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/

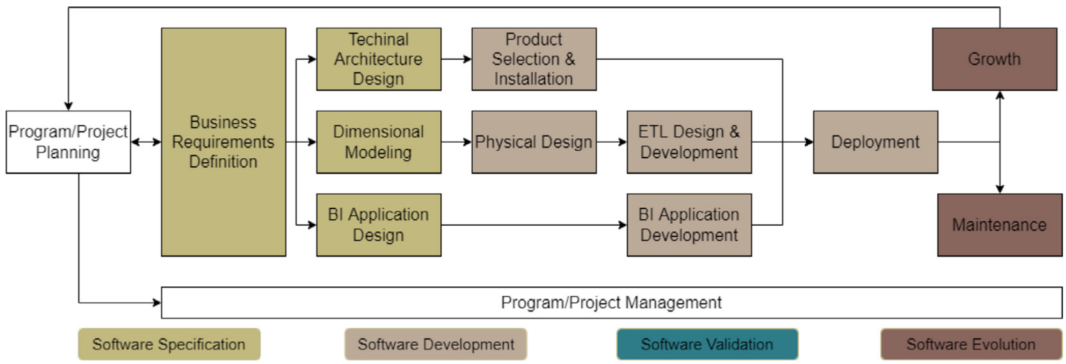

Fig. 1. Kimball's DW/BI lifecycle methodology. Adapted from Kimball et al. (2008).

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

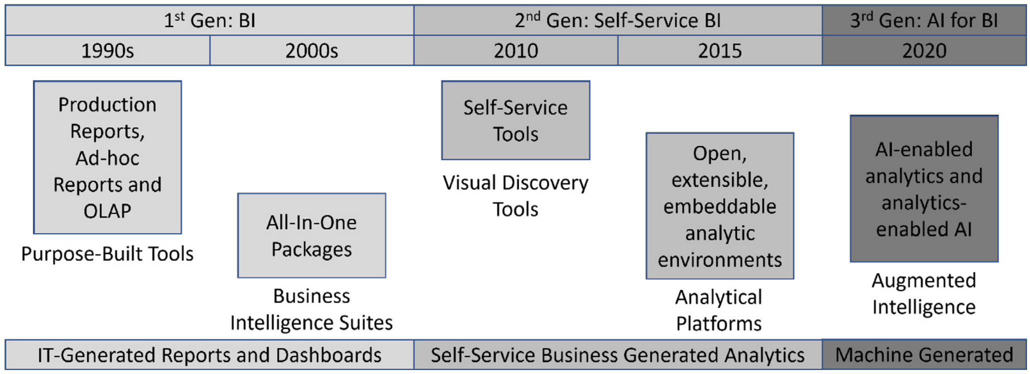

Fig. 2. Evolution of BI. Adapted from Ereth & Eckerson (2018).

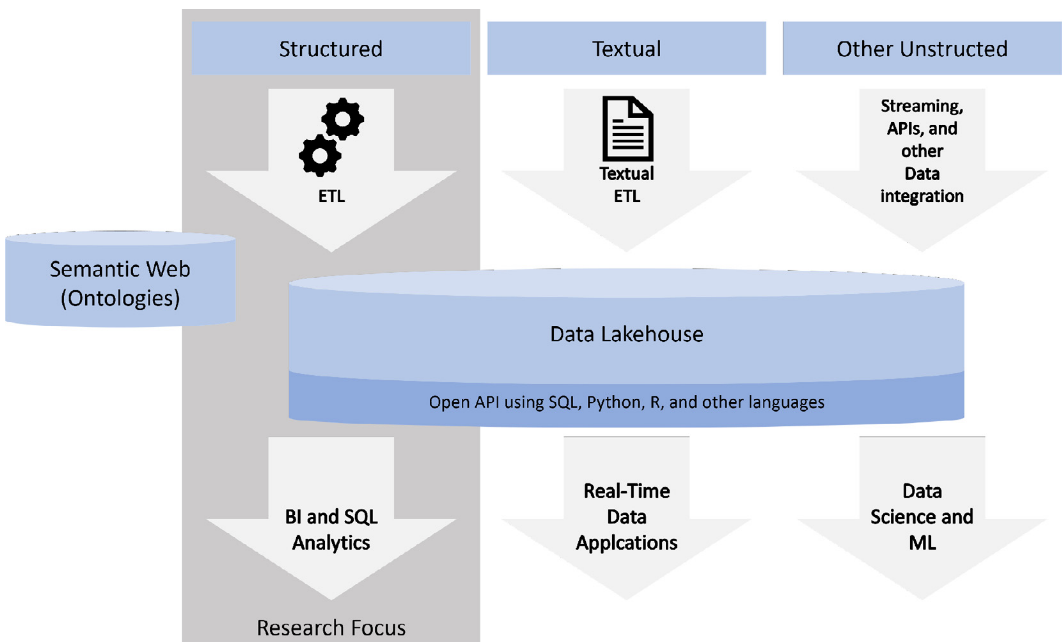

Fig. 3. Data Lakehouse Architecture. This study will focus on the impact of SW on structured data and its analysis. Adapted from Inmon et al. (2021).

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

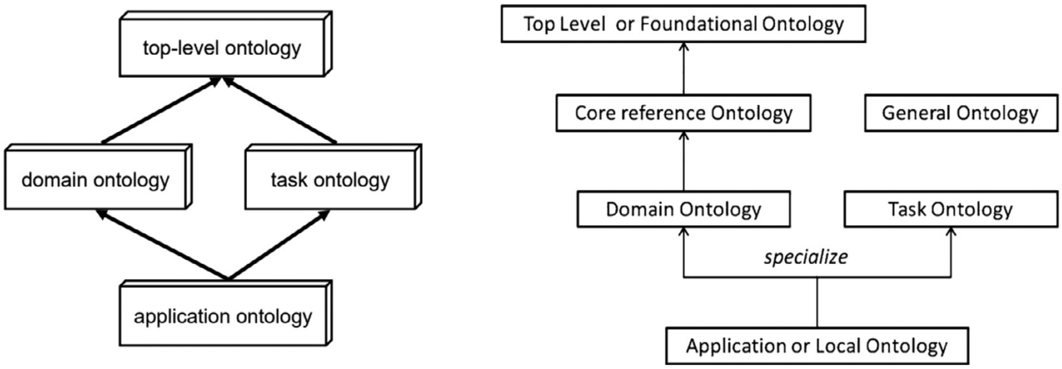

Fig. 4. Ontology types hierarchy based on Scope. Retrieved from Stephan et al. (2007) (left) and Roussey et al. (2011) (right).

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

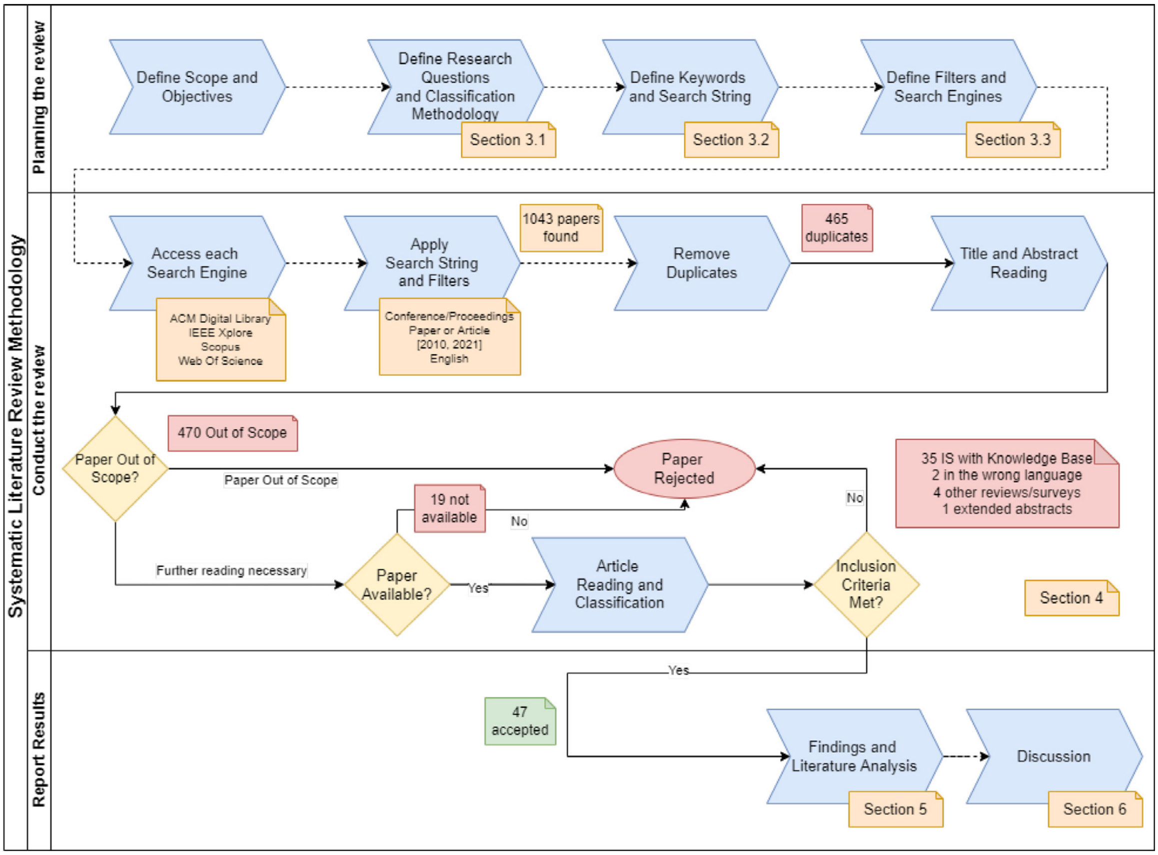

Fig. 5. SLR methodology.

(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

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.

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**

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

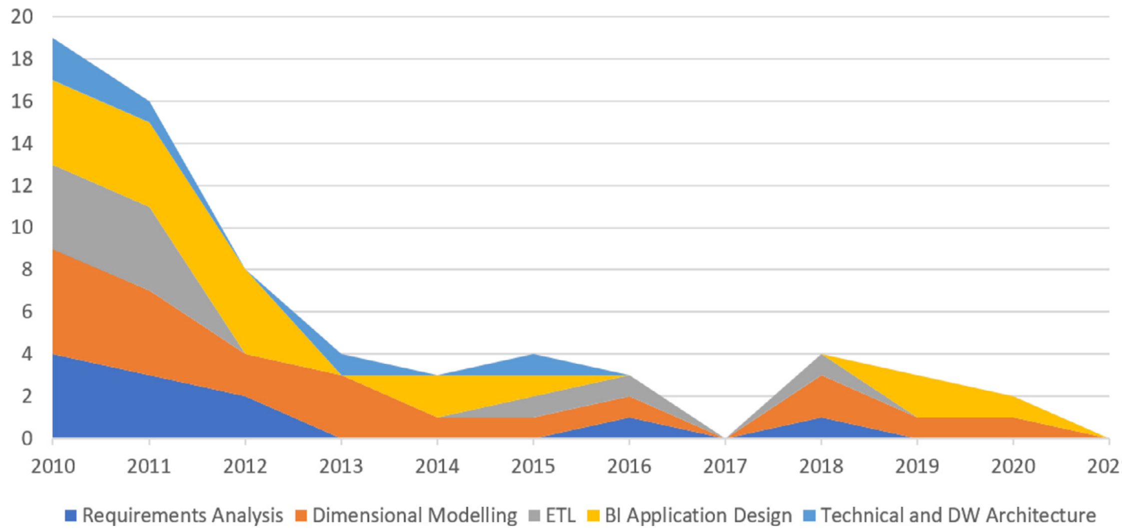

Fig. 6. Evolution of works published per DW/BI Task.

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/

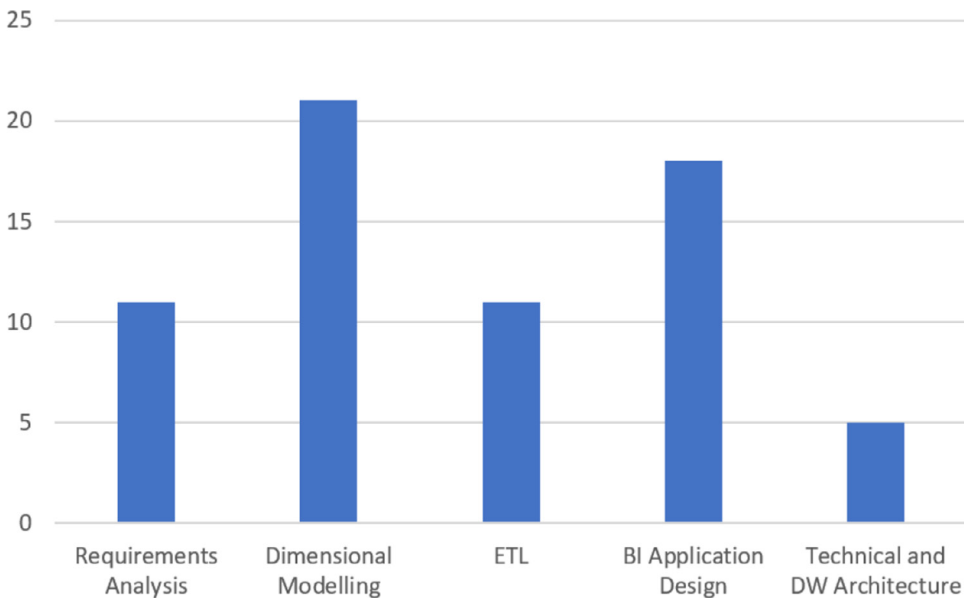

Fig. 7. Number of works per Kimball's DW/BI lifecycle task.

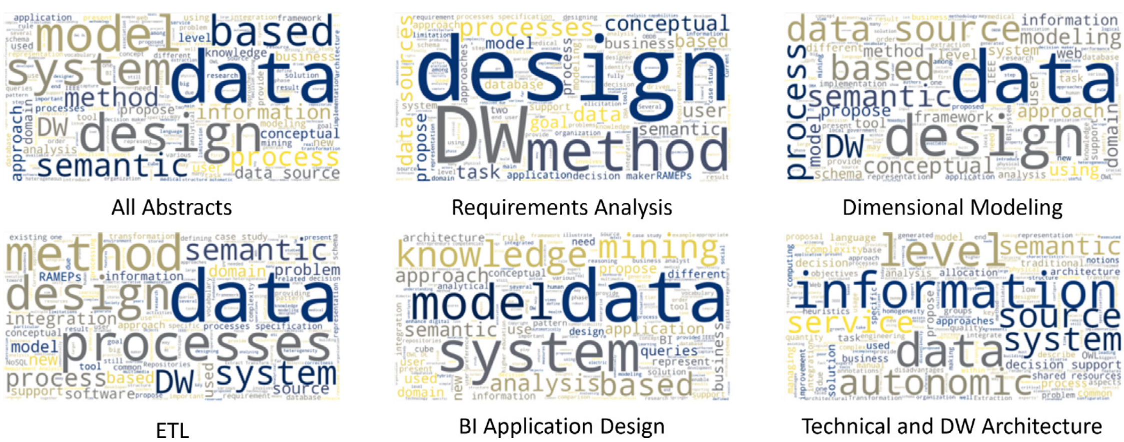

Fig. 8. Abstracts Word clouds.

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.