2015. DOI: 10.1108/K-12-2014-0287.
[25] R. S. Kaplan, N. P. D. K. S. Robert, R. S. Kaplan, and D. P. Norton, *The strategy-focused organization: How balanced scorecard companies thrive in the new business environment*. Harvard Business Press, 2001.
[26] M. E. Porter, "How competitive forces shape strategy," in *Readings in strategic management*, Springer, 1989, pp. 133–143.
[27] W. C. Kim and R. Mauborgne, *Blue ocean strategy, expanded edition: How to create uncontested market space and make the competition irrelevant*. Harvard business review Press, 2014.
[28] A. Osterwalder, "The business model ontology a proposition in a design science approach," 2004.
[29] A. Osterwalder and Y. Pigneur, *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons, 2010, vol. 1.
[30] M. Treacy and F. Wiersema, "Customer intimacy and other value disciplines," *Harvard business review*, vol. 71, no. 1, pp. 84–93, 1993.
[31] R. S. Kaplan, D. P. Norton, *et al.*, "Having trouble with your strategy? then map it," *Focusing Your Organization on Strategy—with the Balanced Scorecard*, vol. 49, no. 5, pp. 167–176, 2000.
[32] A. Osterwalder, Y. Pigneur, G. Bernarda, and A. Smith, *Value proposition design: How to create products and services customers want*. John Wiley & Sons, 2014, vol. 2.
[33] R. J. G. d. Silva, "Aplicação do business model canvas e de balanced scorecard a repositórios digitais," M.S. thesis, 2015.
[34] E. Cardoso, D. Santos, D. Costa, F. Caçador, A. Antunes, and R. Ramos, "Learning scorecard: Monitor and foster student learning through gamification," in *European Knowledge Acquisition Workshop*, Springer, 2016, pp. 55–68.
[35] H. A. Sacoor, E. Arsenio, E. Cardoso, J. Barateiro, *et al.*, "How to ensure the provision of social inclusive public transport bus services? a strategic case study in lisbon," 2021.
[36] N. F. Noy, D. L. McGuinness, *et al.*, *Ontology development 101: A guide to creating your first ontology*, 2001.
[37] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" In *Handbook on ontologies*, Springer, 2009, pp. 1–17.
[38] J. Pan, "Resource description framework," in *Handbook on ontologies*, May 2009, pp. 71–90. DOI: 10.1007/978-3-540-92673-3_3.
[39] X. Su and L. Ilebrekke, "A comparative study of ontology languages and tools," in *International Conference on Advanced Information Systems Engineering*, Springer, 2002, pp. 761–765. DOI: 10.1007/3-540-47961-9-62.
[40] H. A. Hartanto, R. Sarno, and N. F. Ariyani, "Linked warning criterion on ontology-based key performance indicators," in *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)*, IEEE, 2016, pp. 211–216.
[41] F. Bobillo, M. Delgado, J. Gómez-Romero, and E. López, "A semantic fuzzy expert system for a fuzzy balanced scorecard," *Expert Systems with Applications*, vol. 36, no. 1, pp. 423–433, 2009.
[42] R. F. Navarro-Hernandez, A. Perez-Soltero, G. Sanchez-Schmitz, M. Barcelo-Valenzuela, and C. E. Rose, "An ontological model to support the implementation of balanced scorecard in the organizations," in *Proceedings of the 10th World Multiconference on Systemics, Cybernetics and Informatics*, Citeseer, vol. 4, 2006, pp. 324–328.
[43] A. Osterwalder and Y. Pigneur, "An ebusiness model ontology for modeling ebusiness," *BLED 2002 proceedings*, p. 2, 2002.
[44] G. Speckbacher, J. Bischof, and T. Pfeiffer, "A descriptive analysis on the implementation of balanced scorecards in german-speaking countries," *Management accounting research*, vol. 14, no. 4, pp. 361–388, 2003.
[45] P. R. Niven, *Balanced scorecard evolution: A dynamic approach to strategy execution*. John Wiley & Sons, 2014.
[46] G. Lawrie and I. Cobbold, "Third-generation balanced scorecard: Evolution of an effective strategic control tool," *International journal of productivity and performance management*, vol. 53, no. 7, pp. 611–623, 2004.
[47] E. Cardoso, "Performance and quality management of higher education programmes," Ph.D. dissertation, Ph.D. thesis, University Institute of Lisbon (ISCTE-IUL), 2011.
[48] S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure, "Knowledge processes and ontologies," *IEEE Intelligent systems*, vol. 16, no. 1, pp. 26–34, 2001.
[49] M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa, "OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 10, no. 2, pp. 7–34, 2014.
[50] M. Schneider, S. Rudolph, and G. Sutcliffe, *Modeling in owl 2 without restrictions*, 2013. arXiv: 1212.2902 [cs.AI].

CHAPTER 5

Journal Article 4

This chapter presents the application of the Strategy Ontology (artifact #3) in the field of public sector strategy management to enhance the capacity of organizations to make informed data-driven decisions. The Strategy Ontology is demonstrated and evaluated in the context of the DW/BI Strategy Analysis in a Public Organization (LNEC) case study (see Section 1.3.2). As shown in Figure 4.1, this publication (JA4) reports on contributions related to the DSRM's third iteration (Iter. 3). The Strategy Ontology is used with SW techniques to support strategy analysis, including validating the strategy formulation and supporting strategy execution by assessing performance indicators, verifying the design of cause-and-effect relationships between strategic objectives, and monitoring and empirically validating these relationships.

Article details:

- **Title**: Strategic Analysis in the Public Sector Using Semantic Web Technologies;
- **DOI**: http://dx.doi.org/10.1145/3656587;
- **Date**: 2024 (just accepted);
- **Journal**: Digital Government: Research and Practice;
- **Publisher**: ACM.

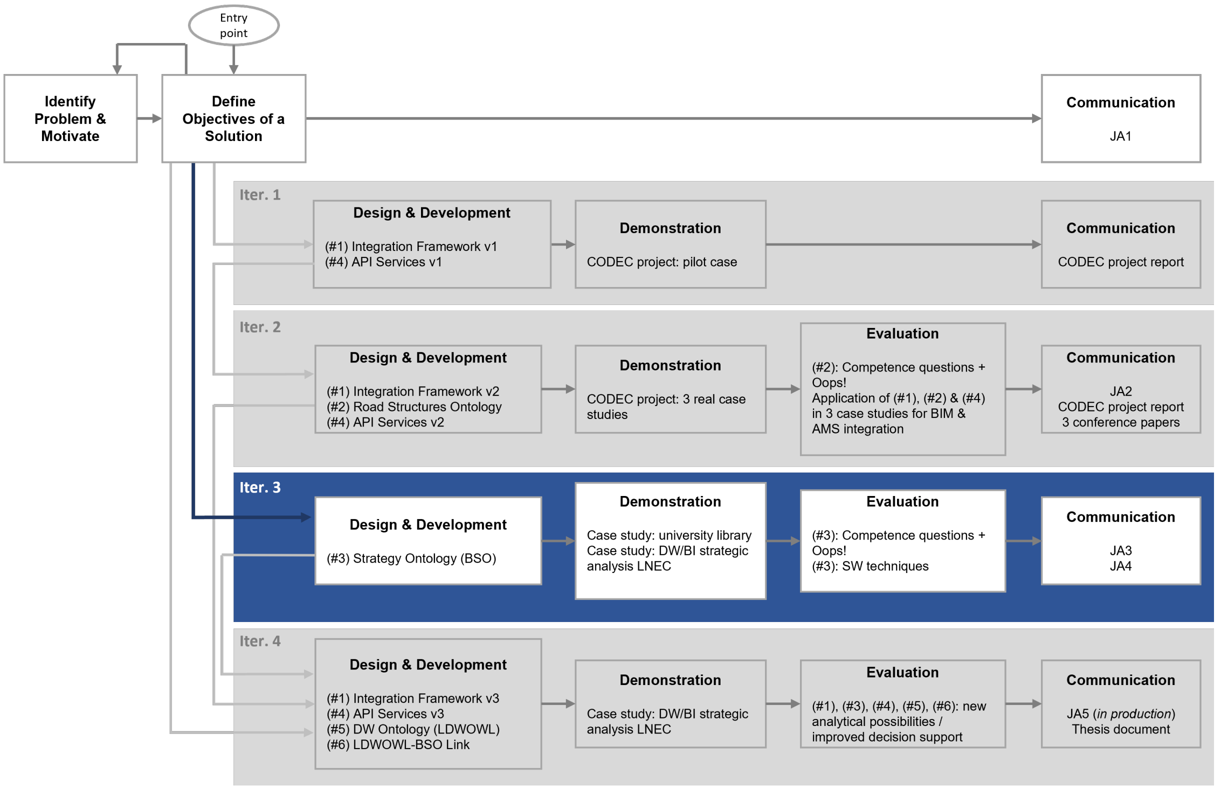


FIGURE 5.1. DSRM's JA4 Communication.

Strategic Analysis in the Public Sector Using Semantic Web Technologies

ANTÓNIO LORVÃO ANTUNES, LNEC - National Laboratory for Civil Engineering, Lisbon, Portugal and ISCTE - Instituto Universitário de Lisboa, Lisbon, Portugal
JOSÉ BARATEIRO, Faculdade de Ciência e Tecnologia, Universidade of Algarve, Faro, Portugal
ELSA CARDOSO, ISCTE - Instituto Universitário de Lisboa, Lisboa, Portugal and CIES-Iscte, Lisboa, Portugal

This article addresses the complex challenges that public organizations face in designing, implementing, and evaluating their strategies, where public interest and regulatory compliance often intertwine with strategic objectives. This research investigates the application of ontologies in the field of public sector strategy management to enhance the capacity of organizations to make informed data-driven decisions, efficiently allocate resources, and effectively navigate the intricate landscape of the public sector. The LNEC - National Laboratory for Civil Engineering's strategy is used as an exploratory case study. Semantic web technologies are used to perform strategy analysis, including validating the strategy formulation and supporting the strategy execution by assessing performance indicators, verifying the design of cause-and-effect relationships between strategic objectives, and monitoring and empirically validating these relationships. The increased interoperability of these technologies enables information sharing across systems and organizations. Following the strategy analysis, recommendations are provided, leading to a more robust and data-driven strategic management approach, enabling accurate, traceable, and continuous monitoring of an organization's strategy. Theoretical and practical implications are discussed, along with limitations and future work. This research offers a blueprint for public sector organizations seeking to optimize their strategies, foster transparency, and deliver more effective services to the public they serve.

CCS Concepts: • **Information systems** → **Decision support systems**; **Semantic web description languages;**

Additional Key Words and Phrases: Strategy Analysis, Public Sector, Semantic Web, Balanced Scorecard, Ontology

ACM Reference Format:
António Lorvão Antunes, José Barateiro, and Elsa Cardoso. 2024. Strategic Analysis in the Public Sector Using Semantic Web Technologies. *Digit. Gov. Res. Pract.* 5, 3, Article 20 (September 2024), 20 pages. https://doi.org/10.1145/3656587

1 INTRODUCTION

Strategic management is a process undertaken by public organizations or other entities to formulate, implement, and evaluate strategies to achieve their long-term objectives and goals [3]. The importance of strategy validation and analysis in the public sector cannot be neglected, as it directly impacts public value, due to the nature of their mission. Organizational performance is significantly impacted by strategic planning in both private and public sectors and can be enhanced by the formalization of the strategic planning process [11]. However, most

Funding: This work was partially supported by the Portuguese Foundation for Science and Technology (grant number 2021.07134.BD).
Authors' addresses: A. Lorvão Antunes, LNEC - National Laboratory for Civil Engineering, Lisboa, Lisboa, Portugal and ISCTE - Instituto Universitário de Lisboa, Lisboa, Portugal; e-mail: aantunes@lnec.pt; J. Barateiro, Faculdade de Ciência e Tecnologia, University of Algarve, Faro, Portugal; e-mail: jebarateiro@ualg.pt; E. Cardoso, Department of Information Science and Technology, ISCTE - Instituto Universitário de Lisboa, Lisboa, Portugal and CIES-Iscte, Lisboa, Lisboa, Portugal; e-mail: elsa.cardoso@iscte-iul.pt.

ACM 2639-0175/2024/09-ART20
https://doi.org/10.1145/3656587

public organizations typically use strategic management systems with low comprehensiveness or formality and are usually decentralized [4].

Král [29] defines performance management as "*a strategic approach to management that provides managers, employees, and owners with the tools and techniques that they then use to plan, monitor, measure, and evaluate the performance of an organization.*" In public administration, the **Balanced Scorecard (BSC)** [22] still remains the most well-known approach to assess performance, due to the balance between non-financial and financial indicators across different perspectives [29, 43]. However, Manes-Rossi et al. [32] only found that 8% of the works in their literature review explore non-financial reporting from a strategy management perspective in the public sector. Furthermore, Král [29] identifies research directions in this field, including continued performance evaluation, use of official quantitative data, and clear and understandable (to policymakers, managers, and stakeholders) performance management systems. Moreover, the author defends that, when performance evaluation methods become overly complex, they often become less practical and less likely to be effectively applied in real-world situations. Sharing this concern, Tawse and Tabesh [43] state that "*the BSC has the potential to improve organizational performance, but to realize that potential it must be effectively implemented.*" The authors provide three recommendations: (1) develop a strategy map to ensure that BSC elements are related with cause-and-effect relationships, (2) ensure top management team commitment and support, and (3) improve key stakeholder engagement.

Semantic Web (SW) technologies, such as ontologies, have been used during the last years to encode knowledge in a way that allows it to be shared, be reused, and, most importantly, become machine-readable [16]. Kalampokis et al. [18] present these technologies as one of the emerging technologies in the public sector, highlighting the advantages of their usage, specifically the shared semantics and interoperability. Ontologies can be used to create a common semantic data model that can be useful to define unified report methods (beneficial for both reporters and readers) [15, 32], integrate or transfer data between public organizations [18], and automate processes using the formalization of knowledge from heterogeneous sources [17]. SW technologies have also been used to foster openness and transparency in the public sector [33]. Pucihar et al. [37] identified ontologies and SW as highly relevant research topics necessary for a holistic and dynamic government, with ontologies being used to improve interoperability and user experience in e-government services [24–28].

This article explores how public organizations, such as the **National Laboratory for Civil Engineering (LNEC)** in Portugal, can employ ontologies to assess their strategy formulation and execution. Furthermore, it investigates how ontologies can assist in the validation of the formulated strategy, evaluation of performance indicators, and validation of cause-and-effect relationships between strategic objectives. In doing so, this research aims to provide tools to enhance the capacity of organizations to make informed data-driven decisions, efficiently allocate resources, and effectively navigate the intricate landscape of the public sector. Understanding the nuances of strategy validation and analysis in this context, where objectives often intertwine with public interest and regulatory compliance, presents a complex challenge to which this article contributes.

This research aims to improve the body of knowledge of public sector strategy management through the use of SW technologies. Additionally, this research can potentially be used to inform public data-driven decision-making and policy definition, in line with what happened during Covid regarding public health policies [15].

The remainder of the article is structured as follows: Section 2 presents background research concerning ontologies and the BSC framework and introduces the **Balanced Scorecard Ontology (BSO)** used in this research. The methodology followed in this research is shown in Section 3. The case study, LNEC, is presented in Section 4, together with its current strategy formulation. Section 5 describes the ontology population process, explaining the mapping between LNEC's strategy and the BSO. Section 6 presents the use of BSO and other semantic technologies to validate and infer over the strategy, followed by strategy analysis, including a set of recommendations, in Section 7; Section 8 encompasses the discussion, which includes theoretical and practical contributions, limitations, and avenues for future work. Finally, Section 9 presents the conclusions.

2 BACKGROUND

This section presents background concepts related to ontologies and the balanced scorecard framework.

2.1 Ontologies

Ontologies are formal and explicit specifications of shared conceptualizations [40], used to represent knowledge pertaining to a specific domain of interest, encompassing its concepts, properties, and relationships. The purpose of ontologies is to facilitate the sharing, reuse, and analysis of knowledge, ultimately promoting interoperability and heterogeneity [35]. According to the **World Wide Web Consortium (W3C)**,[1] these qualities make ontologies an indispensable component of the SW. When populated with individual instances, an ontology is called a Knowledge Base [13].

The **Resource Description Framework (RDF)** is a W3C recommendation designed to enable the creation, exchange, and utilization of web annotations. In RDF, resources are described using triples in the form of *<subject, property, object>* [36]. SPARQL is a W3C query language for querying and manipulating data stored in RDF format, commonly used for querying SW data and knowledge graphs. The W3C **RDF Schema (RDFS)** provides a vocabulary for RDF introducing the concepts of classes and hierarchies. Building on RDFS, the **Ontology Web Language (OWL)** enhances expressiveness by incorporating elements such as disjointness, cardinality, object and data properties, and additional vocabulary. OWL comes in three sublanguages/types: Lite, DL, and Full, each offering varying levels of expressiveness. The choice of an OWL sublanguage depends on the specific problem domain and modeling requirements, with a tradeoff between expressiveness and inference capabilities (reasoning) [41].

2.2 Balanced Scorecard

The BSC model was presented in 1992 by Kaplan and Norton as a system for measuring an organization's performance [22]. Over the years, the BSC has evolved from a performance measurement tool to a strategic management system. Kaplan and Norton emphasized the importance of using the BSC to align the organization's strategy with its performance measures and to drive continuous improvement [19]. Further, the authors introduced the concept of strategy maps, a visual representation of the strategic objectives and the respective cause-and-effect relationships, helping organizations to better understand how their objectives are interconnected and how they can best allocate resources to achieve their goals [23]. The BSC is usually divided into four perspectives: financial, customers, internal processes, and learning and growth. The BSC is "*agnostic to the formulation model used*" [20], meaning that any business strategy formulation can be executed and communicated utilizing the BSC and its elements.

Today, the BSC is seen as a system for communicating and executing strategy [21, 30, 43], which includes elements such as the organization's mission, values, and vision statements; perspectives; strategic themes; strategic objectives; and **Key Performance Indicators (KPIs)** to measure the objectives, targets, and initiatives (projects that need to be executed to achieve the targets). The BSC has been successfully applied in many industries, including Higher Education [6, 8], Healthcare [2, 39, 44], and Tourism [9], improving organizational performance, enhancing strategic alignment, and facilitating communication and coordination across different departments and organizational levels. The BSC is also being used to promote sustainability and corporate social responsibility [10, 14]. Finally, combining the BSC with other systems and tools can lead to a more effective strategy formulation and implementation and improved decision-making [30, 42].

2.3 Balanced Scorecard Ontology

The BSO was developed to describe and store knowledge related to the BSC framework [31], including the strategy map, which presents the long-term strategy, and several quantification frameworks providing a short-term

[1]https://www.w3.org/

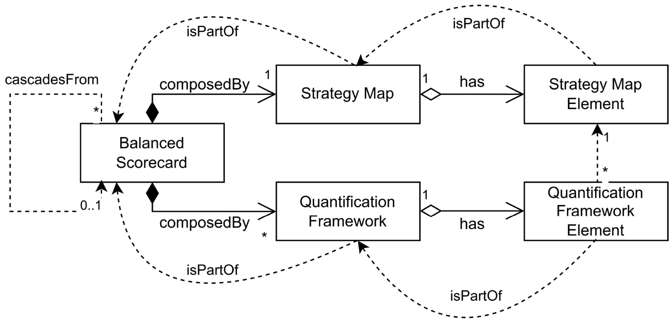


Fig. 1. Balanced Scorecard Ontology concepts.

view of the strategy execution (see Figure 1). The BSO allows any organization to formalize, communicate, align, and execute its BSC-based strategy. The ontology also allows for strategy validation (e.g., ensuring every strategic objective has a performance indicator) and can improve interoperability between performance management systems and the strategy formulation.

A BSC should be defined or cascaded across various organizational levels, enabling managers to formulate corporate, departmental, team, and individual strategies (aligned with the employees' incentive systems). Information requirements vary, as does the granularity (or summarization) of performance indicators and data. In essence, a BSC represents the aggregation of all defined BSCs, starting from the corporate level and cascading down to the lowest organizational level.

Two major components are needed to define a BSC at any strategic level: a Strategy Map and a Quantification Framework, as shown in Figure 1. A Strategy Map presents the long-term view of the strategy, typically including the following elements: strategy statement elements (i.e., vision, mission, and values), strategic objectives, perspectives, and themes. The Quantification Framework offers a shorter-term view containing the tangible indicators, goals, and initiatives needed to translate the strategy into operational terms. The BSO represents and provides information, formalizing these elements as classes (see Figure 2), while the relationships between these elements and their attributes are represented by object and data relationships, respectively. This formalization enables inference on BSC elements and the relationships between these elements (e.g., cause-and-effect relationships between objectives). In addition to the BSC, its components, and elements, the BSO also contains information related to actual values of Performance Indicators, which store corresponding values related to a particular time frame.

3 METHODOLOGY

To explore and evaluate how ontologies can be employed to formulate, validate, and ensure the effectiveness of strategies within public organizations, an exploratory case study based on a public organization was used. This case study used the recently developed BSO to overcome the complex challenges of strategy validation and analysis in the public sector. Lorvão Antunes et al. [31] present the design and development process of the BSO, including the demonstration and evaluation phases based on a real case study of a public faculty library. The current research aims to demonstrate the advantages and impact of SW technologies, specifically the BSO, in strategy management. Figure 3 presents the methodology used, which has the following starting points (inputs):

— **Balanced Scorecard Ontology.** In previous work [31], the BSO was designed and developed to describe and store knowledge related to the BSC framework (see Section 2.3). This formal, structured, and semantically rich representation of the BSC framework ensures consistency in how strategic objectives, performance indicators, and their relationships are defined and interpreted and provides decision-makers with a shared and unambiguous understanding of the BSC components and elements.

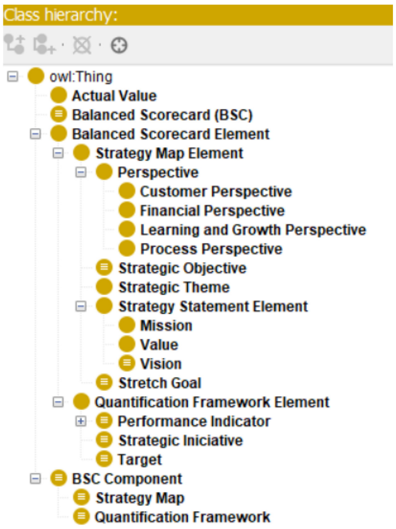


Fig. 2. Balanced Scorecard Ontology class hierarchy in Protégé (https://protege.stanford.edu/).

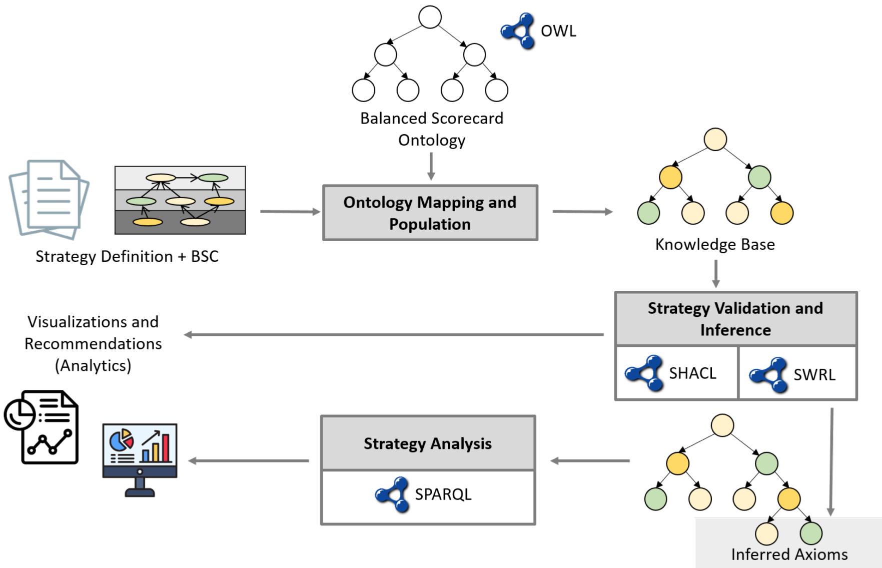


Fig. 3. Research methodology.

— **Strategy Definition and BSC.** The proposed research uses a real public sector case study to provide insights into the impact of semantic technologies in strategy management. A Portuguese public organization (LNEC) is used (see Section 4), having a current strategy recently defined using the BSC framework, allowing it to be represented and analyzed using the BSO.

The remainder of the methodology outlines the tasks that organizations must undertake to use ontologies in the assessment of their strategy formulation and execution. The output of each action can be used to provide managers and stakeholders with information, presented as recommendations (e.g., warnings, errors) and/or visualizations, for example, enabling ontology-driven strategic analysis.

— **Ontology Mapping and Population.** This process is responsible for creating a BSO knowledge base, i.e., an instantiation of the BSO containing the case study's strategy data. Prior to the ontology population, the strategy definition and the BSC need to be analyzed and mapped to BSO entities, and import rules must be created. Section 5 presents the mapping and population process of LNEC's strategy.
— **Strategy Validation and Inference.** Once the knowledge base is obtained, the strategy can be validated. Besides OWL constraints, the SW framework provides other technologies, such as SHACL and SWRL (see Section 6) that are used to validate and infer ontological knowledge. This process is responsible for the detection of errors and warnings related to the BSC's implementation and for obtaining new knowledge related to the strategy and its execution, through a set of logical rules.
— **Strategy Analysis.** Lastly, this methodology defines a set of SPARQL queries that enable the validation of the formulated strategy, evaluation of performance indicators, and validation of cause-and-effect relationships between strategic objectives. The results of these queries can be used by external applications to provide managers and other decision-makers with visualizations and recommendations regarding strategies and their executions.

4 LNEC: A CASE STUDY IN THE PUBLIC SECTOR

The National Laboratory for Civil Engineering (LNEC) in Portugal was established in 1946 to provide specialized services in civil engineering. As a public laboratory, it has been involved in national projects (e.g., dams, communication routes, river and sea hydraulics, and large structures) and international collaborations, performing scientific and technical works in almost 50 countries. Over the years, LNEC expanded its competencies, becoming a hub for research, experimentation, postgraduate education, and community/local services. LNEC underwent several changes in its organizational structure and legal framework, with the most recent reorganization in 2012 and 2013 resulting in improved autonomy in scientific, administrative, and financial matters. In 2021, a BSC approach was used to formulate the strategy for 2021–2027.

As a public institute, LNEC must follow a set of regulations and practices, namely those presented in its organic law and mission letter. Furthermore, LNEC has the duty and responsibility to report on its activities and performance, publishing annual activity plans that include, among other things, evaluations of the objectives and indicators of the **Evaluation and Accountability Framework (QUAR)**. QUAR is a mandatory framework for assessing and monitoring the performance of Portuguese public services,[2] ensuring alignment with strategic objectives, legal requirements, and user satisfaction, while also promoting transparency through public disclosure. QUAR includes various components such as the service's strategic statements (mission and vision), strategic objectives, operational objectives, indicators, targets, and results achieved. It should be noted that strategic objectives, as defined by QUAR, are not quantifiable. Operational objectives must be divided into the following categories: effectiveness, efficiency, and quality. Moreover, operational objectives are measured with a set of weighted indicators, grouped by these categories, which in turn are also weighted to provide an overall organizational performance score.[3]

4.1 LNEC Strategy for 2021–2027

A BSC approach was used to define LNEC's strategy for 2021–2027, including the definition of strategic objectives and indicators used to monitor its execution. The approved strategy map is presented in Barateiro et al.

[2] Article 10, Portuguese Law n.º 66-B/2007, from December 28, 2007.
[3] LNEC QUAR definition for 2023 can be found in Annex III of the activity plan report, available at https://www.lnec.pt/pt/downloads/download.php?id=1037 (in Portuguese).

Table 1. LNEC's Mission, Values, and Vision Statements [1]

Strategy Statement	Description
Mission	LNEC's mission is to undertake, coordinate, and promote scientific research and technological development, aiming for the continuous improvement and the good practice of civil engineering
Values	Excellence; Impartiality; Rigor; Responsibility
Vision	To be a reference in the various fields of civil engineering and related areas

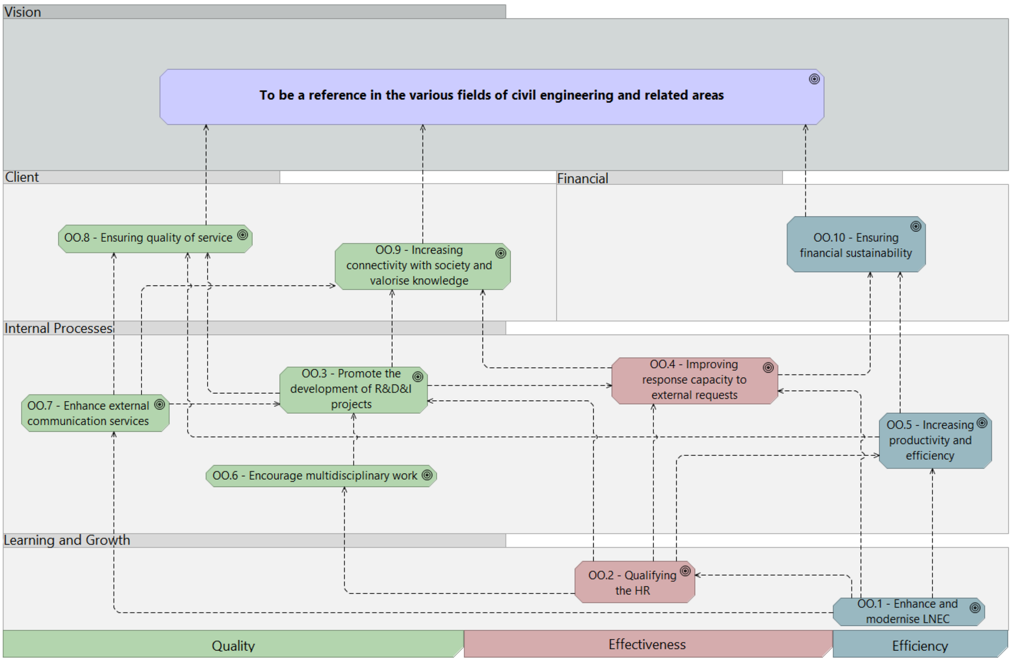


Fig. 4. LNEC's strategy map. Adapted from [1].

[1], designed in Archi[4] and reproduced in Figure 4. The institutional report [1] contains an overview of LNEC's current strategic context using **Political, Economic, Social, Technological, Environmental, and Legal (PESTEL)**; **Strengths, Weaknesses, Opportunities, and Threats (SWOT)**; and TOWS (or reverse SWOT) analyses. Additionally, it also includes a risk management analysis.

The strategy formulation of LNEC includes the mission, values, and vision statements (see Table 1), aligned with QUAR's formulation. The definition of a strategic objective is distinct in a BSC and in QUAR. Contrarily to what the BSC proposes, QUAR strategic objectives are not quantified. Thus, the QUAR strategic objectives could, at most, be considered as BSC strategic themes. However, in the LNEC strategy map, it was chosen to use the three objectives categories prescribed by QUAR as vertical strategic themes, crossing the perspectives of the BSC: Quality, Effectiveness, and Efficiency (see Figure 4).

The LNEC's strategy for 2021–2027 was mapped into 10 strategic objectives (based on QUAR's operational objectives), divided into four perspectives (financial, customers, internal processes, and learning and growth). The strategy map design also comprises the definition of cause-and-effect relationships between the strategic objectives, considering the above-mentioned strategic themes. For each strategic objective, a set of indicators

[4]https://www.archimatetool.com/

was defined, with 22 indicators being defined in total. Some (nine) of these indicators are directly related to QUAR indicators. The yearly activity plans, available on the LNEC's website,[5] present target values and yearly values for QUAR indicators.

5 ONTOLOGY MAPPING AND POPULATION

Instance data was added to the BSO, allowing strategy validation and analysis. The process of incorporating instances into the ontology, referred to as ontology population, was achieved using a Protégé plugin known as Cellfie.[6] Cellfie enabled us to establish a collection of import rules and mappings, utilizing the Manchester OWL Syntax,[7] to translate data from Excel spreadsheets into OWL axioms. Ultimately, a strategy map and a set of quantification frameworks related to LNEC's BSC were imported into the ontology.

5.1 Mapping LNEC's Strategy Map

First, the strategy map and its elements were mapped into ontology entities. The strategy statement elements (mission, vision, and values), BSC perspectives, strategic themes, and strategic objectives were mapped to their respective BSO classes. When appropriate, additional instance metadata was imported using annotation such as RDF Schema's "label" or Dublin Core's[8] "title" and "description." The following properties were used to formalize the strategic objectives relationships:

- **isElementOf**: Relationship between a BSC Element and a BSC Component. In this case, the relationship between a strategic objective and the strategy map
- **contributesToPerspective**: Direct contribution from a Strategic Objective to a Perspective (functional property)
- **contributesToTheme**: Direct contribution from a Strategic Objective to a Strategic Theme
- **hasCauseEffectRelationship**: Direct contribution from a Strategic Objective to another Strategic Objective

5.2 Mapping Quantification Frameworks

Quantification frameworks are defined to evaluate the strategy's execution over a certain time frame, usually a year. Quantification frameworks define indicators used to evaluate the strategy map's strategic objectives. These indicators should be either Lead (i.e., drivers, enablers, predictive) or Lag (i.e., results) indicators. The BSO also defines a number of data relationships to characterize indicators (e.g., data source, acquisition frequency).

LNEC's 22 indicators for the current strategy were added to the ontology, including their codes and description, and related to the quantification frameworks (using the "isElementOf" property). Indicators are connected to strategic objectives using the "isEvaluatedBy" object property. The yearly activity plans publish data concerning 9 indicators from the original 22 introduced in Barateiro et al. [1], all related to QUAR.[9] A quantification framework was created for each activity plan, from 2021 to 2023, and the following relationships were used to complete each framework:

- **hasTarget**: Object property used to relate an indicator with a target. Targets are imported for each of the available indicators with their respective target values and target date (end of the respective year).
- **hasActualValue**: Object property used to relate an indicator with an actual value. Current values for the available indicators are presented in the yearly activity plan and loaded into BSO with their respective date.

[5] https://www.lnec.pt/pt/lnec/instrumentos-de-gestao/documentos-institucionais/
[6] https://github.com/protegeproject/cellfie-plugin
[7] https://www.w3.org/TR/owl2-manchester-syntax/
[8] https://www.dublincore.org/
[9] Quantitative data from a 10th indicator from QUAR is also presented in the activity plans. However, the indicator was divided into two indicators in the current strategy, meaning that the data for these indicators cannot be retroactively obtained.

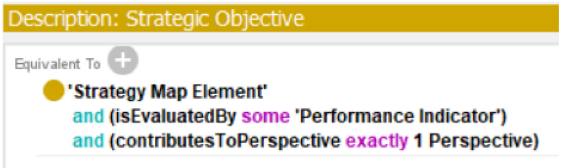


Fig. 5. Strategic objective definition in BSO.

Every year, LNEC defines a set of strategic guidelines to guide LNEC's activities during the year and to help achieve the annual targets. While similar in purpose to BSC's initiatives, the formalization level of these guidelines was considered insufficient when mapping this information into BSO. BSO defines initiatives as strategic projects with a significant organizational commitment (buy-in), with defined dates and resource allocations (e.g., budgets, responsibles), and with impact in at least one strategic indicator of the quantification framework. In this case study, the strategic guidelines are not linked with indicators or strategic objectives, and are usually defined as single-sentence actions. Therefore, they were not loaded as initiatives in the BSO.

6 ONTOLOGY-BASED STRATEGY VALIDATION AND INFERENCE

Once the information from the LNEC's strategy is loaded into the BSO, the information can be validated using SW technologies. Protégé's default reasoner, Hermit,[10] was used to infer new relationships and to detect inconsistencies. No inconsistencies were found using this method. To further explore the benefits of ontological knowledge representation, other technologies were used. **Shapes Constraint Language (SHACL)**[11] and **Semantic Web Rule Language (SWRL)**[12] are two technologies within the SW framework that can be useful in validating the BSO.

SHACL is a constraint language designed to validate the structure and the data of RDF graphs used to represent ontologies. SHACL allows the definition of shapes (templates) that describe the ontology's expected structure. This includes specifying the necessary classes and properties, their cardinality, and the relationships between them. These specifications serve as a safeguard, ensuring that the ontology statements follow the predefined model. Moreover, SHACL can also be used for data validation within the ontology. For example, it enables the definition of constraints that require certain properties to have specific data types, ranges, or formats, which help maintain data integrity and accuracy in the ontology. SHACL also provides a mechanism for checking constraints on the ontology that can identify violations, such as missing data or data that does not conform to defined rules.

The SHACL shapes can be used to validate information in the BSO. For example, a strategic objective is defined in the BSO as shown in Figure 5, stating that any strategic objective needs to be related to at least one value indicator, which would be the case if any of the QUAR strategic objectives were included in the strategy map. In that case, Hermit would not classify the ontology as inconsistent due to the Open World Assumption used in semantic languages such as OWL. SHACL could be used to warn users that these objectives should be evaluated by a set of indicators. Listing 1 presents an example of a shape that ensures that any strategic objective is evaluated by at least one indicator. The QUAR strategic objectives would be identified with a warning message (see Figure 6) since they do not have any indicator.

SWRL is an expressive rule language that allows users to define complex semantic rules for reasoning over ontologies. SWRL can also be used to check for inconsistencies or ensure data quality in the ontology. However, SWRL is known for its inference capabilities that can help derive implicit information from an ontology. This can

[10] http://www.hermit-reasoner.com/
[11] https://www.w3.org/TR/shacl/
[12] https://www.w3.org/submissions/SWRL/

SHACL constraint violations: 4

Severity	Message	FocusNode
http://www.w3.org/ns/shacl#Warning	A Strategic Objective must be evaluated by at least one Indicator	https://www.iscte-iul.pt/ontologies/BSO#OE.1
http://www.w3.org/ns/shacl#Warning	A Strategic Objective must be evaluated by at least one Indicator	https://www.iscte-iul.pt/ontologies/BSO#OE.2
http://www.w3.org/ns/shacl#Warning	A Strategic Objective must be evaluated by at least one Indicator	https://www.iscte-iul.pt/ontologies/BSO#OE.3
http://www.w3.org/ns/shacl#Warning	A Strategic Objective must be evaluated by at least one Indicator	https://www.iscte-iul.pt/ontologies/BSO#OE.4

Fig. 6. SHACL constraints violations resulting from Listing 1's shape.

Listing 1. SHACL Shape Example

```
ObjectiveIndicatorValidation
        a sh:NodeShape ;
        sh:targetClass Strategic_Objective;
        sh:property [
                sh:path isEvaluatedBy;
                sh:minCount 1;
                sh:severity sh:Warning;
                sh:message "A Strategic Objective must be evaluated by at least one
                    Indicator";].
```

be valuable in identifying missing or implied relationships between entities in the BSO. For example, the layout of perspectives in the strategy map is defined in BSO using the "isBaseFor" object relationship. This means that if perspective A "isBaseFor" perspective B, perspective A should be visually presented below perspective B. This relationship also implies that perspective A is a driver for perspective B, in the sense that the performance of strategic objectives in perspective A drives the performance of objectives in perspective B.

Listing 2. SWRL Rule for isBaseFor Inference

```
contributesToPerspective(?o1, ?p1) ^ contributesToPerspective(?o2, ?p2) ^
hasCauseEffectRelationship(?o1, ?o2) ^ differentFrom(?p1, ?p2) ->
isBaseFor(?p1, ?p2)
```

While the "isBaseFor" object property was not explicitly stated when loading LNEC's information to the ontology, SWRL can be used to infer this connection between perspectives. Using the rule shown in Listing 2, the "isBaseFor" object property is inferred based on the cause-and-effect relationships of objectives related to each perspective. The "isBaseFor" object property is asymmetric and irreflexive, and, although it is theoretically transitive, it was not formalized as such because OWL reasoners could not infer over complex properties [38]. However, using SWRL, a transitive rule can be used to achieve the same effect, which allows the rule engine to correctly infer this relationship between perspectives. The results of this inference can be seen in Figure 7, where the "isBaseFor" object property was correctly inferred between the learning and growth and the other perspectives.

In summary, SWRL and SHACL play complementary roles in validating the BSO. Leveraging both technologies ensures that the ontology is not only semantically correct and complete but also structurally and data-wise compliant with the BSO model, ultimately leading to a more reliable and accurate representation of a BSC.

Over the years, Kaplan and Norton have perfected an adaptable performance management framework that allows executives and managers to design and use the BSC with the level of detail required by their organizational strategy. However, due to this adaptability, there is not a standard formalization of rules defining what constitutes a complete and well-defined BSC. To this end, the BSO was defined as a way to formalize the BSC framework, which can now be complemented with SWLR rules and SHACL shapes.

Caldeira [5] presents a simple set of rules for strategic objectives and cause-and-effect relationships between them in the Portuguese civil services. Cardoso [7] also presents a set of similar rules, but generic for any BSC,

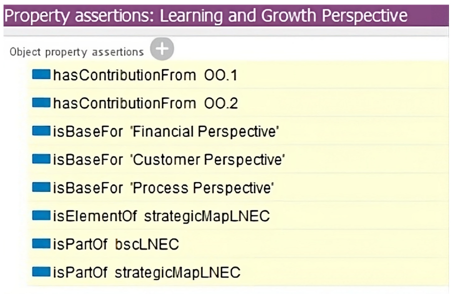


Fig. 7. Learning and growth perspective inferred properties.

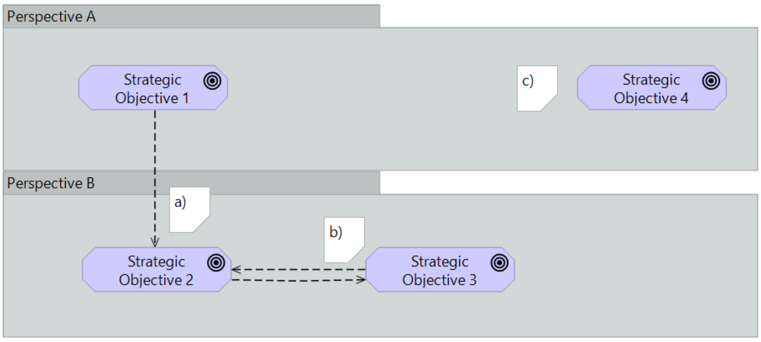


Fig. 8. Set of validations for strategy map's cause-and-effect relationship. Adapted from Caldeira [5] and Cardoso [7].

with dos and don'ts of the cause-and-effect relationships definition. These guidelines (see Figure 8) are important since they represent strategic hypotheses in terms of the expected impact of the performance of each objective. Monitoring these relationships is challenging and time-consuming and has been appointed for many years as one of the limitations of the BSC [34, 43]. Using the ontology (OWL), SWRL, and SHACL, any strategy, such as the one from the LNEC's case study, can now be validated against these rules. Table 2 showcases how each rule can be validated using the aforementioned technologies.

7 STRATEGY ANALYSIS

The populated ontology and the inferred axioms were exported from Protégé and imported to GraphDB,[13] a linked data environment compliant with W3C standards (i.e., RDF, OWL, SPARQL). Once stored in this semantic graph database, the ontology can be queried or updated using SPARQL endpoints. While not strictly necessary for this analysis, GraphDB allows the ontology to be accessed by external applications, such as Power BI,[14] a **Business Intelligence (BI)** tool that allows users to create interactive reports and dashboards, which can be used to visualize the strategy execution.

The BSO can be now be queried to obtain information either explicit or implicit in the ontology. Queries such as the ones presented in Listing 3 and 4 allow the user to explore information related to the ontology. The first query returns information related to the main elements of the strategy map, namely the strategic objectives, perspectives, and strategic themes. The second query returns any strategic objective that is directly or indirectly

[13] https://graphdb.ontotext.com/
[14] https://powerbi.microsoft.com/

Table 2. Validation of Rules of Cause-and-effect Relationships in a Strategy Map

Rule	Description	Technologies	Validation
a	Cause-and-effect relationships should not be defined in a descendent direction, since they contradict the logic of the model.	OWL, SWRL	SWRL infers "isBaseFor" object property based on cause-and-effect relationships between strategic objectives (see Listing 2). If there is a downward relationship, the rule will infer that perspective A "isBaseFor" perspective B and vice versa. Due to the asymmetric characteristic of this property, the reasoner will find the ontology to be in an inconsistent state.
b	Cause-and-effect relationships should not be defined with two-directional arrows, because it becomes impossible to distinguish the cause from the effect.	OWL	hasCauseEffectRelationship object property is asymmetric and irreflexible.
c	"Orphans" objectives, i.e., not linked to any objective, should not exist, because they do not express how they can contribute to the strategy.	SHACL	Shape for strategic objectives ensures that the objective is either in range or in the domain of at least one hasCauseEffectRelationship property.

Adapted from Cardoso [7].

(transitive property) affected by a certain objective (the notation of a class name between angle brackets (<>), e.g., $< Strategic\_Objective >$ is used to define any instance of that class).

Listing 3. SPARQL Query for Strategic Objectives

```
SELECT ?perspective ?objectiveID ?theme
WHERE{
    ?objectiveID rdf:type Strategic_Objective;
                 contributesToPerspective ?perspective;
                 contributesToTheme ?theme.
} ORDERBY ?perspective ?objectiveID
```

Listing 4. SPARQL Query for Strategic Objectives' Cause-and-effect Relationships

```
SELECT ?objectiveID
WHERE{
    <Strategic_Objective> rdf:type Strategic_Objective;
              isElementOf ?StrategicMap;
              hasCauseEffectRelationship+ ?objectiveID.
    ?StrategicMap hasElement ?objectiveID.}
```

Table 3 presents the Quantification Framework for 2022, based on the activity plan for the same year. For each objective, it includes the indicator's target value, the actual value achieved, and the result status. The results reveal the performance against the predefined targets, with some indicators labeled as failed (for falling short of expectations) and others marked as success (for meeting or exceeding their targets). There are no target or actual values available for 13 indicators, and three objectives (OO.4, OO.7, and OO.8) lack associated indicators for evaluation. It should be noted that although indicators have been defined for these objectives, no target or actual values are publicly available. Indicator 5 (related to the monthly average of research grants) associated with objective OO.2 (Qualify HR) also had a substantial shortfall compared to its target (only achieving 10% of the target). However, notwithstanding the failed result, indicators 8 (OO.3) and 22 (OO.10) achieved more than 95% of their corresponding targets.

Access to real data enables a data-driven assessment and in-depth analysis of both the execution of a strategy and the quality of its formulation, i.e., evaluating whether the strategy was well defined and if it is delivering the intended results toward the fulfillment of the vision (future position). For instance, when managers

Table 3. Analysis of Performance Indicators Based on the 2022 Activity Plan

Objective	IndicatorID	Indicator	Target Value	Actual Value	Result
OO.1	Ind.1	Investment in research infrastructure/total expenditure	0.09	0.064	failed
OO.2	Ind.5	Number of contracts for junior researchers and LNEC research grants (monthly average)	40	4	failed
OO.3	Ind.8	Number of strategic research studies in partnership with other entities	65	62	failed
OO.3	Ind.9	Percentage of external funding for Strategic Research relative to total expenses	0.075	0.093	**success**
OO.5	Ind.11	Number of technical publications (reports, technical notes, opinions, etc.) per researcher	3.7	3.07	failed
OO.6	Ind.13	Total revenue from internal processes/total revenue	280	375	**success**
OO.9	Ind.18	Own revenue from contract activities linked to research projects/total expenditure	70	86	**success**
OO.9	Ind.19	Number of scientific and technical events organized or co-organized by LNEC	110	135	**success**
OO.10	Ind.22	Percentage of self-financing amount relative to total expenses	0.45	0.449	failed

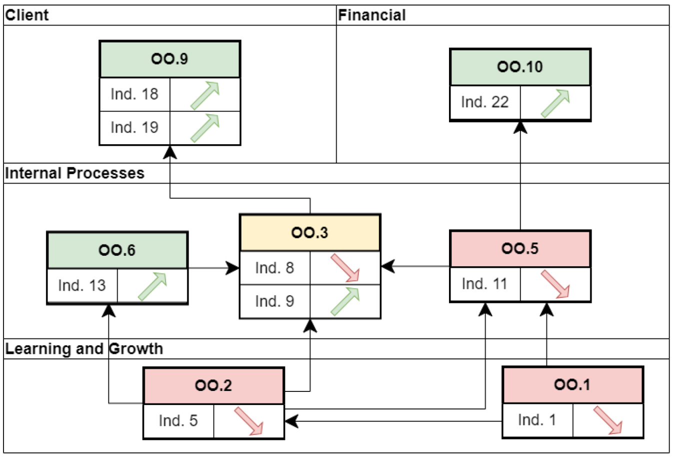


Fig. 9. LNEC's strategy map with trend analysis between 2021 and 2022 (excluding objectives that are not being measured).

establish cause-and-effect relationships among strategic objectives, they are essentially formulating hypotheses about how these objectives and their associated indicators are interconnected (e.g., if customer satisfaction improves, it should lead to increased sales; if production costs decrease, it should result in higher profits). By linking the LNEC's strategy map with its quantification frameworks, the set of strategic hypotheses can now be empirically validated. Figure 9 displays the strategic objectives, cause-and-effect relationships, and measured indicators,[15] together with their growth trend, related to the evolution shown between the 2021 and 2022 quantification frameworks. The color of strategic objectives and indicators represent their growth trend: green—positive, red—negative, and yellow—inconclusive. Without defining the weights of the impact for each indicator, OO.3's trend growth cannot be determined.

When a strategic objective is linked to another, this relationship can be extended to their respective indicators. That is, if the performance of the indicator measuring the "cause" objective increases, it is expected that

[15]Note that Figure 9 is a simplification since indicators are not part of the strategy map or its visualization.

the performance of the indicators measuring the "effect" objective will also improve. The cause-and-effect relationships assessment need to take into account the weights of indicators measuring a particular objective. In Figure 9, if a particular indicator demonstrates an upward trend across these quantification frameworks, signifying an increase in value, it is anticipated that the indicators related to the affected objective will exhibit a similar trend. It is worth noting that, in the LNEC's case, this analysis is simplified since all indicators have a positive polarity. The BSO enables the definition of the indicator's polarity using a data property. The same can be expected in the opposite situation, as shown in Figure 9 for all OO.1 cause-and-effect relationships. However, all other objectives and indicators present an unexpected behavior, with some type of disparity in cause-and-effect relationships. There can be several underlying factors contributing to these observed disparities:

— **Inaccurate cause-and-effect relationships formulation.** The cause-and-effect relationship is not proven by the data itself, meaning that the anticipated connections and relationships between strategic objectives could not be substantiated by the data (e.g., relationships between OO.2 and OO.6 or OO.5 and OO.10). This can happen because the cause-and-effect link between the objectives does not exist after all, or because this relationship is not as direct as initially formulated (a third strategic objective might be needed to correctly depict the organization's reality).
— **Inaccurate or incomplete objective evaluation.** While cause-and-effect relationships may exist between objectives, the accurate evaluation of a strategic objective can be compromised by the absence of an indicator or misformulation of the indicator (wrong level of detail/aggregation or wrong context). For example, the cause-and-effect relationships from OO.2 to OO.6 or OO.5 to OO.10 might exist, but the associated indicators do not allow an accurate or complete evaluation of the objectives.
— **Impact of multiple cause-and-effect relationships.** Strategic objectives and their indicators can be influenced by a multitude of sources. However, it is important to recognize that not all indicators respond uniformly to their strategic objective's cause-and-effect relationships. The impact of a cause-and-effect relationship may be partly observed in some of the indicators related to a strategic objective, while the remaining indicators are affected by different cause-and-effect links originating from other strategic objectives and indicators. Furthermore, even if the cause-and-effect relationships between these indicators could be verified, there may be a significant variation in the impact of each cause in the indicator. This complexity is especially pronounced in the analysis of OO.3's cause-and-effect relationships, for example, as the direct impact of each indicator lacks formalization, especially the weights of each relationship. For example, Ind. 8 and 9 are affected by OO.2, OO.5, and OO.6. Observing the Ind.9 growth trend, different scenarios can be possible. For instance, Ind. 9's growth trend can be fully explained by OO.6's performance, with OO.2 and OO.5 only affecting Ind. 8; or Ind. 9's may be impacted by all of the objectives, with OO.6's performance having a higher impact on this indicator than the other two objectives combined. A similar logic can be applied to the cause-and-effect relationship between OO.3 and OO.9.
— **Data issues.** In some cases, organizations may lack comprehensive data to fully evaluate the impact of cause-and-effect relationships. This can be due to data gaps, measurement challenges, or limited historical information. Incomplete or incorrect data can hinder the accurate assessment of the performance of strategic objectives.

Finally, the knowledge contained in the ontology and all the queries and analyses presented in this section can be used by external applications to create visualizations and provide recommendations. GraphDB allows the creation of access points that can be used for applications via REST API or similar software interfaces. The information in Table 3 was retrieved for all available quantification frameworks and exported from GraphDB to PowerBI. Figure 10 presents a dashboard that can be used for strategy execution analysis, based on the BSO knowledge. The average target completion percentage is presented for each strategic objective evolution, together with the evolution of each indicator with the respective target defined in the quantification frameworks (filtered for 2021 and 2022).

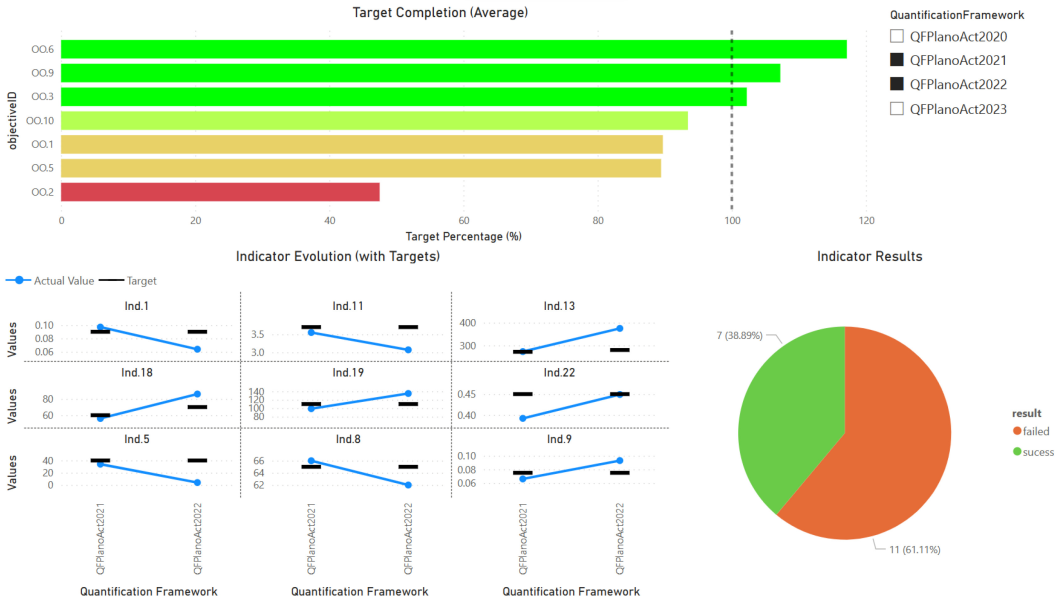


Fig. 10. Dashboard for strategy execution analysis in power BI.

7.1 Recommendations

The absence of data for a significant portion of the defined indicators has a notable adverse impact on the effectiveness of the strategic analysis. Incomplete data can hinder the ability to assess strategy success and the true performance of strategic objectives and their associated indicators. It may result in an incomplete or inaccurate understanding of the organization's progress toward its goals, potentially leading to sub-optimal decision-making.

On the other hand, the frequency of mandatory reporting, while providing a consistent basis for analysis, might be too long for an adequate and timely performance evaluation. This can be a limitation that affects an organization's ability to make real-time decisions and adapt as needed to achieve their intended targets. The strategy execution evaluation should monitor and report on real data managed by organizational information systems. Ideally, the values should be retrieved from information systems, such as BI systems, and loaded into the ontology using an automated or semi-automated process. Implementing a comprehensive data collection and respective reporting mechanisms enables an accurate and continuous evaluation of the strategy execution, leading to the realization of an integrated strategic management and execution system, and allowing for timely adjustments and interventions throughout the year.

Furthermore, it is imperative that LNEC's vision can be quantifiable and measured. The BSO states that an organizational vision contains a quantifiable stretch-goal, with a well-defined niche, time limit, and target. The stretch-goal establishes a performance indicator and a target value with a clear time frame to achieve it, enabling a clear quantification of the vision statement for the specific strategy cycle. Without a quantifiable stretch-goal, managers cannot measure how close or how far they are to their future expected position, and therefore cannot take action to ensure the long-term organization's success. The niche clarifies the scope of the strategy, defining the boundaries that will guide the organization's actions.

There is room for improvement in the identification and definition of strategic guidelines. The formalization of these guidelines as actual BSC initiatives would provide real action plans that operationalize the strategy to

Table 4. Strategic Guidelines Impact by Indicator

Objective	Indicator	Growth %	Trend	Number of Guidelines
OO.1	Ind.1	−33.85	decreasing	2
OO.2	Ind.5	−88.24	decreasing	1
OO.3	Ind.8	−6.06	decreasing	3
OO.3	Ind.9	39.46	increasing	3
OO.5	Ind.11	−13.52	decreasing	0
OO.6	Ind.13	37.36	increasing	1
OO.9	Ind.18	53.57	increasing	0
OO.9	Ind.19	36.36	increasing	3
OO.10	Ind.22	14.62	increasing	0

help achieve a set of targets, leading to an effective strategy execution. These strategic projects must have a clear resource allocation, such as a budget, and outline the strategic objectives and indicators that will be impacted. By importing strategic initiatives to the ontology, the impact of each initiative could be validated both in its formulation and influence on indicator performance. For example, a SHACL shape similar to the one presented in Listing 1 would reveal which initiatives do not impact any indicators. Initiatives' impact on each indicator could be analyzed, similar to the analysis in Table 4, which shows the indicator performance trend between 2021 and 2022, and the number of strategic guidelines defined in the quantification framework for 2022 associated with each indicator.[16] There is only one indicator (Ind. 11, from objective OO.5) that has a downward tendency and is not the target of a strategic guideline. Furthermore, defining budget allocation for strategic initiatives would enable a real and accurate analysis of the organizational commitment to each objective, providing greater insight into the growth trend of each indicator.

The observation of a downward growth trend in the base perspective (learning and growth perspective) in Figure 9 may be a cause for concern. This may indicate a potential lack of investment in critical areas that serve as the foundation for achieving other strategic objectives. Neglecting these foundational elements can have cascading effects on overall performance and hinder future achievements.

In conclusion, addressing these types of issues is essential to enhance the effectiveness of strategic analysis and improve the ability to achieve strategic objectives. These recommendations can contribute to a more robust and data-driven strategic management approach, increasing the organization's ability to adapt and improve.

8 DISCUSSION

This research explores how public organizations, such as LNEC, can employ ontologies to assess their strategy formulation and execution. The BSO was used to validate the formulated LNEC's strategy, using the OWL reasoner, SHACL, and SWRL to find inconsistencies and allow inference over the strategic knowledge. The use of these technologies ensures that the ontology is not only semantically correct and complete but also structurally and data-wise compliant with the BSO model, ultimately leading to a more reliable and accurate representation of the BSC framework.

Furthermore, it investigates how the ontology can be used to assess the implementation and execution of the strategy. The quantification framework analysis can be done to evaluate performance indicators and validate the cause-and-effect relationships between strategic objectives. Formulating a strategy, especially the cause-and-effect relationships from the BSC, is a complex and subjective process that relies on the creation of hypotheses based on managers' knowledge, insights, and "gut feelings." With the BSO, the design of these relationships can now be improved, since the ontology enables their monitoring and empirical validation. Lastly, SW technologies offer increased interoperability that can be used to share information across systems and organizations.

[16]This association was not provided by any official LNEC document. It was obtained manually to provide an example for this analysis.

8.1 Contributions to the Theory

The present study contributes to the existing literature by addressing various identified gaps associated with strategy analysis in the public sector, specifically the low comprehensiveness and formality of strategic management systems [4]. This research explores how ontologies can be used to overcome this gap, enabling data-driven decision and improving strategy formulation and evaluation. As shown in previous research [15, 17, 18, 32], SW technologies are one of the emerging technologies in the public sector, providing shared semantics and enabling interoperability across public organizations.

The BSO ontology was previously developed to formalize the BSC [31], which is the most recognized approach for performance assessment in public administration [29, 43], providing insight into financial and non-financial objectives and indicators, addressing the often overlooked non-financial aspects of the public sector [32]. The BSO was developed to enable accurate, traceable, and continuous monitoring and improvement of the strategy execution, based on a data-driven approach. Existing studies, such as those by Kumar et al. [30] and Tawse and Tabesh [43], also emphasize the importance of combining the BSC with other systems and tools for an effective implementation, which can be facilitated through the increased formalization and interoperability from the ontology.

Using the proposed methodology (see Section 3), SW technologies can be used to support strategy management as long as the organization's strategy is formulated using a BSC. Formalizing knowledge through techniques such as ontologies offers several benefits, including enhanced interoperability between systems, knowledge validation and inference, and improved communication through semantics [40]. These benefits can be leveraged to address identified gaps in the literature, such as ensuring effective implementation of the BSC as a strategic management system and fully harnessing the framework's benefits [31, 43]. Additionally, SW technologies enable continuous performance evaluation, creating clear and understandable performance management systems (minimizing misunderstandings inside and outside the organization) [29].

8.2 Contributions to the Practice

Using the BSO and SW technologies can offer multiple advantages for managers and decision-makers. Primarily, it ensures the alignment between the BSC and the organization's overarching goals. The ontological structured representation allows managers to assess whether elements within the BSC align with the organization's strategy, preventing the misallocation of resources to nonessential or redundant indicators and objectives.

The BSO provides an unambiguous representation of the BSC framework, promoting a shared understanding of the strategy, strategic objectives, and indicators among all stakeholders. This enhanced clarity can improve communication and alignment across the organization, spanning different organizational levels or departments and serving as a valuable tool for knowledge transfer within the organization. By formalizing the cascading impact of each BSC element, the contribution of individual or departmental objectives to the overall organizational strategy can be clarified, offering a deeper understanding of the strategic framework to employees and stakeholders, potentially acting as a motivational factor.

Moreover, SW technologies play a crucial role in supporting compliance and governance initiatives by enhancing external communication and alignment. SW technologies streamline documentation and reporting and enable managers to verify organizational adherence to regulatory requirements and compliance with relevant standards, such as those established by policymakers. This is particularly significant for policy compliance, reporting, and performance evaluation in the context of international policies and public administration. For example, the European Commission can provide a set of rules (formalized using SWRL and SHACL) that can be validated against an organization's strategy and its execution.

The European Green Deal[17] is a perfect example of the potential application of this research contribution. The European Green Deal defines several policy initiatives for climate neutrality, including legal obligations, such

[17] https://www.consilium.europa.eu/en/policies/green-deal/

as the European climate law, where the EU and its member states are committed to cutting net greenhouse gas emissions in the EU by at least 55% by 2030, compared to 1990 levels.

Furthermore, integrating the BSO into the strategic decision-making process ensures the alignment of all decisions with the organization's mission, vision, and strategic objectives. This proactive approach prevents decisions that may not contribute to the organization's long-term success. The BSO facilitates dependency analysis, providing insights into how changes in a specific indicator or objective may impact the overall strategy. Additionally, using an ontological model can streamline the integration of data from various sources into the BSC model, making the collection, analysis, and reporting of performance indicators more efficient. The automation of these processes is crucial for supporting real-time or near-real-time monitoring of performance indicators and data-driven decision support. This contribution is particularly significant given the growing importance of leveraging data in strategic decision-making processes within an evolving business environment, as highlighted by Grant [12].

8.3 Limitations and Future Work

This research explored how public organizations can take advantage of SW technologies to assess their strategy formulation and execution, based on a real case study of a public Portuguese organization (LNEC). While there are no perceived barriers to generalizing and implementing the proposed methodology in other scenarios, the adaptability of this work must be explored in other contexts or industries (public or private). This is a necessary step to ensure that the only restriction to the reproducibility of a similar analysis using the BSO is the translation of the organizational strategy into a BSC.

Additionally, it is essential to note that while most of the methodology processes can be automated (namely, the use of SW technologies for validation and analysis), the ontology mapping and population is still a semi-automated process since knowledge is usually retrieved from non-structured data. While import rules (or other existing ontology population methods) can be defined based on a data template, data related to the BSC is usually only available in non-structured documents, leading to a case-by-case mapping process. Following the same premise, the BSC framework's adaptable nature poses challenges in employing SW technologies such as SWRL and SHACL, as their application will depend on the level of detail and validation rules required by each manager or organization (see Section 6).

As discussed in Section 7, ontologies and SW technologies can be used to enable strategy validation and analysis for managers and decision-makers. For example, Figure 10 shows how external applications can benefit from ontology knowledge to provide end-users with strategic information, using a simple visualization. However, additional work is required to fully showcase the potential of these technologies as part of a fully integrated solution or framework where managers can interact, monitor, analyze, and receive alerts or even analytical recommendations regarding their strategy implementation and execution.

Furthermore, data access presents a notable challenge regarding strategy evaluation. As emphasized earlier, monitoring and analyzing the strategy execution requires actual data managed by the organizational information systems. However, establishing a seamless relationship between this data encompassing values collected for each indicator and their ontological representation is intricate. The complexity arises from variations in indicators defined at different levels of detail. Values should be extracted from information systems, like BI systems, and integrated into the ontology through an automated or semi-automated process. Overcoming these limitations is crucial for ensuring the robustness and applicability of the ontology and methodology in supporting an effective strategy execution and decision-making.

9 CONCLUSION

This article presents an applied research based on the impact and potential of SW technologies, such as ontologies, in the assessment of strategy formulation and execution in public sector strategy management. The LNEC was used as a representative case study in this research, which ultimately tries to enhance organizational performance and enable accurate, traceable, and continuous monitoring of an organization's strategy.

The BSO, an ontology design to describe and store knowledge related to the BSC framework, was used to validate the strategy formulation from LNEC. When complemented by semantic technologies such as SHACL and SWRL, the BSO can be used to validate any set of rules and ensure that the ontology is consistent and is structurally and data-wise compliant with the BSC model. The BSO can also be used to evaluate performance indicators and monitor or validate cause-and-effect relationships between strategic objectives. Lastly, the BSO increases the interoperability of strategic information.

In public sector organizations, like LNEC, the efficient use of resources directly impacts society at large, so there is a pressing need to enhance strategic decision-making and resource allocation. This research showcases the potential of ontology-driven strategic analysis to enhance organizational efficiency, adaptability, and decision-making capabilities while ensuring a shared understanding of strategies and data. Ultimately, this research offers a blueprint for public sector organizations seeking to optimize their strategies (i.e., more informed, efficient, and impactful strategies), foster transparency, and deliver more effective services to the public they serve.

Automating strategy analysis and connecting it to data is a critical need in today's dynamic business landscape. It promises enhanced efficiency, real-time insights, and the ability to handle the complexity of modern organizations. Public sector organizations, such as LNEC, stand to benefit significantly from this approach, as it ensures that resources are allocated effectively, strategies remain adaptable, and decisions are data driven.

REFERENCES

[1] José Barateiro, Paula Couto, António Antunes, António Sebastião, and Maria Alzira Santos. 2021. *Enquadramento Estratégico do LNEC—Proposta de Mapa Estratégico.* Technical Report. National Laboratory for Civil Engineering (LNEC), Lisbon, Portugal. In Portuguese.

[2] Frida Betto, Alberto Sardi, Patrizia Garengo, and Enrico Sorano. 2022. The evolution of balanced scorecard in healthcare: A systematic review of its design, implementation, Use, and review. *International Journal of Environmental Research and Public Health* 19, 16 (2022), 10291.

[3] John Bryson and Bert George. 2020. Strategic management in public administration. In *Oxford Research Encyclopedia of Politics.*

[4] John M. Bryson. 2018. *Strategic Planning for Public and Nonprofit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement.* John Wiley & Sons.

[5] Jorge Caldeira. 2009. *Implementação do Balanced Scorecard no Estado.* Edições Almedina, Portugal.

[6] Mark Anthony Camilleri. 2021. Using the balanced scorecard as a performance management tool in higher education. *Management in Education* 35, 1 (2021), 10–21.

[7] E. Cardoso. 2011. *Performance and Quality Management of Higher Education Programmes.* Ph. D. Dissertation. University Institute of Lisbon (ISCTE-IUL).

[8] Elsa Cardoso and Maria José Trigueiros. 2007. Using the balanced scorecard as a tool for performance management of higher education degrees. In *Proceedings of the 13th International Conference of European University Information Systems (EUNIS'07).* 27–29.

[9] Tahniyath Fatima and Saïd Elbanna. 2020. Balanced scorecard in the hospitality and tourism industry: Past, present and future. *International Journal of Hospitality Management* 91 (2020), 102656. https://doi.org/10.1016/j.ijhm.2020.102656

[10] Raquel Fernandez-Gonzalez, Felix Puime-Guillen, and Jorge Eduardo Vila-Biglieri. 2023. Environmental strategy and the petroleum industry: A sustainability balanced scorecard approach. *Journal of Petroleum Exploration and Production Technology* 13, 2 (2023), 763–774.

[11] Bert George, Richard M. Walker, and Joost Monster. 2019. Does strategic planning improve organizational performance? A meta-analysis. *Public Administration Review* 79, 6 (2019), 810–819.

[12] Robert M. Grant. 2021. *Contemporary Strategy Analysis.* John Wiley & Sons.

[13] Nicola Guarino, Daniel Oberle, and Steffen Staab. 2009. What is an ontology? In *Handbook on Ontologies.* Springer, 1–17.

[14] Erik G. Hansen and Stefan Schaltegger. 2016. The sustainability balanced scorecard: A systematic review of architectures. *Journal of Business Ethics* 133 (2016), 193–221.

[15] Teresa M. Harrison and Theresa A. Pardo. 2020. Data, politics and public health: COVID-19 data-driven decision making in public discourse. *Digital Government: Research and Practice* 2, 1 (2020), 1–8.

[16] Pascal Hitzler. 2021. A review of the semantic web field. *Communications of the ACM* 64, 2 (2021), 76–83.

[17] Karuna Pande Joshi and Srishty Saha. 2020. A semantically rich framework for knowledge representation of code of federal regulations. *Digital Government: Research and Practice* 1, 3 (2020), 1–17.
[18] Evangelos Kalampokis, Nikos Karacapilidis, Dimitris Tsakalidis, and Konstantinos Tarabanis. 2023. Understanding the use of emerging technologies in the public sector: A review of horizon 2020 projects. *Digital Government: Research and Practice* 4, 1 (2023), 1–28.
[19] Robert S. Kaplan and David P. Norton. 1996. Linking the balanced scorecard to strategy. *California Management Review* 39, 1 (1996), 53–79.
[20] Robert S. Kaplan and David P. Norton. 2006. *Alignment: Using the Balanced Scorecard to Create Corporate Synergies*. Harvard Business Press.
[21] Robert S. Kaplan and David P. Norton. 2008. *The Execution Premium: Linking Strategy to Operations for Competitive Advantage*. Harvard Business Press.
[22] Robert S. Kaplan and David P. Norton. 1992. *The Balanced Scorecard: Measures That Drive Performance*. Vol. 70. Harvard Business Review US.
[23] Robert S. Kaplan and David P. Norton. 2001. *The Strategy-focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Harvard Business Press.
[24] Ralf Klischewski. 2003. Semantic web for e-Government. In *International Conference on Electronic Government*. Springer, 288–295.
[25] Ralf Klischewski and Martti Jeenicke. 2004. Semantic web technologies for information management within e-government services. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.* IEEE, 10–pp.
[26] Ralf Klischewski and Stefan Ukena. 2007. Designing semantic e-Government services driven by user requirements. *Proceedings of the 6th International EGOV Conference on Ongoing Research, Project Contributions and Workshops*.
[27] Ralf Klischewski and Stefan Ukena. 2009. A value network analysis of automated access to e-government services. (2009).
[28] Ralf Klischewski and Stefan Ukena. 2010. E-government goes Semantic Web: How administrations can transform their information processes. In *Semantic Technologies for E-Government*, T. Vitvar, V. Peristeras, and K. Tarabanis (Eds.). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-03507-4_5
[29] Martin Král. 2022. 20-year history of performance measurement in the local public sector: A systematic review. *International Journal of Public Administration* 45, 9 (2022), 726–740.
[30] Jitender Kumar, Neha Prince, and H. Kent Baker. 2022. Balanced scorecard: A systematic literature review and future research issues. *FIIB Business Review* 11, 2 (2022), 147–161.
[31] António Lorvão Antunes, Elsa Cardoso, and José Barateiro. 2023. The balanced scorecard ontology: A semantic approach to enhance strategy management. Preprint available at SSRN: https://ssrn.com/abstract=4592135 or http://dx.doi.org/10.2139/ssrn.4592135.
[32] Francesca Manes-Rossi, Giuseppe Nicolò, and Daniela Argento. 2020. Non-financial reporting formats in public sector organizations: A structured literature review. *Journal of Public Budgeting, Accounting & Financial Management* 32, 4 (2020), 639–669.
[33] Petar Milić, Nataša Veljković, and Leonid Stoimenov. 2018. Semantic technologies in e-government: Toward openness and transparency. In *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues*. Springer, 55–66.
[34] Hanne Norreklit. 2000. The balance on the balanced scorecard a critical analysis of some of its assumptions. *Management Accounting Research* 11, 1 (2000), 65–88.
[35] Natalya F. Noy and Deborah L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. *Technical Report. Stanford University*, 25.
[36] Jeff Pan. 2009. Resource description framework. In *Handbook on Ontologies*. Springer, 71–90. https://doi.org/10.1007/978-3-540-92673-3_3
[37] Andreja Pucihar, Kristina Bogataj, and Maria Wimmer. 2007. Gap analysis methodology for identifying future ICT related eGovernment research topics–case of "ontology and semantic web" in the context of eGovernment. (2007).
[38] Michael Schneider, Sebastian Rudolph, and Geoff Sutcliffe. 2013. Modeling in OWL 2 without restrictions. In *Proceedings of the 10th International Workshop on OWL: Experiences and Directions (OWLED 2013) (CEUR-WS)*, 1080 (2013).
[39] Ishani B. Soysa, Nihal P. Jayamaha, and Nigel P. Grigg. 2019. Validating the balanced scorecard framework for nonprofit organisations: An empirical study involving Australasian healthcare. *Total Quality Management & Business Excellence* 30, 9–10 (2019), 1005–1025.
[40] Rudi Studer, V. Richard Benjamins, and Dieter Fensel. 1998. Knowledge engineering: principles and methods. *Data & Knowledge Engineering* 25, 1–2 (1998), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6
[41] Xiaomeng Su and Lars Ilebrekke. 2002. A comparative study of ontology languages and tools. In *International Conference on Advanced Information Systems Engineering*. Springer, 761–765. https://doi.org/10.1007/3-540-47961-9-62
[42] Enrico Supino, Federico Barnabè, Maria Cleofe Giorgino, and Cristiano Busco. 2019. Strategic scenario analysis combining dynamic balanced scorecards and statistics. *International Journal of Productivity and Performance Management* 69, 9 (2019), 1881–1902.
[43] Alex Tawse and Pooya Tabesh. 2023. Thirty years with the balanced scorecard: What we have learned. *Business Horizons* 66, 1 (2023), 123–132.
[44] Stephen Victor and Ayesha Farooq. 2022. Combining the use of data analytics and balanced scorecard to enhance healthcare delivery: A study. *Journal of Health Management* 24, 2 (2022), 248–255.

Received 30 September 2023; revised 8 February 2024; accepted 20 March 2024

CHAPTER 6

Linking Data to Strategy

This chapter describes work related to the fourth DSRM's design interaction (Iter. 4), shown in Figure 6.1. Specifically, this chapter describes the design and development process of two artifacts: (1) the DW Ontology, called Light Data Warehouse Ontology (LDWOWL) (artifact #5), to provide a semantic representation of the DW/BI; and (2) the LDWOWL-BSO Link (artifact #6), as a semantic link between the DW information and the domain-ontology within the framework. Further, the refinement of the Integration Framework (version 3 of artifact #1) and API Services (version 3 of artifact #4) are also detailed. The new version of the Integration Framework aims to support the integration, alignment, and traceability between strategy and the organizational information systems necessary for providing data to the BSC's performance indicators. The formalization of the organizational strategy is ensured by the domain-specific Strategy Ontology (artifact #3). A fifth journal article related to this doctoral thesis is in production. It is titled "Application of Semantic Web Techniques in DW/BI Systems for Strategy Analysis" and highlights the work presented in this chapter and its associated research outputs.

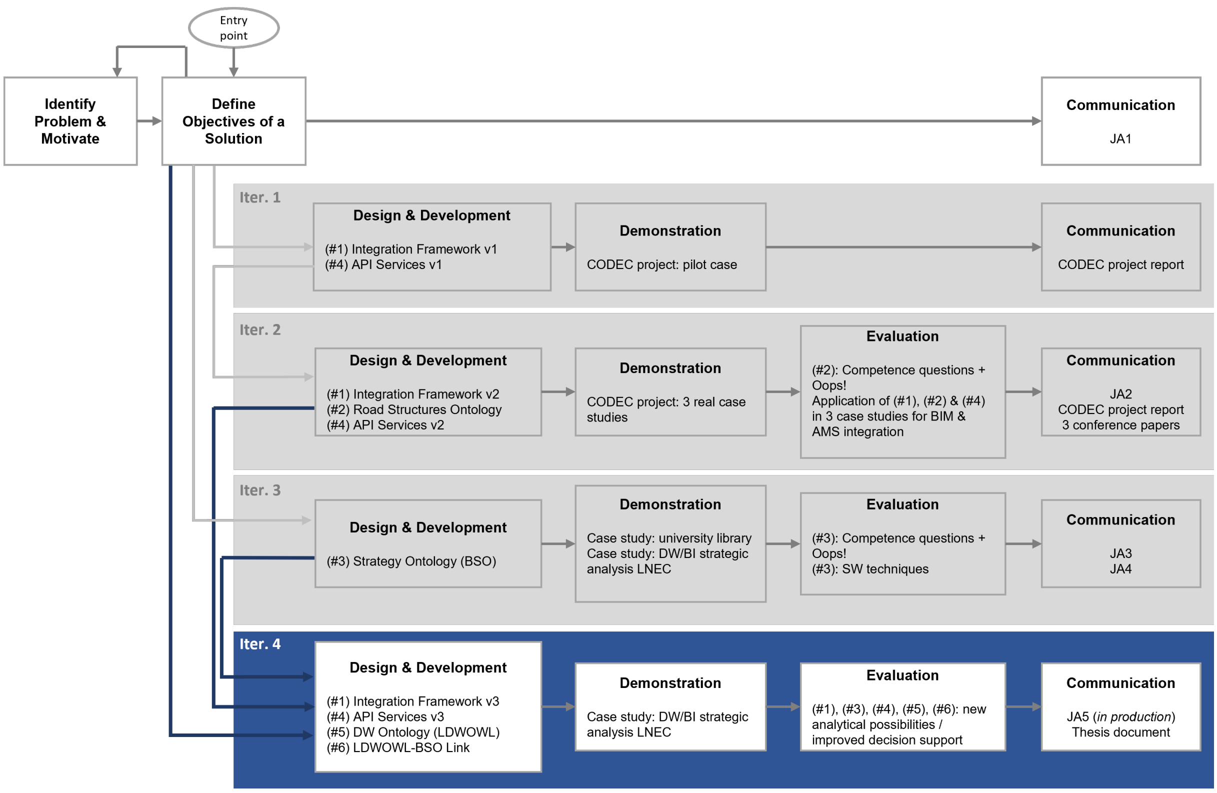


Figure 6.1. DSRM's JA5 Communication.

The Strategy Ontology (artifact #3), named BSO, was presented in Chapter 4. The BSO was designed in an endeavor to bridge the gap between strategy management and data related to the BSC framework. The BSO provides a structured framework to store and analyze knowledge related to the BSC, incorporating information about the strategic components and elements used for evaluating strategy execution. Specifically, the suggested formalization of the BSC framework provides a semantic layer to facilitate the integration, alignment, and traceability of strategic models with organizational information systems, which are essential for supplying data to evaluate the BSC's performance indicators. In today's fast-paced business environment, organizations are often forced to continuously adapt to changes, which may lead to a misalignment between the planned and executed strategies. This reinforces the need and relevance of establishing traceability and monitoring capabilities between strategic models and organizational information systems, by automating strategy analysis and connecting it to data [15]. Existing studies, such as those by Kumar et al. [35] and Tawse and Tabesh [61], also emphasize the importance of combining the BSC with other systems and tools for effective implementation.

The BSO can be seen as an additional semantic layer seamlessly integrated into the Business Intelligence part of the Execution Premium Process [28], which proposes the key steps for effectively implementing a BSC (see Figure 6.2). In a comprehensive BSC implementation, the BSO facilitates the data optimization phase ("Monitor and Learn" and "Test and Adapt" processes), enabling accurate, traceable, and continuous monitoring and improvement of the strategy execution based on a data-driven approach.

Monitoring and analyzing strategy execution requires access to real-time data managed by organizational information systems like the DW/BI system. Automating strategy analysis and connecting it to data is a critical need in today's dynamic business landscape. It promises enhanced efficiency, real-time insights, and the ability to handle the complexity of modern organizations. However, the ontology mapping and population processes related to the BSO, including migrating values related to performance indicators, are currently semi-automated, since knowledge is primarily retrieved from non-structured data sources. Additionally, data access poses a significant challenge in strategy evaluation, as establishing a seamless relationship between this data, containing the necessary information to evaluate performance indicators, is intricate since indicators are defined at different levels of detail. Overcoming these complexities entails extracting values from systems like BI systems and integrating them into the ontology through automated or semi-automated processes, which is vital for ensuring the robustness and applicability of the ontology and methodology in supporting effective strategy execution and decision-making.

6.1. Integration Framework

The proposed solution for aligning the DW/BI system with the organizational strategy is presented in Figure 6.3. The technical architecture presented showcases the new components added to the Integration Framework, leading to its third version. The Integration Framework (version 3) takes advantage of SW technologies to formulate, validate, and

FIGURE 6.2. The Execution Premium Process. Adapted from Kaplan and Norton [28].

ensure the effectiveness of strategies. Furthermore, SW technologies are used to materialize the relationship between strategic goals and DW data, providing a reliable source for data-driven insights necessary to enhance and support strategy management. The framework is comprised of a Semantic Layer and an Abstraction layer, ensured by API Services (version 3 of artifact #4), a DSS (DW/BI system), and BI Applications.

The semantic layer contains a set of ontologies used to represent both the strategic knowledge and the DW structure. The BSO, presented in Chapter 4, contains information about the strategy, formalized using a BSC approach, and provides a structured and machine-readable representation of the organizational goals, indicators, and targets. The Light Data Warehouse Ontology (see Section 6.2) is a new ontological artifact designed and developed to represent the DW's conceptual and logical models. This ontology allows users to define the measures and context of analysis related to organizational processes, and relate these entities to their logical representation, namely, fact tables, dimension tables, facts, and dimensional attributes. Lastly, a link set was defined to relate BSO entities to LDWOWL classes, presented in Section 6.3.

Upon storing the ontologies in a semantic graph database, users can query and update these knowledge bases through SPARQL endpoints. However, effective utilization of these endpoints requires BI users to possess a comprehensive understanding of SW techniques, such as ontologies, their entities and relationships, and the SPARQL query language as a

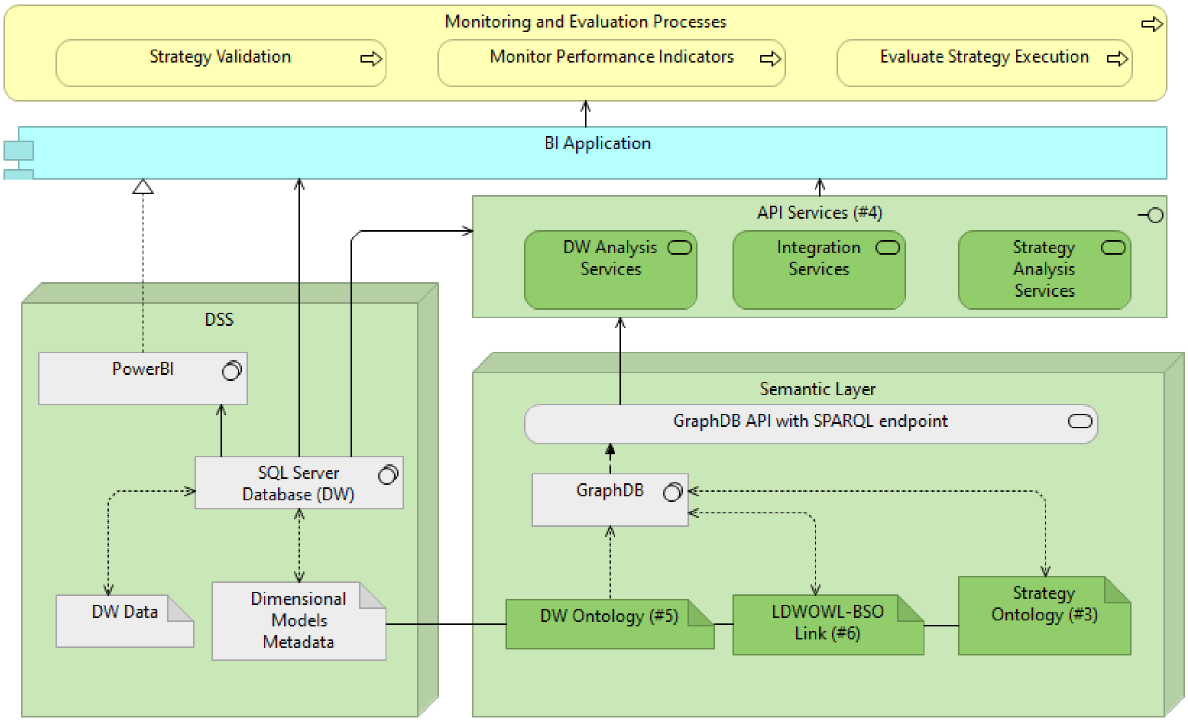


FIGURE 6.3. Integration Framework (version 3 of artifact #1) components used in Iter. 4. Research artifacts are highlighted in darker green.

prerequisite to access this knowledge. The development of abstraction layers, such as the Strategic Analysis Services, DW Analysis Services, and Integration Services APIs, was undertaken to enhance user interaction with ontologies, ensuring accessibility for users as well as external applications. These API Services were intentionally designed to facilitate future expansion and enhancement in response to evolving needs, providing flexibility for various services and use cases associated with the underlying ontologies.

As stated before, the DW is an integrated repository of structured data related to an organization. Typically, DW uses dimensional modeling to store data and provide simplified analytical and decision-support capabilities to business users, through BI applications [32, 31]. This framework enables managers and decision-makers to validate and analyze their strategies based on ontological knowledge and SW technologies supported by DW data. BI applications can benefit from ontology knowledge to provide end-users with strategic information. By using visualizations and recommendations regarding strategies and their executions, these applications can help managers validate the formulated strategies, evaluate performance indicators, and validate cause-and-effect relationships between strategic objectives. This fully integrated solution should allow managers to interact, monitor, analyze, and receive alerts or even analytical recommendations regarding their strategy implementation and execution.

6.2. DW Ontology (LDWOWL)

The DW Ontology (artifact #5), named the Light Data Warehouse Ontology (LDWOWL), was developed as a means to semantically connect DW systems' conceptual and logical

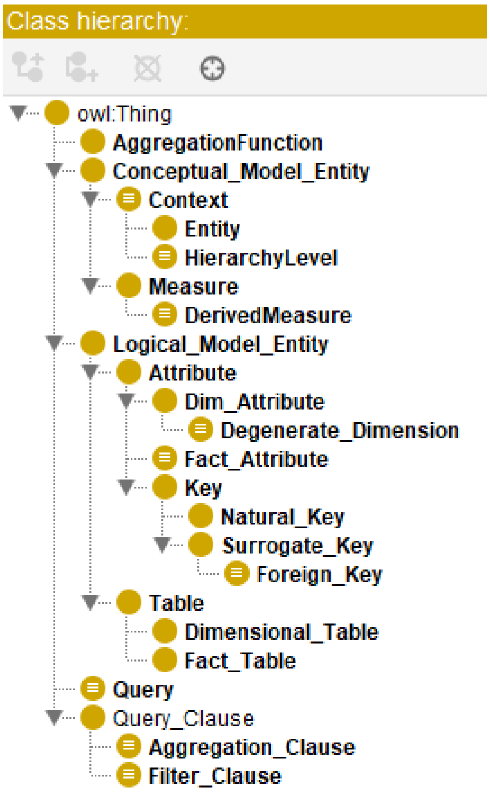


FIGURE 6.4. LDWOWL Class Hierarchy.

metadata to other domains (such as BSO's strategy). The ontology was developed in Protégé to support the DW requirements and star-schema analysis. LDWOWL entities (see Figure 6.4) are used to represent a DW's conceptual model, logical models, and BI typical queries.

As outlined in Chapter 2, ontologies are used during the dimensional modeling process to streamline dimensional design, discover business entities and their relationships, and find potential facts and dimensions from each data source. Consequently, much of the existing literature positions ontologies either as the primary resource for the DW or as an intermediary layer between the source system and the ETL process. However, there is an absence of ontologies capable of conceptually formalizing dimensional models. While ontologies effectively capture business/domain-specific entities and their relationships (e.g., Client, Sale), they overlook the fundamental concepts that form the basis of dimensional models (e.g., Dimension table, Fact table). Recognizing this gap, the LDWOWL was designed and developed to contribute to this particular challenge.

6.2.1. Conceptual Model

Business requirements definition is a pivotal part of the design of a DW/BI system (as proposed by Kimball's DW/BI lifecycle methodology [32]). The elicitation of business concepts is key during this process, where business users and DW/BI experts are tasked with identifying and classifying business entities as either context (dimensions) or measures (facts). Process-driven DW design methodologies, such as BEAM - Business

Event Analysis & Modelling [12], focus on identifying business events within the organization and describing them. Methods, such as the BI Model Canvas, are used to collect event information related to products/services (*What*), time (*When*), places (*Where*), persons/organizations (*Who*), motivations (*Why*), or process information (*How*). In addition to the event context, measures (*How many*) are also identified, allowing event performance to be measured. The resulting conceptual model can now be stored in the LDWOWL (see Figure 6.4).

As stated before, in a DW/BI system, concepts can be either measures or context. Measures are indicators, such as quantities, currency, or amounts of time, measured during the execution of events. Since the individual value of each transaction is not (usually) of analytical interest, these values should be aggregated into measures. The same value can be used for multiple measures, depending on the aggregation function used. For example, in a sales context, Total Sales Value and Average Sales Value are two different conceptual measures that are obtained by applying different aggregation functions (in this case, sum and average) to the same measure (Sales Value). Measures can also be derived to obtain Derived Measures (e.g., ratios), and can be classified according to their additivity (additive, semi-additive, or non-additive).

Measures are aggregated according to a given context. Users can use the "hasContext" object property from LDWOWL to evaluate the analytical context of each measure, which can be further specified following BEAM's 7W's (e.g., the "*When*" information can be associated with a measure using the "hasTemporalContext" object property). Furthermore, context typically exhibits well-defined hierarchical relationships, which are used within DW/BI systems to construct aggregate or derived models and guide model exploration within BI applications. In LDWOWL, these top-down hierarchies between contexts are defined using "bottomHierarchyLevelFor" and "topHierarchyLevelFor" (e.g., $< Date >$ bottomHierarchyLevelFor $< Month >$), also ensuring the conformity between these. Moreover, LDWOWL can take advantage of other Semantic Web standards, such as SWRL[1], to add inference rules such as the one presented below:

$$\text{Measure}(m) \land \text{hasContext}(m, c1) \land \text{bottomHierarchyLevelFor}(c1, c2) \rightarrow \text{hasContext}(m, c2)$$

This rule allows the ontology to infer relationships between measures and hierarchy levels based on the asserted relationships between measures and base context (or lower hierarchy levels). Base context, such as Date, Location, Client or Collaborator is stored as "Entity". However, an Entity can be used as a Hierarchical level of one or more entities (e.g., Location can be used to aggregate clients or collaborators).

[1]https://www.w3.org/submissions/SWRL/

6.2.2. Logical Model

The logical model of a star schema represents how the data will be stored and organized within the DW. The metadata of these models can be stored within LDWOWL, allowing them to be associated with and validated against the conceptual model.

Dimension Tables represent context information about analytical-relevant business entities (e.g., Date, Client). Entities with an explicit hierarchical relationship are usually represented within the same table (e.g., Product, Model, and Brand are represented within the Product dimension table). In addition to the dimensional attributes (used to aggregate and filter facts), dimensional tables have a surrogate key (primary key) and can contain several natural keys (which unequivocally identify entities in source systems). Facts are stored in the center of the star schema, called the Fact Table. Facts are typically numeric values representing measurable and quantitative data. The fact table also contains foreign keys for each of the associated dimensional tables, and degenerated dimensions (context information stored in the fact table, such as transaction identifiers).

Concepts must be related to an attribute using the "hasLogicalAttribute" object property, ensuring that they are represented in the DW. Measures are related to fact attributes from the logical model[2], while context information is related to dimensional attributes (usually stored in dimensional tables). Furthermore, a default aggregation attribute can be defined for each context concept (e.g., "Stock Keeping Unit (SKU)" for Product, "Brand name" or "Brand Code" for Brand), allowing users to aggregate by a certain context without specifying a desired attribute.

6.2.3. BI Queries

A Query concept can also be stored within the LDWOWL, following the typical BI query SQL pattern. A BI Query focuses on a measure ("queriesMeasure"), which is typically aggregated using an SQL aggregation function (e.g., "SUM", "COUNT", "AVG"). The query defines how this measure should be:

(a) aggregated over some context ("aggregateBYContext"), leading to a GROUP BY clause in SQL;
(b) filtered for some context ("filterFORContext"), leading to a WHERE clause in SQL.

The ontology defines how a certain measure is typically analyzed, relating the measure with a set of context entities (at any hierarchical level), and providing knowledge related to the representation of these concepts in the DW's logical model. Filter clauses allow the definition of filter values, allowing the storage of filter information (such as "equal to" or "greater than" a certain value) in a formal way (as shown in Figure 6.5). The query is related to a measure ("Total Expenditure"), which is aggregated by a context ("Month")

[2]If avoidable, non-addictive measures (such as ratios) should not be represented in a fact table. The fact table should, instead, store the addictive measures that are used to calculate these derived measures.

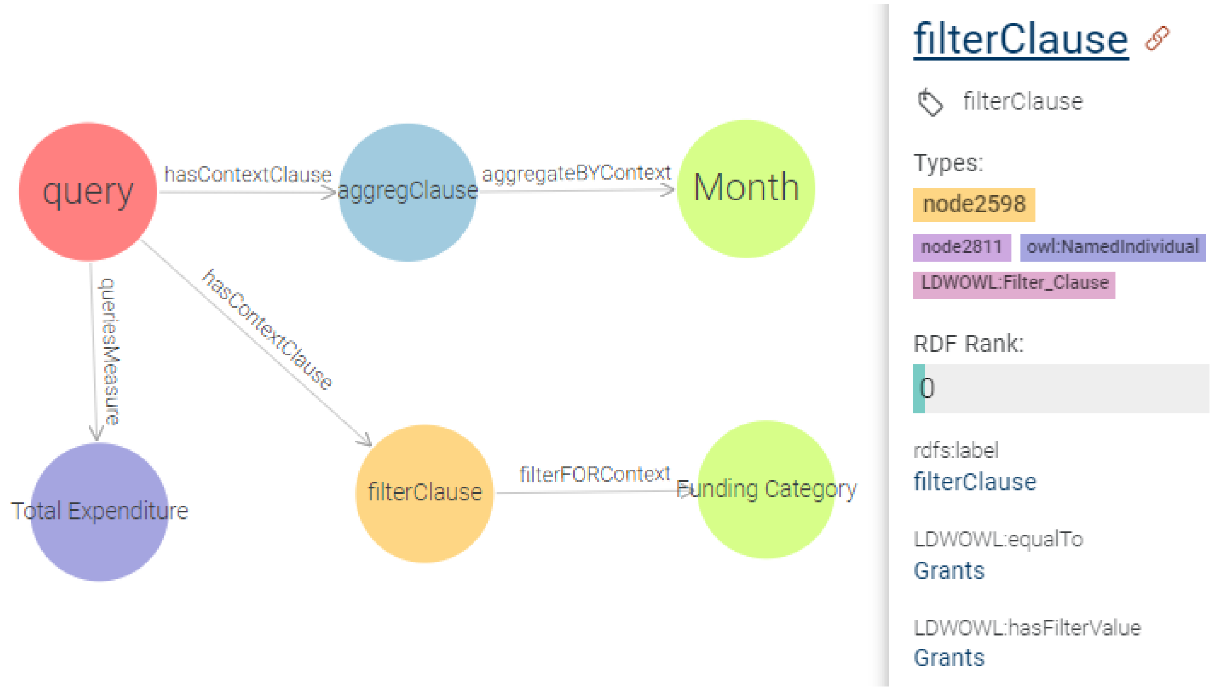


FIGURE 6.5. LDWOWL Query Example (GraphDB visualization).

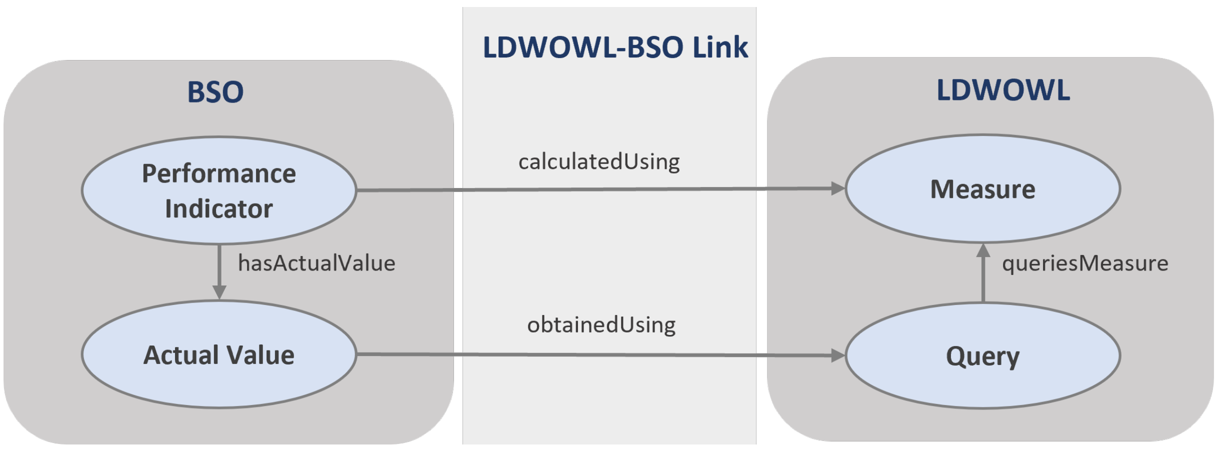


FIGURE 6.6. LDWOWL - BSO Link.

and filtered by a second context ("Funding Category"). The filter is defined using the "equalTo" data property, which, in this case, filters the measure for "Grants".

6.3. LDWOWL - BSO Link

As previously mentioned, the LDWOWL was developed as a means to semantically connect the DW to other domains, namely with the BSO's strategy. The LDWOWL - BSO link set (artifact #6) was developed to link these two ontologies, defining the object properties needed to relate BSO entities to LDWOWL entities, effectively linking strategy to data, as shown in Figure 6.6.

The LDWOWL's Measure and Query entities serve as anchor points between the BSO and the DW information. The "calculatedUsing" object property relates performance indicators to conceptual measures. These measures are represented in the DW through facts (which can be directly available in transaction fact tables or be made available by designing derived tables, such as periodic snapshots). This measure can then be analyzed according to a certain context defined in the conceptual model.

An actual value related to the execution of a performance indicator can now be connected to an LDWOWL Query by using the "obtained using" object relationship. The query defines the context needed to aggregate and filter a measure, allowing it to explicitly define how a certain value for a performance indicator should be calculated, regardless of its detail level. This enables users to formally detail and retrieve the necessary data to evaluate their strategy, without the knowledge and complexity of the logical model.

6.4. API Services

Once stored in a semantic graph database, such as GraphDB, ontologies can be queried or updated using SPARQL endpoints, allowing the BSO and LDWOWL ontologies to be accessed by external applications. However, access to the SPARQL endpoint requires BI users to have a good understanding of the ontology (its entities, relationships, etc.) and the query language used to retrieve information, namely SPARQL. Both the BSO and the LDWOWL REST APIs were developed to provide an abstraction layer for users to interact with the ontologies. The new API Services (version 3 of artifact #4) were developed in Phyton using FastAPI[3] and SPARQLWrapper[4]. These services do not encompass an exhaustive list of services or use cases associated with the underlying ontologies. They have been intentionally developed to allow them to be expanded and improved as necessary.

6.4.1. BSO API

The BSO services were based on the competency questions initially defined in its Ontology Requirements Specification Document, presented in Chapter 4 and use cases developed for strategy analysis, namely those presented in Chapter 5. Table 6.1 presents a summary of currently available API services related to the BSO.

6.4.2. LDWOWL API

Similarly to the BSO API, a set of services was developed to access and update knowledge related to the LDWOWL. These services include obtaining the measure context, the aggregation (hierarchical) paths for context and the available attributes for a given context. Furthermore, the API also allows the retrieval of existing queries, query context (aggregation and filter context), and the default attributes necessary to run the query (and ensure that they are connected in the logical model, i.e., that the fact table with the necessary fact is related to the context's dimensions). Moreover, the service allows the transformation of typical BI queries to their respective SQL queries, which then allows the DW to be queried and data to be retrieved. However, this method is currently limited, only working with single fact queries and assuming filter clauses to always be in conjunction.

[3]https://fastapi.tiangolo.com/
[4]https://sparqlwrapper.readthedocs.io/

Table 6.1. BSO API Services.

Services	Description
BSO - Balanced Scorecard	Services related to the BSC class. Allows to obtain a list of existing BSCs from the ontology, information related to them, such as strategy elements (Mission, values and vision), time-horizon and organizational level, and their components (strategy map and quantification frameworks);
BSO - Strategy Map	Services related to the strategy map. Contains endpoints to get information concerning strategic objectives, perspectives (and their order) and strategic themes pertaining to a strategy map;
BSO - Strategic Objective	These services allow the individual analysis of strategic objectives. Given a strategic objective, information related to the objective perspective, strategy themes and performance indicators can be obtained. A list of objectives influenced directly or indirectly by cause-and-effect relationships can also be retrieved;
BSO - Performance Indicator	Services related to a performance indicator, namely the type (lead or lag), properties (e.g., data acquisition frequency), and defined targets;
BSO - Strategy Execution	Services related to actual values of performance indicators (including the addition of new values) and initiatives
BSO - Cascading	Services related to BSC cascading (how BSCs and strategic objectives are related across organizational levels)
BSO - Strategy Analysis	Currently, two different services are provided, related to use cases developed for LNEC's strategy analysis. First, given a quantification framework, a complete overview is provided, detailing strategy objectives, performance indicators, related targets, and actual values, allowing users to evaluate the strategy performance. Second, a comparison between frameworks, which is useful for validating cause-and-effect relationships, as shown in Chapter 5.

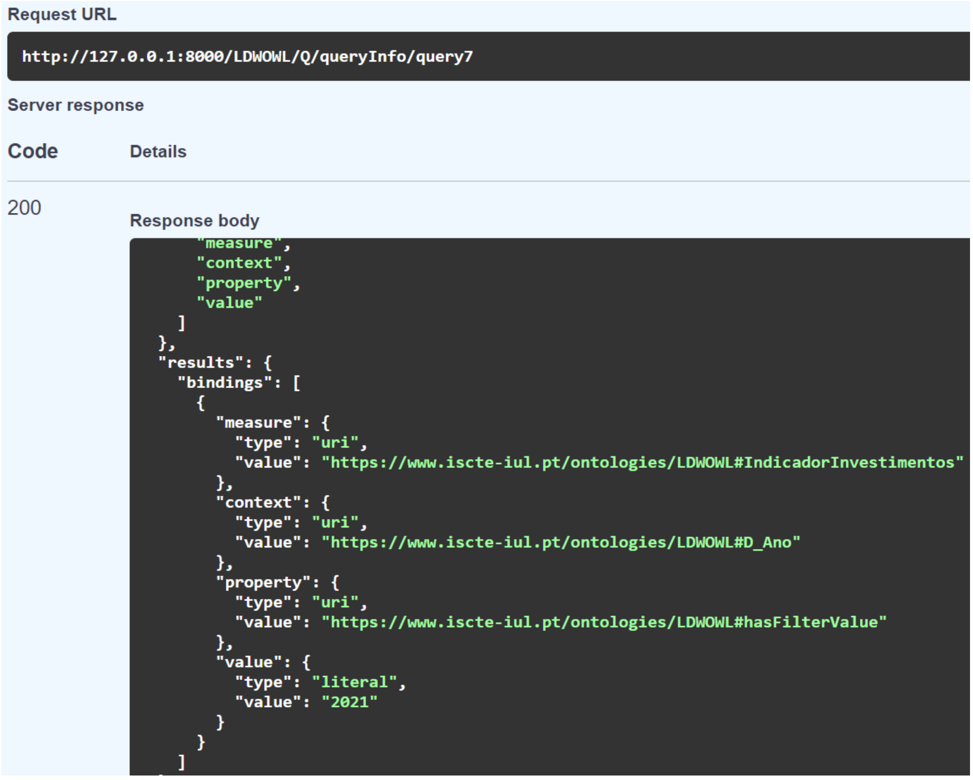
Figure 6.7 presents the query information, including measure and context information, resulting from a GET service from this API, shown in Swagger[5]. The specific query uses the yearly context ("D_Ano") to filter the investment indicator ("IndicadorInvestimentos", in Portuguese) for the year 2021. Another service can be used to obtain the SQL query text, based on the information from the previous query information, as illustrated in Figure 6.8. The investment indicator for 2021 is calculated by dividing two facts from a derived table, filtered by the month dimension for the year in question. The logical model stored within the LDWOWL is used to fill in information regarding tables and surrogate keys, as well as the necessary fact and dimensional attributes.

6.4.3. Integration API

A set of services was developed to take advantage of the LDWOWL—BSO link and showcase the integration between the DW/BI and the BSO's strategy. These services are designed to validate this integration and enable effective strategy execution and decision-making by automating the ontology population process.

The first set of services enabled by the LDWOWL - BSO link concerns the validation of strategy/data alignment. By taking advantage of this integration, users can validate if all of the performance indicators are connected to conceptual measures or if the queries

[5]https://swagger.io/




```
      "measure",
      "context",
      "property",
      "value"
    ]
  },
  "results": {
    "bindings": [
      {
        "measure": {
          "type": "uri",
          "value": "https://www.iscte-iul.pt/ontologies/LDWOWL#IndicadorInvestimentos"
        },
        "context": {
          "type": "uri",
          "value": "https://www.iscte-iul.pt/ontologies/LDWOWL#D_Ano"
        },
        "property": {
          "type": "uri",
          "value": "https://www.iscte-iul.pt/ontologies/LDWOWL#hasFilterValue"
        },
        "value": {
          "type": "literal",
          "value": "2021"
        }
      }
    ]
```


FIGURE 6.7. Query Measure and Context Information from LDWOWL API.


```
{
  "head": {
    "vars": [
      "queryText"
    ]
  },
  "results": {
    "bindings": [
      {
        "queryText": {
          "type": "literal",
          "value": "SELECT (SUM(factTable.despesaInv)/SUM(factTable.despesaTotal) ) as indInv\nFROM DM.derived_table as factTable, DM.d_mes as D_MES\nWHERE factTable.ID_MES = D_MES.ID_MES and D_MES.ANO = 2021"
        }
```


FIGURE 6.8. Resulting SQL Query from LDWOWL API.

for retrieval of their actual values are formalized in the ontology. Furthermore, since the conceptual and logical models are related in the LDWOWL, the query for each value can itself be validated, i.e., if all attributes needed relating to the fact and analysis context exist in the DW and if their tables are related through surrogate keys.

The second set of services showcases how this solution can be used to enable the automatic extraction of the values of performance indicators from the DW and populate them into the ontology.

Given a BSO performance indicator, BI queries are obtained from the LDWOWL for any missing actual value related to that performance indicator. These BI queries are then transformed into SQL queries and used to retrieve data from the DW, which, in turn, is used to populate the BSO. A similar service can also be used to validate existing actual values already in the ontology. As shown before, the BSO API can provide strategy analysis services that take advantage of this new information retrieved from the DW. By using DW/BI system data, this approach enables reliable and data-driven strategic decision support for users and managers.

6.5. Demonstration and Evaluation

The Integration Framework (version 3), as proposed in Section 6.1, aims at aligning the DW/BI system with the organizational strategy, by taking advantage of SW techniques, such as ontologies. The DW/BI Strategy Analysis in a Pubic Organization case study, presented in Chapter 5, was used to demonstrate the integration framework, which allows the retrieval of performance indicators' actual values from the DW into the BSO, due to the semantic integration between the LDWOWL and BSO. The application of this framework in our research is supported by two key prerequisites, both of which have already been met:

(1) The organization has already developed a Data Warehouse. The framework operates on the premise that the DW was designed following a dimensional modeling approach and contains all the necessary data for evaluating organizational performance.
(2) The organizational strategy was formulated using a BSC and is formalized using the BSO. As stated before, the case study's strategy was already populated into the BSO, which enabled a set of semantic-based analyses related to the validation of the strategy formulation and evaluation of organization performance, as shown in Chapter 5.

In the previous work, data related to the evaluation of performance indicators was imported from non-structured data, namely the annual reports published by the public entity, involving a significant level of human intervention in the data extraction and ontology population process. Three significant problems arise from the existing approach: (1) The monitoring frequency of these indicators may not provide managers and decision-makers with timely insights to adapt to evolving business conditions; (2) The availability of data is limited to mandatory reported indicators, neglecting other important performance indicators that are currently not being evaluated by this solution; and, (3) Manual intervention is an error-prone process that can introduce data problems, impacting the quality of the decision-making process.

To demonstrate the integration framework, data regarding the evaluation of Ind. 2 - Modernization and Valorization Project Expenditure Ratio (%) was automatically retrieved from the DW and populated into the BSO knowledge base containing LNEC's strategy. As indicated in Chapter 5, the data on indicators whose reporting is not mandatory, such as Ind. 2, was not available. The actual values of the performance indicators are retrieved via the Integration API service, which identifies the necessary BI queries related to that indicator, obtains the actual values from the DW and fills in the ontology with those values. This service was used to obtain data relating to a performance indicator with missing monitoring information. Among other capabilities, this solution can be used to automatically validate existing ontological knowledge with data from the DW and to obtain performance evaluation data more frequently.

Once this information is contained in the ontology, external applications, such as BI applications, can take advantage of the existing APIs to create reports and visualizations and execute data-driven analyses related to organizational strategy. Figure 6.9 shows a simple visualization regarding the evaluation of two strategic objectives from LNEC's strategy, together with their respective performance indicators, including Ind. 1 and Ind. 2 for the strategic objective OO.1 - Enhance and Modernize LNEC, and Ind. 5 for strategic objective OO.2 - Qualifying the HR[6]. Furthermore, the visualization also presents targets and trend lines for each performance indicator. With sufficient data from the DW, this approach can be used to obtain data on all the case study's performance indicators, enabling LNEC's strategy to be fully evaluated and analyzed.

However, it's important to note that in Figure 6.9, Ind.2 lacks defined targets as they were not populated into BSO during the research described in Chapter 5. The strategy evaluation process is hindered without the correct formalization of the strategy and its performance indicators, which can be ensured by the BSO. Managers and decision-makers can leverage these formalisms to enhance their analysis and exploration by providing additional context and knowledge within the BI environment. When integrated with real-time data managed by organizational IS, strategic information, including targets, initiatives, and cause-and-effect relationships of strategic objectives, plays a pivotal role in supporting data-driven decision-making, ensuring a reliable and effective strategy execution.

6.6. Discussion

This chapter explores the application of SW techniques in DW/BI systems for strategy analysis and management. In previous research, the BSO (presented in Chapter 4) was used to validate and infer over strategic knowledge of a public organization. Chapter 5 explored how organizations can employ these techniques to assess their strategy formulation and execution. The BSO was used to evaluate performance indicators and validate the cause-and-effect relationships between strategic objectives while ensuring a more reliable and accurate representation of the BSC framework (i.e., ensuring that the organization's strategy formulation is semantically, structurally, and data-wise compliant with the BSO model).

Ontology-based strategy evaluation and analysis can be hindered by a lack of data, as shown in Chapter 5. Access to actual data managed by organizational systems, such as DW/BI systems, is essential to provide a reliable base for strategy evaluation and analysis.

[6]More information related to these performance indicators and strategic objectives can be found in Chapter 5, Section 7 - Strategy Analysis.

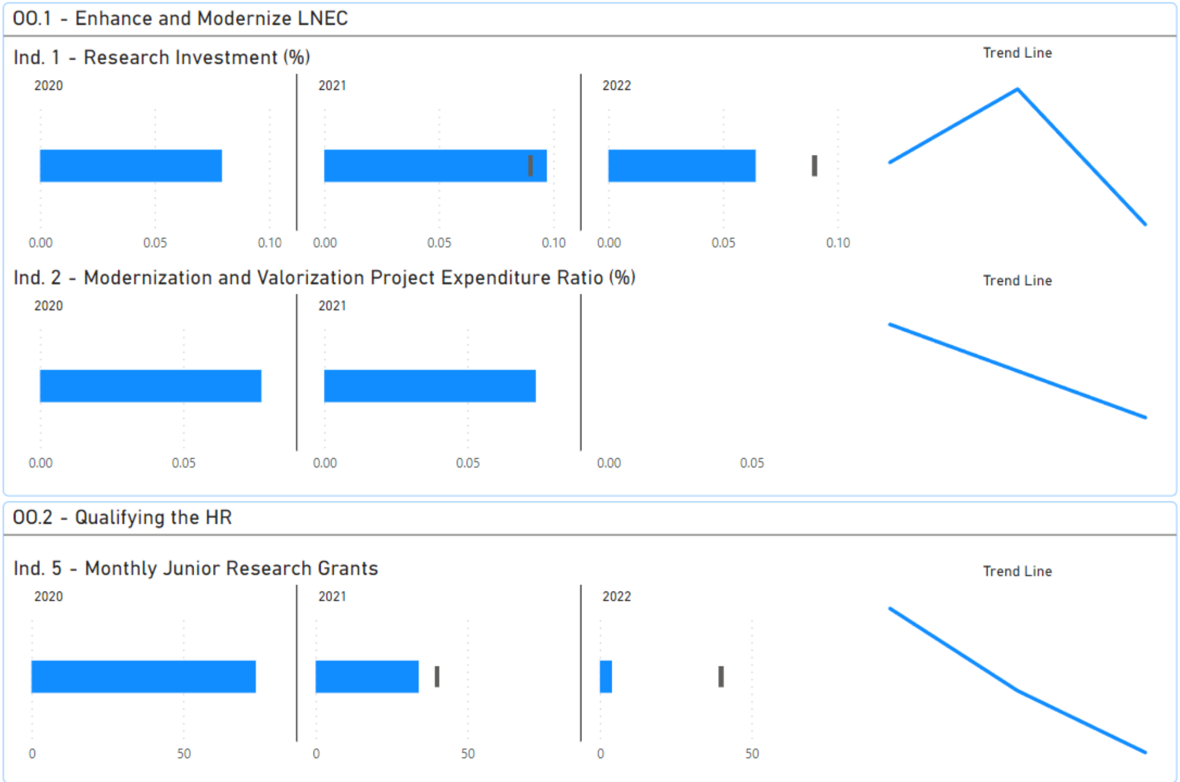


FIGURE 6.9. PowerBI Visualization: OO.1 and OO.2 Evaluation.

However, establishing a seamless relationship between this data, encompassing values collected for each performance indicator, and their ontological representation (in the BSO) is an intricate and complex challenge to which this research contributes. The third version of the Integration Framework (artifact #1) is proposed for aligning a DW/BI system with the organization strategy (see Section 6.1), by using SW technologies to materialize the relationship between strategic entities and DW data. By establishing this relationship, the process of extracting values from IS and integrating them into the ontology can now be automated, providing a reliable source for data-driven insights necessary to enhance and support strategy management. By bridging the gap between data and strategy, organizations can enhance their ability to monitor and analyze strategy execution accurately, thus facilitating more informed decision-making processes.

6.6.1. Contributions to the practice

The alignment of IS, namely DW/BI systems, with strategic knowledge can augment previously identified advantages that the BSO and SW technologies offer to managers and decision-makers. First, formulating a strategy, particularly the cause-and-effect relationships from the BSC framework, often depends on managers' insights, knowledge, and intuition. The proposed framework facilitates the integration of performance data into the BSO, enabling the monitoring and empirical validation of these relationships. This proactive approach prevents decisions that may not contribute to the organization's long-term success, preventing the misallocation of resources to the fulfillment of non-essential or redundant strategic indicators and objectives. Furthermore, by using actual data related to strategic execution, the proposed solution can enable dependency analysis processes, providing insights into how changes in a specific indicator or objective may impact the overall strategy.

Second, the use of ontologies, such as BSO and the LDWOWL, provides increased interoperability, enabling the sharing of information and data across systems and organizations. This interoperability is key in an integrated solution such as the one proposed in this research, making the collection, analysis, and reporting of performance indicators more efficient. The automation of data integration from organizational IS is crucial for supporting real-time or near-real-time monitoring of performance indicators and data-driven decision support.

Finally, industry, government, and university organizations around the world often need to adhere to regulatory requirements and comply with relevant standards, such as those established by policymakers (e.g., the European Commission). SW technologies can play a crucial role in supporting compliance and governance initiatives by enhancing external communication and alignment, by leveraging their semantic interoperability to streamline documentation and reporting. For example, public organizations in Portugal, such as LNEC, are subject to a mandatory framework for assessing and monitoring the performance of Portuguese public services. The proposed solution can facilitate the automatic evaluation and reporting of indicators for compliance frameworks. Furthermore, due to the semantic integration between the BSO and LDWOWL (see Section 6.3), managers can define BI queries to answer each performance indicator, enabling the report and analysis of all performance indicators related to the organizational strategy. By using LDWOWL's DW conceptual model, the proposed integration framework facilitates the formulation of BI queries, ensuring and validating that each performance indicator evaluation can be evaluated.

CHAPTER 7

Conclusions

This final chapter presents the conclusions of this doctoral thesis. It summarizes this research's main results and artifacts, highlighting their innovative aspects and potential impact on the IS research field. Additionally, it discusses how the research questions were addressed. Finally, an overview of the limitations and challenges encountered during this work is presented, identifying future research directions for academics and practitioners.

7.1. Research Summary

This research started with the primary goal of exploring the use and integration of SW techniques, such as ontologies, with DSS, namely DW/BI systems, to enhance the decision-making process. Due to their semantic formalization, and inference qualities, ontologies are used in IS to cope with the growing need for sharing and reusing data and knowledge in various research areas. The integration of these knowledge-based artifacts into DSS can provide new sources of information, enable new analytical possibilities, and facilitate the decision-making process.

Chapter 2 presented a systematic literature review aiming at exploring the incorporation and impact of ontologies in the DW/BI system, analyzing how ontologies can be used to enrich and facilitate the design, development, and operation of the DW/BI system. An identified unexplored research gap entails leveraging ontologies containing domain-specific knowledge to extend the set of available data, enriching and enhancing the exploration phase in BI applications, taking advantage of the ontologies' interoperability to improve or facilitate the decision-making process.

To address this gap, a set of research artifacts was developed throughout this doctoral thesis to represent and explore domain-specific concepts (related to real-world case studies) and take advantage of this knowledge in a DSS, providing new analytical possibilities. Following the DSRM process, presented in Figure 1.4, this doctoral thesis entails four design and development iterations to address the research objectives outlined for this work.

The first contribution of this research is the development of a framework to support, explore, and validate domain-specific information, such as strategy, in BI environments. This framework, called Integration Framework (artifact #1), was designed to delineate the essential components and requirements to integrate ontological knowledge with DSS, enabling ontology-supported analysis and exploration in DSS.

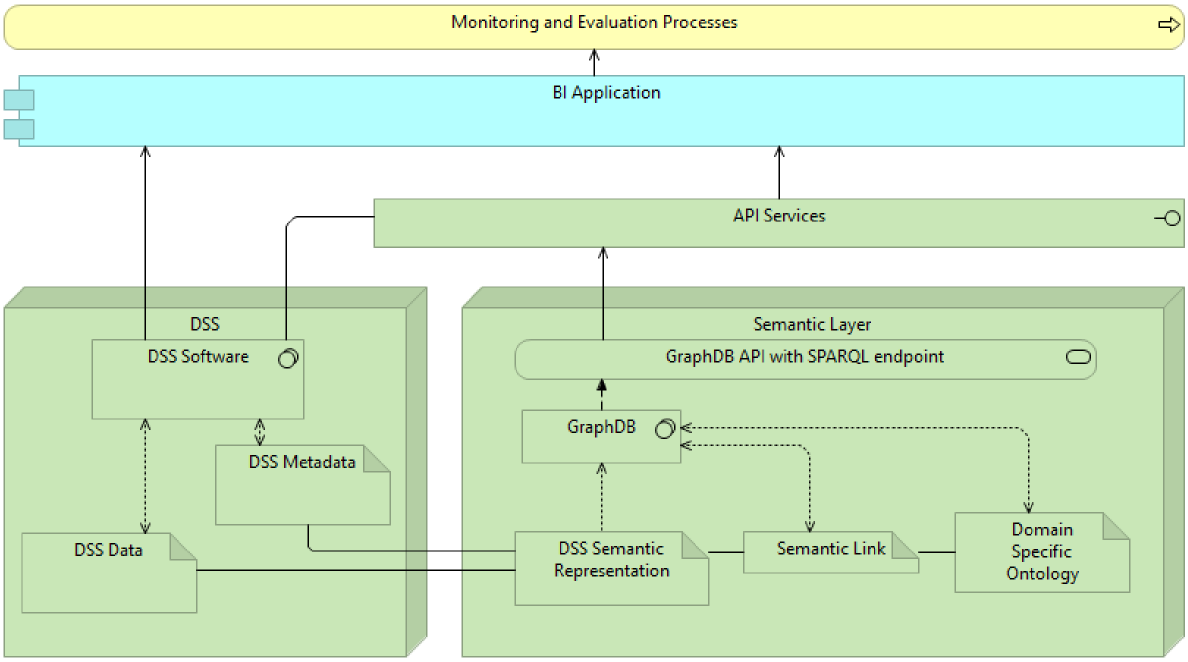


FIGURE 7.1. Integration Framework High-level Architecture.

Figure 7.1 describes the high-level technical architecture of this framework using ArchiMate[1], which was initially conceptualized in Section 1.4.3 in Figure 1.5. The Integration Framework comprises the DSS, the Semantic Layer, the Abstraction layer (API Services), and the BI application used to support the monitoring and evaluation processes. The DSS component is a traditional DSS system, providing BI Applications as their exploration environment. Data is integrated in the Semantic Layer, where domain-specific knowledge is formalized within an ontology and linked to the DSS semantic representation through a semantic link, enabling the enrichment of the DSS information. The DSS component depends on the case study, while the Semantic Layer is composed by the ontologies and semantic links. The Semantic Layer uses GraphDB as the ontology repository, enabling a SPARQL endpoint (GraphDB API). The API Services provides an integrated access point to exploit the Semantic Layer. The API Services can be accessed by any kind of application, as, in this case, the BI applications.

The Integration Framework (artifact #1) was designed and developed throughout this doctoral thesis, being initially proposed during Iter. 1 (version 1), and incrementally developed during Iter. 2 (version 2) and Iter. 4 (version 3). A set of new components was added in each version of the artifact to support monitoring and evaluation processes in the different application scenarios.

The first DSRM iteration (Iter. 1) entailed the development of Integration Framework (version 1 of artifact #1) and API Services (version 1 of artifact #4). Figure 7.2 presents the architecture of the first version of the Integration Framework. The existing DSS component provides a 3D visualization of an IFC Road Model, while the Semantic Layer

[1]https://www.opengroup.org/archimate-forum/archimate-overview

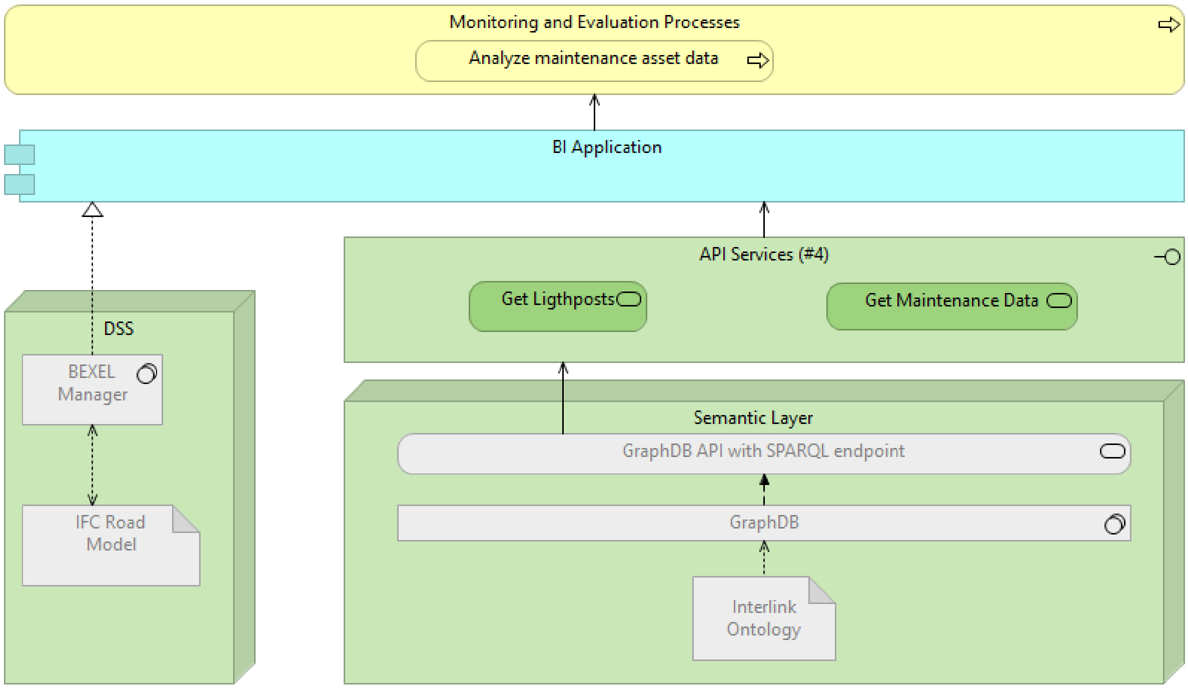


FIGURE 7.2. Integration Framework version 1.

uses an existing ontology[2] (Interlink ontology) from a previous research project, stored in GraphDB. The developed API Services (version 1) use the GraphDB API's SPARQL endpoint to retrieve the maintenance asset data from the ontology, enabling access within the 3D visualization BI environment. These artifacts were used and demonstrated as a proof-of-concept to assess the feasibility of using ontological knowledge within a DSS environment, providing asset managers with maintenance information for light posts.

Following the demonstration of the proof-of-concept, Iter. 2 starts with new design and development phases for the Integration Framework (version 2 of artifact #1) and the corresponding API Services (version 2 of artifact #4). The Integration Framework, detailed in Figure 7.3, was demonstrated and evaluated in the context of the CoDEC project (see Section 1.3.1) to link operational data with BIM environments, facilitating the decision-making process in the European Highways industry. The existing DSS uses 3D BIM models (represented in IFC) to provide users with information regarding road structures, usually supporting decisions in the early phases of these structures' life-cycle. Typically, asset management data required to support decision-making in the operational & management phase is managed by AMS that do not interoperate with BIM. By taking advantage of SW techniques, the Integration Framework was used to deliver and integrate this operational data with BIM, allowing users to access operational data in a familiar BIM-based decision support environment.

A domain-specific ontology was used as the base for data exchange between BIM and AMS in the semantic layer of the Integration Framework. The Road Structures Ontology (artifact #2), named Engineering Structures Ontology in Chapter 3 (JA2), was designed

[2]https://wpd1.s3.eu-central-1.amazonaws.com/index.html

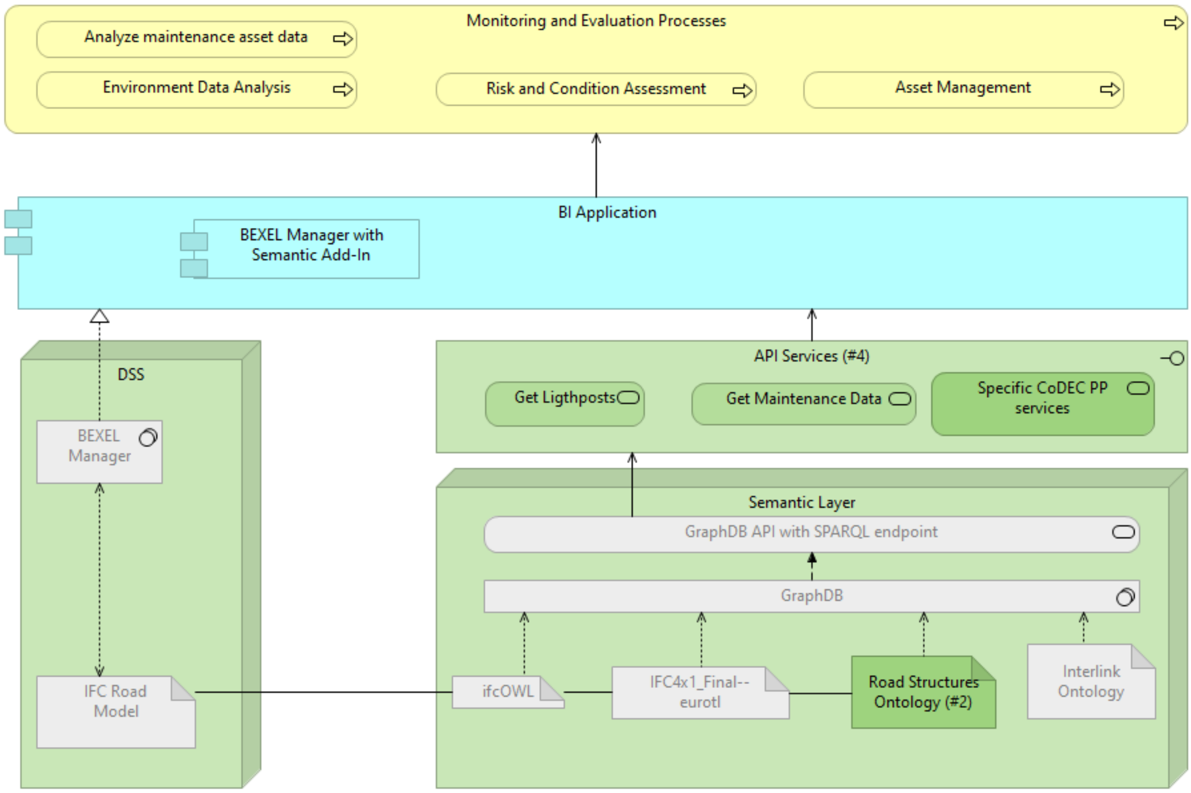


FIGURE 7.3. Integration Framework version 2.

to model and represent structures, structural elements, and operational data, such as sensor information and pavement properties, based on a shared conceptualization provided by several European NRAs. The ontology allows data to be encoded in a homogeneous way, enabling semantic and data interoperability between the NRAs and their systems. The integration between the domain-specific ontology and the DSS is ensured using an existing semantic representation of the DSS (ifcOWL[3]), with the necessary semantic link (IFC4x1_Final–eurotl[4]), developed within the context of the Interlink project. Specific services, such as retrieval of inspection data, were added to the API Services (version 2 of artifact #4) to provide access to this integrated knowledge from the Semantic Layer and support the CoDEC pilot projects through the integration of AMS into BIM.

The third research iteration (Iter. 3) comprised the design and development of the Strategy Ontology (artifact #3), with the objective of representing and exploring the domain-specific concepts related to organizational strategy. This artifact was introduced in Chapter 4 as the Balanced Scorecard Ontology (BSO). The BSO provides a structured framework to store and analyze knowledge related to the BSC, providing a semantic layer to facilitate the integration, alignment, and traceability of strategic models with organizational information systems. This artifact was demonstrated and evaluated using a public university library case study, showing that the BSO is able to formalize BSC

[3]https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl/
[4]https://wpd1.s3.eu-central-1.amazonaws.com/IFC4x1_Final_doc/index-en.html

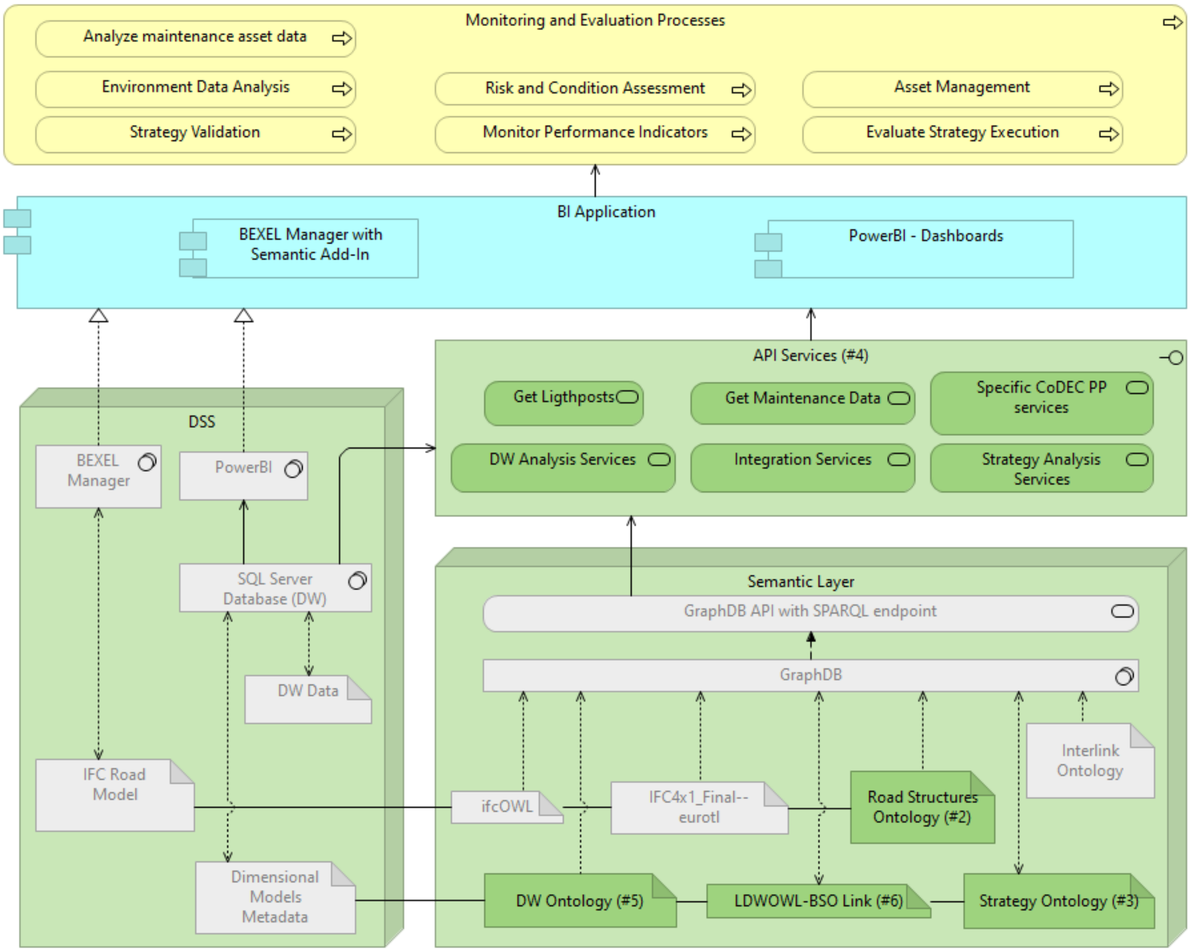


FIGURE 7.4. Integration Framework version 3.

knowledge, validate BSC elements and relationships, and infer new knowledge related to them.

The practical effectiveness and applicability of the BSO were further demonstrated and evaluated in the DW/BI Strategy Analysis in a Public Organization case study (see Section 1.3.2). Chapter 5 explored how public organizations, such as LNEC, can leverage ontologies to support and validate their strategy management processes. The BSO is used, together with a set of SW techniques, namely SPARQL, SHACL, and SWRL, to validate the strategy formulation, support the evaluation of performance indicators, and verify the established cause-and-effect relationships between strategic objectives.

The work developed during Iter. 3, presented in Chapters 4 and 5, was crucial in exploring and showcasing the main benefits of using SW techniques for domain-specific analysis, namely strategic validation and analysis. BI applications were used during the research to provide users with information and visualizations regarding their strategy and its execution (translated into performance indicators). However, without a formalized connection to organizational systems, the approach had clear limitations (e.g., error-prone process, low data acquisition frequency), mostly caused by the heavy manual intervention in the ontology population process.

TABLE 7.1. Evolution of the Integration Framework artifact design.

	Version 1	Version 2	Version 3
BI Application	BEXEL Manager	BEXEL Manager with Semantic Add-In	PowerBI - Dashboards
API Services	Get Light-posts and Get Maintenance Data	Specific CoDEC PP services	DW, Strategy Analysis, and Integration Services
DSS Repository	IFC Model	IFC Model	DW
DSS Semantic Representation	N/A	ifcOWL	DW Ontology (#5)
Semantic Link	N/A	IFC4x1_Final–eurotl	LDWOWL - BSO Link (#6)
Domain-Specific Ontology	Interlink Ontology	Road Structures Ontology (#2)	Strategy Ontology (#3)
Decision Support Processes	Analyze Maintenance Data	Environment Data Analysis Risk and Condition Assessment Asset Management	Strategy Validation Monitor Performance Indicators Evaluate Strategy Execution
Demonstration Case Study	CoDEC Pilot Case	CoDEC Pilot Projects	DW/BI Strategy Analysis in a Public Organization

To overcome the above-mentioned limitation, the fourth and final DSRM iteration of this doctoral thesis (Iter. 4) focused on the design and development of a new version of the Integration Framework (version 3) aimed at enabling the integration of strategic knowledge with DW/BI system data. Figure 7.4 illustrates the final version of the Integration Framework (version 3), which includes all the components designed and developed throughout this research (highlighted in the figure). Chapter 6 described the refinement of the Integration Framework and the API services (version 3 of artifact #4), and the design and development of two new artifacts, namely the DW-Ontology (LDWOWL) (artifact #5) and the LDWOWL - BSO Link (artifact #6). These artifacts were used as components of the Integration Framework and serve as the DSS semantic representation and semantic link, respectively. By establishing this relationship, the process of extracting values from the DW/BI system and integrating them into the domain-specific ontology can now be automated. The demonstration and evaluation phases of Iter. 4 were based on the DW/BI Strategy Analysis in a Public Organization case study (see Section 1.3.2). By integrating strategic knowledge (formalized with the Strategy Ontology, artifact #3) with DW/BI system data, the Integration Framework enables more informed and reliable data-driven decision-making processes, improving the organization's ability to monitor and analyze its strategy.

Table 7.1 summarizes the incremental evolution of the Integration Framework (artifact #1). While there is no direct relation between the different application scenarios, there are no perceived barriers to the semantic integration between domain-specific ontologies and other components.

7.2. Discussion and Contributions

This doctoral thesis aimed to advance the use of SW techniques in DSS, particularly focusing on integrating ontologies into DW/BI systems. Integrating ontologies into DSS promises to enhance decision-making and reveal new insights from organizational data. Through SW technologies, organizations can enrich BI exploration, improve interoperability, and align their DSS with domain-specific knowledge. The discussion section of each chapter elaborated on both theoretical and practical contributions, offering valuable insights into the primary findings of this research. These findings and contributions are used to address and answer the research questions posed at the beginning of this doctoral thesis (refer to Section 1.4.1). The remainder of this section discusses the extent to which each research question was addressed.

RQ1: How can Semantic Web technologies complement current BI systems?

RQ1 explored how Semantic Web technologies can complement existing BI systems. The systematic literature review presented in Chapter 2 explores the utilization of ontologies within DW/BW systems, focusing on their incorporation, integration, and impact on various tasks throughout the DW/BI lifecycle; hence, answering RQ1.

Ontologies play a crucial role in multiple DW/BI lifecycle tasks, notably dimensional modeling and requirement analysis. They streamline dimensional design by identifying business entities, relationships, and facts [52, 18, 4]. Requirement analysis benefits from ontologies by supporting requirements elicitation and resolving ambiguity [30, 2]. In ETL processes, ontologies aid in the configuration and instantiation of ETL patterns [43]. Ontologies also enhance metadata integration from multimedia or NoSQL databases, improving the overall integration process of non-structured data [63, 49]. Exploration of BI models can leverage ontologies and their semantics, enabling inference capability and interoperability. Works like [48, 39] transform dimensional models into OWL ontologies or RDF Data Cubes, facilitating knowledge extraction and data analysis. The literature cites diverse reasons for employing SW techniques in DW/BI systems. Motivations include addressing data/semantic heterogeneity, enhancing interoperability, facilitating integration, and providing semantic content for requirement and data analysis as detailed in the discussion section of Chapter 2.

Increased data and semantic interoperability can be key to enriching and enhancing decision-making in existing DSS, facilitating the integration of DSS data with domain-specific knowledge contained in external sources. This unexplored research gap led to the definition of RQ2.

RQ2: How can the interoperability between DSS and other systems enhance decision-making using Semantic Web technologies at strategic and operational levels?

RQ2 delved into the potential benefits of interoperability between DSS and other systems using SW technologies in enhancing decision-making processes. Many organizations have already implemented DSS to support their analytical and decision-making processes. However, these systems typically can not handle all organizational data, which is typically distributed across various structured and unstructured sources, such as operational systems, documents, and strategy reports. Additionally, the emergence of new systems and data sources in post-production can require costly updates and new ETL tasks to integrate this additional information into the existing DSS.

Ontologies and other SW techniques promote interoperability and heterogeneity through the sharing, reuse, and analysis of knowledge. These SW techniques can facilitate seamless data exchange and integration across heterogeneous systems by providing a common framework for data representation and its semantics. This shared conceptualization and formalization of knowledge facilitates the integration of data from diverse sources with DSS, providing a comprehensive and holistic perspective for decision-making. Moreover, due to their semantically rich formalization, ontologies can capture the rich meaning and context of data, allowing managers and decision-makers to better understand, analyze, and infer over complex domain-specific knowledge, leading to more informed and data-driven decisions, both strategically and operationally.

The Integration Framework (artifact #1) was designed and developed during this doctoral thesis to address and explore this RQ. This artifact allows the integration of DSS with other systems, enabling DSS data to be enriched and complemented with data or information from external sources (and vice-versa). Knowledge from other sources, such as operational systems or documents (e.g., strategy reports), can now be formalized and semantically integrated with the DSS, providing additional data and information that can be used by any BI application to enhance the decision-making process.

RQ3: To what extent can the use of Semantic Web technologies improve the interoperability between DSS and operational systems?

The findings related to RQ3 were obtained during DSRM's Iter. 2 and are discussed in Chapter 3. The Integration Framework (artifact #1) was used to integrate BIM and AMS, through the use of SW representations, namely the Road Structures Ontology (#2). Furthermore, the API Services (artifact #4) was developed to provide access to this SW representation in the BIM environment. CoDEC project (see Section 1.3.1) was the case study used to study the impact of SW-enabled interoperability between DSS and operational systems.

At an operational level, data and information can be shared across systems, providing users with the additional necessary information and knowledge for decision-making. The integration between systems enables data from operational systems and other sources to be accessed and analyzed in the decision-support environment of their DSS, improving the efficiency of the decision-making process. Moreover, ontology-based analysis can be enabled in the decision-support environment, allowing users to take advantage of the knowledge representation and inference provided by these encoded shared conceptualizations and other SW technologies. These technologies can be used to validate domain-specific knowledge against the DSS repository data (e.g., ensure that each structural element is represented by a BIM model entity), or automatically infer new knowledge, allowing, for example, risk evaluation to be automatically performed based on previous inspections or sensor data analysis, which can be automatically displayed as an alert within the BIM environment.

During the demonstration and evaluation phases of Iter. 2, the integration of external data from operational systems (AMS) into the DSS environment was demonstrated by leveraging research artifacts #1, #2, and #4, enhancing the DSS's analytical capabilities.

RQ4: How can the strategic elements of a BSC be formalized, ensuring their alignment with the various organizational levels (strategic, tactical, and operational)?

RQ4 focused on effectively representing strategic elements, particularly related the Balanced Scorecard framework. This research question was directly addressed during this work's third iteration of DSRM, through the design and development of the Strategy Ontology (artifact #3). The Balanced Scorecard Ontology, introduced in Chapter 4, provides a formal, structured, and semantically rich representation of the BSC framework, ensuring consistency in how strategic objectives, performance indicators, and their relationships are defined and interpreted, and providing decision-makers with a shared and unambiguous understanding of the strategy. This knowledge representation can capture the complex inter-dependencies and cause-and-effect relationships between various components, providing a deeper understanding of how they impact one another. The BSO also formalizes the cascading of BSCs, ensuring their alignment across the various organizational levels (strategic, tactical, and operational).

The BSO also enables automated reasoning, providing logical inferences that can help to identify implicit relationships or conflicts within the proposed BSC model. For example, it can provide rules that enable the detection of wrongfully defined strategic objectives or alerts when certain indicators are irrelevant to the organization's strategy (i.e., are not being used to evaluate any objective or long-term goal).

RQ5: To what extent can the use of Semantic Web technologies improve the interoperability between DW/BI systems and the organizational strategy?

The Integration Framework (version 3 of artifact #1) and its components were demonstrated in the DW/BI Strategy Analysis in a Public Organization (LNEC) case study (see Section 1.3.2), uncovering findings related to RQ5.

The Strategy Ontology (artifact #3) is a key component of the Integration Framework, enabling the formalization of the organizational strategy. Following the work presented in Chapter 4, which described the formalization of the BSC framework, additional demonstrations and evaluations were conducted concerning the BSO. Chapter 5 showcased how ontologies and other SW techniques can effectively support strategy management processes, such as performance indicators evaluation and monitoring and validation of cause-and-effect relationships between strategic objectives. These findings allowed us to explore how existing DW/BI systems can leverage this formalization to enhance the decision-making process, and how DW/BI systems can be a vital component to support data-driven strategy management processes.

Chapter 6 demonstrated the integration of DW/BI systems with strategic knowledge, describing the fourth iteration of DSRM (Iter. 4) of this doctoral thesis. The use of SW techniques streamlines data sharing across systems and organizations, facilitating efficient data collection, analysis, and reporting of performance indicators. The Integration Framework (version 3 of artifact #1) provides a semantic layer that integrates, aligns, and traces strategic models with organizational DW/BI. This alignment allows for the automatic retrieval of performance indicator values, increasing monitoring frequency with reduced human intervention and minimizing the probability of errors. This interoperability also supports compliance and governance initiatives by improving external communication and alignment with regulatory requirements, enabling automatic evaluation and reporting of compliance indicators. Furthermore, this integration enables empirical data-driven validation of the strategy formulation and ensures the alignment between data and performance indicators.

Moreover, the integration enriches conventional performance analysis conducted through DW/BI systems by incorporating strategic context, such as targets and objectives. This enhancement provides managers with deeper insights into how decisions align with broader organizational goals, improving the decision-making process's efficiency, quality, and timeliness. Overall, the integration of SW techniques in DW/BI systems empowers managers with user-friendly information and tools, enabling informed, data-driven strategic decisions aligned with the organizational strategy.

7.3. Future Work and Limitations

This section outlines future research directions, addressing limitations not covered in this doctoral thesis.

This doctoral thesis contributes to addressing the gap identified during the SRL, presented in Chapter 2, namely the use of ontologies to support and improve the analysis and exploration of existing DW/BI. However, other research gaps were not covered, such as the use of ontologies to integrate DW/BI systems (structured data) with unstructured data. From an IS research perspective, there is a clear interest in creating an integrated ecosystem that enables the analysis of both structured and unstructured data [25]. As

Ravat and Zhao [50] state, whether the DW coexists or is part of a Data Lake architecture is still a matter of debate. However, information should always flow between the two, and metadata management systems should be in place to allow users to find the relevant data and cross-reference information as transparently and directly as possible. Ontologies could provide a missing connection point between DW data and other data types that are inside or outside the system/architecture, preventing data swamps [20, 41]. This interoperability could, for example, be ensured through the metadata representation of each repository.

The ontology population processes performed during this research are dependent on manual intervention, including the mapping between sources and ontology entities. While this research provides solutions for the automatic population of data related to performance indicators (see Chapter 6), enabled by the Integration Framework (version 3 of artifact #1), most of the domain-specific knowledge was imported using Cellfie in Protégé, requiring manual intervention. This semi-automated process was performed during Iter. 2 to populate the Road Structures Ontology (artifact #2) with structural and operational data. The mapping between this information and the semantic representation of the BIM model, ensured by ifcOWL, is done manually. The Strategy Ontology (artifact #3) was populated during Iter. 3, based on non-structured data related to strategic reports, following the same semi-automated process. It is imperative to develop automated tools or algorithms to simplify the process of ontology population, both in structured and unstructured data, and improve the efficiency of these solutions.

This research explored the application of the artifacts in two real-world case studies: CoDEC research project (civil engineering) and DW/BI Strategy Analysis in a Public Organization. While there are no perceived barriers to generalizing and implementing the proposed artifacts in other scenarios, the adaptability of these solutions must be explored in other contexts or industries. For example, the application of the domain-specific artifacts can be further studied by extending the Road Structure Ontology (artifact #2) for other highway structures or using the Strategy Ontology (artifact #3) to assess strategy formulation and execution in other organizations (public or private).

Furthermore, while the Integration Framework (artifact #1) is aimed at integrating domain-specific ontologies with the DSS, ontologies can also be used to increase the interoperability between the existing domains within the same semantic layer. There is no perceived barrier to using, for example, the Strategy Ontology to enable the evaluation of AMS data provided by the Road Structures Ontology based on performance indicators with defined targets, as long as the required mappings or semantic links are developed. The same principle can be applied to integrate strategic information with Enterprise Architecture (EA) models, such as ArchiMate, to ensure an alignment between strategy and other EA layers, such as business, application, and infrastructure. This alignment would ensure the integration between strategic business vision down to the IT infrastructure, enabling EA inter-layer analyses with the domain-specific knowledge.

The work presented in Chapters 4, 5 and 6 demonstrated how ontologies and SW technologies can be used to enable strategy validation and analysis for managers and decision-makers. These works showed how external applications can benefit from ontology knowledge to provide end-users with strategic information. However, additional work is required to fully showcase the potential of these technologies as part of a fully integrated BI solution, where managers can interact, monitor, analyze, and receive alerts or analytical recommendations, regarding the execution of the strategy. These BI solutions can also be supported or enriched with natural language processing or AI techniques, allowing users, for example, to define BI queries through natural language tuned by knowledge from the underlying ontologies, namely the DW Ontology (LDWOWL) (artifact #5).

Moreover, as stated in Chapter 6, LDWOWL was designed to formally and semantically store information concerning the DW's conceptual and logical models as a means to relate this knowledge to other domains (such as strategy). Notwithstanding, this "light" ontology, as the name suggests, does not cover the entire scope of the DW/BI domain but is limited to some aspects of requirements analysis and dimensional modeling, particularly for star schemas. For example, concepts such as aggregated and derived fact tables[5], while taken into account, were not tested, and some advanced concepts (e.g., Bridges[6]) and variable depth hierarchies (i.e., ragged hierarchies) were not taken into account. Moreover, although queries followed the typical BI structure, the ontology does not cover the full range of BI query complexity. For example, the ontology does not currently cover logical connectives, such as conjunctions and disjunctions between multiple filter clauses. Hence, there are many opportunities for further exploration in future research.

The current development of the LDWOWL was idealized as a proof-of-concept. A formal ontology development methodology was not followed during the scope of this work, and the ontology was not formally validated (with tools such as OOPS! - Ontology Pitfall Scanner! [46]) or evaluated (by using competency questions). This ontological artifact's full scope, impact, and potential should be researched and analyzed in isolation from the presented framework. Further, the use of SW techniques, such as SWRL, SHACL, and SPARQL, for knowledge validation and inference should continue to be explored.

Kaplan and Norton [28] introduced the Execution Premium Process (see Figure 6.2) as a comprehensive framework for organizations to operationalize their strategies. This model emphasizes the critical importance of aligning organizational processes, resources, and actions with strategic objectives. It enables organizations to effectively execute their strategies, fostering accountability and promoting continuous improvement. The need to ensure that strategies remain comprehensive, transparent, and flexible is still at the forefront of BSC research, which points to integrated reporting mechanisms as one of the growing trends [36]. In fact, the effective execution of the Execution Premium Process relies heavily on access to accurate, timely, and relevant data from various sources, requiring a reliable monitoring of KPIs and strategic objectives. Furthermore, achieving

[5]Derived fact tables include Accumulating Snapshots, Periodic Snapshot, and Merged fact tables [3].
[6]Multiple-to-multiple relationships between facts and dimensions, or between dimensions and attributes.

alignment across different departments and organizational levels is critical for the successful execution of the strategy. The artifacts produced in this doctoral thesis contribute to a holistic view of an organization, which can lead to a more realistic and cost-effective implementation of the strategy.

References

[1] CoDEC Project Report - Deliverable D3A: Pilot projects report and consolidated implementation resources, 2021. https://www.codec-project.eu/Resources/projectreports.

[2] B. Aadil, A.A. Wakrime, L. Kzaz, and A. Sekkaki. Automating data warehouse design using ontology. In *Proceedings of 2016 International Conference on Electrical and Information Technologies (ICEIT)*, pages 42–48, 2016. doi: 10.1109/EITech.2016.7519618.

[3] Christopher Adamson. *Star schema the complete reference*. McGraw Hill Professional, 2010.

[4] Glenda Amaral and Giancarlo Guizzardi. On the application of ontological patterns for conceptual modeling in multidimensional models. In *Advances in Databases and Information Systems*, pages 215–231, Cham, 2019. Springer International Publishing. doi: 10.1007/978-3-030-28730-6_14.

[5] António Lorvão Antunes, Elsa Cardoso, and José Barateiro. Incorporation of ontologies in data warehouse/business intelligence systems-a systematic literature review. *International Journal of Information Management Data Insights*, 2(2):100131, 2022.

[6] José Barateiro, Paula Couto, António Antunes, António Sebastião, and Maria Alzira Santos. Enquadramento estratégico do lnec - proposta de mapa estratégico. Technical report, National Laboratory for Civil Engineering (LNEC), Lisbon, Portugal, 09 2021. In Portuguese.

[7] Sukalpa Biswas, John Proust, Tadas Andriejauskas, Alex Wright, Carl van Geem, Darko Kokot, António Antunes, Vânia Marecos, José Barateiro, Shubham Bhusari, et al. Codec: Connected data for road infrastructure asset management. In *IOP Conference Series: Materials Science and Engineering*, volume 1202, page 012002. IOP Publishing, 2021.

[8] Sukalpa Biswas, John Proust, Tadas Andriejauskas, Alex Wright, Carl Van Geem, Darko Kokot, António Antunes, Vânia Marecos, José Barateiro, Shubham Bhusari, et al. Demonstrating connectivity and exchange of data between bim and asset management systems in road infrastructure asset management. In *International Road Federation World Meeting & Exhibition*, pages 379–392. Springer, 2021.

[9] John Bryson and Bert George. Strategic management in public administration. In *Oxford Research Encyclopedia of Politics*. 2020.

[10] John M Bryson. *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement*. John Wiley & Sons, 2018.

[11] Betsy Burton, Lee Geishecker, K Schelegel, Bill Hostmann, Tom Austin, Gareth Herschel, Alex Soejarto, and Nigel Rayner. Business intelligence focus shifts from tactical to strategic. *Retrieved from Gartner database (G00139352)*, 2006.

[12] Lawrence Corr and Jim Stagnitto. *Agile data warehouse design: Collaborative dimensional modeling, from whiteboard to star schema*. DecisionOne Consulting, 2011.

[13] Aaron Costin, Alireza Adibfar, Hanjin Hu, and Stuart S Chen. Building information modeling (bim) for transportation infrastructure–literature review, applications, challenges, and recommendations. *Automation in construction*, 94:257–281, 2018.

[14] Bert George, Richard M Walker, and Joost Monster. Does strategic planning improve organizational performance? a meta-analysis. *Public Administration Review*, 79(6): 810–819, 2019.

[15] Robert M Grant. *Contemporary strategy analysis*. John Wiley & Sons, 2021.

[16] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993. doi: 10.1006/knac.1993.1008.

[17] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In *Handbook on ontologies*, pages 1–17. Springer, 2009.

[18] M. Gulic. Transformation of owl ontology sources into data warehouse. *2013 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1143–1148, 2013.

[19] Shivam Gupta, Arpan Kumar Kar, Abdullah Baabdullah, and Wassan AA Al-Khowaiter. Big data with cognitive computing: A review for the future. *International Journal of Information Management*, 42:78–89, 2018. doi: 10.1016/j.ijinfomgt.2018.06.005.

[20] Ahmed A Harby and Farhana Zulkernine. From data warehouse to lakehouse: A comparative review. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 389–395. IEEE, 2022.

[21] Teresa M Harrison and Theresa A Pardo. Data, politics and public health: Covid-19 data-driven decision making in public discourse. *Digital Government: Research and Practice*, 2(1):1–8, 2020.

[22] Timo Hartmann and Amy Trappey. Advanced engineering informatics - philosophical and methodological foundations with examples from civil and construction engineering. *Developments in the Built Environment*, 4:100020, 2020. ISSN 2666-1659. doi: https://doi.org/10.1016/j.dibe.2020.100020. URL `https://www.sciencedirect.com/science/article/pii/S2666165920300168`.

[23] John Hebeler, Matthew Fisher, Ryan Blace, and Andrew Perez-Lopez. *Semantic web programming*. John Wiley & Sons, 2011.

[24] Pascal Hitzler. A review of the semantic web field. *Commun. ACM*, 64(2):76–83, jan 2021. ISSN 0001-0782. doi: 10.1145/3397512.

[25] B. Inmon, M. Levins, and R. Srivastava. *Building the Data Lakehouse*. Technics Publications, 2021. ISBN 9781634629669.

[26] Karuna Pande Joshi and Srishty Saha. A semantically rich framework for knowledge representation of code of federal regulations. *Digital Government: Research and Practice*, 1(3):1–17, 2020.

[27] Evangelos Kalampokis, Nikos Karacapilidis, Dimitris Tsakalidis, and Konstantinos Tarabanis. Understanding the use of emerging technologies in the public sector: A review of horizon 2020 projects. *Digital Government: Research and Practice*, 4(1): 1–28, 2023.

[28] Robert S Kaplan and David P Norton. *The execution premium: Linking strategy to operations for competitive advantage*. Harvard business press, 2008.

[29] Robert S Kaplan, David P Norton, et al. *The balanced scorecard: measures that drive performance*, volume 70. Harvard business review US, 1992.

[30] Selma Khouri and Bellatreche Ladjel. A methodology and tool for conceptual designing a data warehouse from ontology-based sources. In *Proceedings of the ACM 13th International Workshop on Data Warehousing and OLAP*, page 19–24, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450303835. doi: 10.1145/1871940.1871946.

[31] Ralph Kimball and Margy Ross. *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons, 2013.

[32] Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker. *The data warehouse lifecycle toolkit*. John Wiley & Sons, 2008.

[33] Darko Kokot. Asset management approach for transport infrastructure networks: The am4infra project. In *Airfield and Highway Pavements 2019: Design, Construction, Condition Evaluation, and Management of Pavements*, pages 374–381. American Society of Civil Engineers Reston, VA, 2019.

[34] Martin Král. 20-year history of performance measurement in the local public sector: a systematic review. *International Journal of Public Administration*, 45(9):726–740, 2022.

[35] Jitender Kumar, Neha Prince, and H Kent Baker. Balanced scorecard: A systematic literature review and future research issues. *FIIB Business Review*, 11(2):147–161, 2022.

[36] Satish Kumar, Weng Marc Lim, Riya Sureka, Charbel Jose Chiappetta Jabbour, and Umesh Bamel. Balanced scorecard: trends, developments, and future directions. *Review of Managerial Science*, pages 1–43, 2023.

[37] Francesca Manes-Rossi, Giuseppe Nicolò, and Daniela Argento. Non-financial reporting formats in public sector organizations: a structured literature review. *Journal of Public Budgeting, Accounting & Financial Management*, 32(4):639–669, 2020.

[38] ELISABETTA Marcovaldi and MAURIZIO Biccellari. Asset data dictionary. *Deliverable D3*, 1, 2018.

[39] A. Matei, K.-M. Chao, and N. Godwin. OLAP for multidimensional semantic web databases. *Lecture Notes in Business Information Processing*, 206:81–96, 2015. doi:

10.1007/978-3-662-46839-5_6.

[40] Petar Milić, Nataša Veljković, and Leonid Stoimenov. Semantic technologies in e-government: Toward openness and transparency. *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues*, pages 55–66, 2018.

[41] Fatemeh Nargesian, Erkang Zhu, Renée J Miller, Ken Q Pu, and Patricia C Arocena. Data lake management: challenges and opportunities. *Proceedings of the VLDB Endowment*, 12(12):1986–1989, 2019.

[42] Natalya F Noy, Deborah L McGuinness, et al. Ontology development 101: A guide to creating your first ontology, 2001.

[43] B. Oliveira and O. Belo. An ontology for describing etl patterns behavior. In *DATA 2016 - Proceedings*, pages 102–109, 2016. doi: 10.5220/0005974001020109.

[44] Jeff Pan. Resource description framework. In *Handbook on ontologies*, pages 71–90. 05 2009. doi: 10.1007/978-3-540-92673-3_3.

[45] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[46] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2):7–34, 2014.

[47] Daniel J Power. *Decision support basics*. Business Expert Press, 2009.

[48] N. Prat, J. Akoka, and I. Comyn-Wattiau. Transforming multidimensional models into OWL-DL ontologies. In *Proceedings - International Conference on Research Challenges in Information Science*, 2012. doi: 10.1109/RCIS.2012.6240451.

[49] M. Pticek and B. Vrdoljak. Semantic web technologies and big data warehousing. In *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) - Proceedings*, pages 1214–1219, 2018. doi: 10.23919/MIPRO.2018.8400220.

[50] Franck Ravat and Yan Zhao. Data lakes: Trends and perspectives. In *Database and Expert Systems Applications*, pages 304–313. Springer International Publishing, 2019. ISBN 978-3-030-27615-7. doi: 10.1007/978-3-030-27615-7_23.

[51] Petar Ristoski and Heiko Paulheim. Semantic web in data mining and knowledge discovery: A comprehensive survey. *Journal of Web Semantics*, 36:1–22, 2016. doi: 10.1016/j.websem.2016.01.001.

[52] Oscar Romero and Alberto Abelló. A framework for multidimensional design of data warehouses from ontologies. *Data Knowledge Engineering*, 69(11):1138–1157, 2010. ISSN 0169-023X. doi: 10.1016/j.datak.2010.07.007.

[53] Catherine Roussey, Francois Pinet, Myoung Ah Kang, and Oscar Corcho. An introduction to ontologies and ontology engineering. In *Ontologies in Urban Development Projects*, pages 9–38. Springer London, London, 2011. ISBN 978-0-85729-724-2. doi:

10.1007/978-0-85729-724-2_2.

[54] Pegdwendé Sawadogo and Jérôme Darmont. On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, 56(1):97–120, 2021. doi: 10.1007/s10844-020-00608-7.

[55] Ramesh Sharda, Dursun Delen, Efraim Turban, J Aronson, and T Liang. *Business intelligence and analytics*. Pearson Edition Limited, 10th edition, 2015.

[56] Barry Smith. Ontology. *Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166, 2003.

[57] Grimm st Stephan, Hitzler st Pascal, and Abecker st Andreas. Knowledge representation and ontologies. *Semantic Web Services: Concepts, Technologies, and Applications*, pages 51–105, 2007. doi: 10.1007/3-540-70894-4_3.

[58] Rudi Studer, V Richard Benjamins, and Dieter Fensel. Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2):161–197, 1998. doi: 10.1016/S0169-023X(97)00056-6.

[59] Xiaomeng Su and Lars Ilebrekke. A comparative study of ontology languages and tools. In *International Conference on Advanced Information Systems Engineering*, pages 761–765. Springer, 2002. doi: 10.1007/3-540-47961-9-62.

[60] Xiaomeng Su and Lars Ilebrekke. A comparative study of ontology languages and tools. In *International Conference on Advanced Information Systems Engineering*, pages 761–765. Springer, 2002. doi: 10.1007/3-540-47961-9-62.

[61] Alex Tawse and Pooya Tabesh. Thirty years with the balanced scorecard: What we have learned. *Business Horizons*, 66(1):123–132, 2023.

[62] E. Turban, R. Sharda, and D. Delen. *Decision Support and Business Intelligence Systems*. USA: Pearson Education, Inc., 9th edition, 2010.

[63] Andrei Vanea and Rodica Potolea. Semantically enhancing multimedia data warehouses - using ontologies as part of the metadata. In *Proceedings of the 13th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 163–168. INSTICC, SciTePress, 2011. ISBN 978-989-8425-53-9. doi: 10.5220/0003434701630168.

[64] Hugh Watson and Barb Wixom. The current state of business intelligence. *Computer*, 40:96 – 99, 10 2007. doi: 10.1109/MC.2007.331.

[65] Pabasara Upalakshi Wijeratne, Chathuri Gunarathna, Rebecca Jing Yang, Peng Wu, Keith Hampson, and Ammar Shemery. Bim enabler for facilities management: A review of 33 cases. *International Journal of Construction Management*, 24(3):251–260, 2024.