WILEY   **WIREs**
DATA MINING AND KNOWLEDGE DISCOVERY

OVERVIEW

Using Semantic Web technologies in the development of data warehouses: A systematic mapping

Ricardo Gacitua[1] | Jose N. Mazon[2*] | Ania Cravero[1]

[1]Department of Computer Science and Informatics, Universidad de La Frontera, Temuco, Chile

[2]Department of Software and Computing Systems, Universidad de Alicante, Alicante, Spain

Correspondence
Ricardo Gacitua, Department of Computer Science and Informatics, Universidad de La Frontera, Temuco, Chile.
Email: ricardo.gacitua@ufrontera.cl

[*]**Present address**
Jose N. Mazon, Department of Computer Science and Informatics, Universidad de Alicante, Alicante, Spain.

The exploration and use of Semantic Web technologies have attracted considerable attention from researchers examining data warehouse (DW) development. However, the impact of this research and the maturity level of its results are still unclear. The objective of this study is to examine recently published research articles that take into account the use of Semantic Web technologies in the DW arena with the intention of summarizing their results, classifying their contributions to the field according to publication type, evaluating the maturity level of the results, and identifying future research challenges. Three main conclusions were derived from this study: (a) there is a major technological gap that inhibits the wide adoption of Semantic Web technologies in the business domain;(b) there is limited evidence that the results of the analyzed studies are applicable and transferable to industrial use; and (c) interest in researching the relationship between DWs and Semantic Web has decreased because new paradigms, such as linked open data, have attracted the interest of researchers.

This article is categorized under:
    Fundamental Concepts of Data and Knowledge > Information Repositories
    Fundamental Concepts of Data and Knowledge > Knowledge Representation
    Algorithmic Development > Web Mining

KEYWORDS
data warehouses, Semantic Web, systematic mapping

1 | INTRODUCTION

To include external web data in the traditional data warehouse (DW) systems, and the traditional online analytical processing (OLAP) processes, is a promising way for broadening traditional business intelligence (BI) analysis Abelló et al. (2013). In the context of BI, a DW is used to collect, organize and store subject-oriented, integrated, time-variant, and nonvolatile data (Inmon, 1992). Traditionally, a DW has been defined as a historical data repository containing data collected from a wide variety of heterogeneous sources by means of extraction–transformation–loading (ETL) processes (Kimball & Ross, 2002). On the other hand, multidimensional (MD) processing, also called OLAP, is an approach for responding to MD analytical queries. Research focusing on DWs and OLAP has led to the creation of important technologies for the design, administration, and use of information systems in decision-making support. Part of the interest in, and success of this field, can be attributed to the demonstrated need for software and tools that help improve data analysis and administration. This is mainly due to the large quantity of information that is being accumulated by corporations as well as scientific databases.

Traditional BI tools, such as OLAP, have been successfully applied to large amounts of data coming from operational databases. However, there is a trend whereby DWs are becoming more and more dynamic, with updates occurring almost in real-time, and with the inclusion of more complex types of data (Henschen, 2015). This situation has forced traditional BI to

open its gates to external data in order to encompass a more heterogeneous and open analysis scenario (Chen, Chiang, & Storey, 2012). Current research envisions that Semantic Web technologies are required for realizing the next generation of DWs (Abelló et al., 2015), as an increasing quantity of semantically annotated data is available over the Internet.[1] To include Semantic Web information in a traditional OLAP analysis process is therefore a promising way to augment traditional BI analyses (Trujillo & Maté, 2012). The Semantic Web is an extension of the Web proposed by the World Wide Web Consortium.[2] Its intended objective is to facilitate the creation of technologies that publish legible data for informatic applications, which are implemented by adding semantic metadata and ontologies to the web (Shadbolt, Berners-Lee, & Hall, 2006). In practical terms, the strength of the Semantic Web lies in its ability to aggregate the semantic annotations of web-published content so that the information can be effectively retrieved and processed, either by humans or by machines, for a wide variety of tasks (Hendler, 2001). The above goals is achieved through the use of a diverse variety of software technologies, such as ontologies (Coral, Francisco, & Mario, 2006) and markup languages (e.g., RDF and OWL) (Saha, 2007), which allow semantic annotations to be added to resources that can either be very simple, or very complex annotations, depending on the requirements.

Although the Semantic Web and DWs, in the context of BI applications, have gone in different research directions over the last few years, some recent results show that the convergence of these two fields is not only inevitable but also beneficial for both sides (Golfarelli, Rizzi, & Cella, 2004; Lather, 2012). The exploration and use of Semantic Web technology has therefore attracted attention from researchers studying DW development (Berlanga, Aramburu, Llidó, & García-Moya, 2014; Golfarelli et al., 2004). Recently, the concept of the web of linked data has appeared as a mechanism for making all data (Klyne & Carroll, 2004) available using the HTTP protocol, as happens with HTML documents (Bizer, Heath, & Berners-Lee, 2009). Linked open data (LOD) have been introduced as a promising paradigm for opening up data because it facilitates data integration on the web (Bizer et al., 2009). Similar to data warehousing approaches, LOD can be prepared to enable sophisticated data analysis. As one of the main problems of DWs is data integration, some researchers propose to use data exchange standards like RDF, to host structured content to publish DW content as linked data. In the business and enterprise domain, there is still a gap between conceptual approaches for modeling architectures for linked data, systems, data models, as well as their implementation, operationalization and execution (Abramowicz, Auer, & Heath, 2016). In addition, businesses need to overcome the inevitable tension between the value they traditionally assign to proprietary data, and the value of opening up.

The proposals of the Semantic Web and data storage come from very different areas. Thus, the studies that have been published have approached the integration of these topics through varied perspectives. As the development of these areas has advanced, there has been increased growth in the number of published reports and results. Hence, summarizing and providing a general analysis of the integration of these topics is quickly becoming necessary. As a result, it is difficult to know the current state of these proposals. Essentially, the impact of these research studies, as well as the maturity level of their results, is still unclear. For example, classifying the scientific contributions made on this topic is useful for researchers, as the topics that address the Semantic Web and DWs have not necessarily been developed in a related way. Indeed, each topic has its own separate conferences and journals. Providing a general summary of the progress in integrating these two areas is therefore useful in summarizing and classifying research as well as in establishing future research challenges. The systematic mapping (MP) study is a secondary study method that has recently attracted considerable attention (Petersen, Feldt, Mujtaba, & Mattsson, 2008; Petersen, Vakkalanka, & Kuzniarz, 2015) largely because it offers a specific way of systematically reviewing the literature in regards to a particular topic (Kitchenham, 2012). A systematic map structures the type of research reports that have been realized, as well as a categorization of their published results. It usually delivers a visual summary—a map—of their results. A systematic map of the literature has been recommended mainly for research areas with a lack of relevant primary studies, as is the case for the reported use of Semantic Web technologies in DW development. In recent years, there has been an increase in the number of reports related to DWs and the Semantic Web that take into account methodological proposals and techniques, among others. This goes along with the fact that these reports stem from a variety of research areas, such as Databases, Artificial Intelligence, and Natural Language Processing, and the fact that the relationship between the Semantic Web and DWs is a problem of current interest. Given the massive amount of existing data (i.e., Big Data) and its use on the web, it becomes necessary to discern the type of research that is being conducted with respect as to how Semantic Web technologies are being used in DW development. A systematic map allows for the categorization of the results and for a summary to be presented graphically.

The objective of this study is to examine recently published research articles that consider the use of Semantic Web technologies in DW development in the business domain. We conducted an SM study in order to study the literature with the objectives of summarizing their results, evaluating the maturity level of those results and identifying challenges for future research in this field. More specifically, we focused on the following research question (RQ): *What is the state of the recent research covering the use of Semantic Web technology in DW development?* We expounded upon the general objectives by examining more specific objectives, which were:

1. Synthesizing evidence in order to propose relevant suggestions for practical applications of Semantic Web technologies in DW development.
2. Creating a classification for the published research on DWs and the Semantic Web.
3. Identifying research trends, open problems, and areas for improvement within a research body that considers both DWs and the Semantic Web together.

The article is structured in the following manner: Section 2 summarizes the motivation and contribution of this article. Section 3 offers a general methodology and the basis of the research focus. Section 4 presents the results and a discussion of the main findings of the study. In addition, a review of some industrial tools and projects within the broad topic of DWs and the Semantic Web is provided. Section 5 presents the research challenges posed by this study. Finally, Section 8 offers the conclusions of the study and recommendations for future work.

2 | MOTIVATION AND CONTRIBUTION

The main contributions of this article can be summarized as follows:

1. The identification of the main proposals that use Semantic Web technology in DW development.
2. The identification of the Semantic Web technologies that are being used.
3. A classification framework for the contributions of scientific studies.
4. A discussion of the gaps identified in the articles reviewed.
5. An outline detailing future research challenges.

In accordance with Kitchenham, Budgen, and Brereton (2011), the contributions of a systematic map (as the one presented here) are addressed by the following people:

- Beginner researchers who are initiated in the use of the Semantic Web in the development of DWs.
- Experienced researchers who may do so as qualitative reference work that saves time for subsequent studies, provides an understanding of the existing literature on specific topics, and allows them to identify the need for conducting additional research in specific areas. This is particularly important for highly multidisciplinary fields such as the intersection between DWs and the Semantic Web, where different subcommunities publish in different publication venues, and even experienced researchers are not necessarily aware of all the facets and contributions in the broader research field.
- Industrial actors who need a thorough introduction and overview of the research field of study will find it useful. A mapping study allows industry to get an overview of state-of-the-art innovations, and to identify trends and clusters of research studies that are suitable and applicable for their particular business use, aiding communication and knowledge transfer between academia and industry. Given the relative immaturity of research on the use of Semantic Web technologies in the development of DWs, and the rapidly changing technologies, the relevance for industry cannot be overstated, since, for example, standards, tools, and case studies that are of particular interest to industry are not yet sufficiently present in the literature.

3 | LITERATURE REVIEW METHOD

As a research area matures, there is often a strong increase in the number of publications and research results. Thus, the need to summarize and provide general overviews of the topic of interest begins to become more relevant. The method used in this research is an SM study, which provides a methodical and objective approach for identifying the nature and extent of the empirical study data that are available to answer a particular RQ (Kitchenham et al., 2010). SM aims to organize the research undertaken rather than to answer detailed RQs (Kitchenham, 2012). The main objective of an SM study is to provide a synthesis of the research area and identify the quantity, the type of research, and the available results. Indeed, it often becomes necessary to visualize the frequency of publications over time in order to determine trends. We used ideas from da Mota Silveira Neto, Carmo Machado, McGregor, de Almeida, and de Lemos Meira (2011) in order to use an extended SM process. This extended process includes topics not covered by Petersen et al. (2008), such as: a protocol and a classification schema. Figure 1 depicts the SM process. The basic steps of the process are: definition of the protocol and the RQs, execution of a search of relevant articles, filtration of articles, search of key concepts used in summaries and, finally, the extraction of data and the processes of mapping. Each process produces output. The final output of the process is the systematic map.
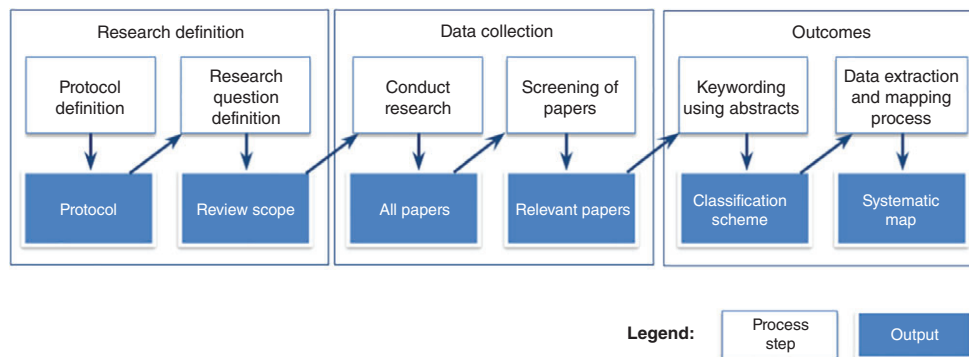
FIGURE 1 The systematic mapping process

The following sections describe the stages that were realized.

3.1 | First phase: research definition

This section presents the first phase of the mapping study process, in which the protocol and RQs are defined.

3.1.1 | Protocol

In this study, the purpose of the protocol is to manage the research objectives, and to define clearly how they can be achieved by defining the RQs and planning how the sources and selected studies will be used to respond to these questions. We adopted this task following systematic review guidelines. The first activity in this study was to develop a protocol, that is, a strategy, defining the basic mapping study procedure.

3.1.2 | Research questions

The objectives of SM are reflected by the RQs. The RQs were framed by four criteria:

- *Population*: The scientific literature that shows the relationship between the Semantic Web and DWs.
- *Study design*: Describes the relationship with a focus on the study. In this case, it is defined as the use of Semantic Web technologies in DW development.
- *Intervention*: Any study that shows some content of Semantic Web technology is used during some stage of the DW development process, such as ETL or OLAP.
- *Outcomes*: This refers to the quantity and type of related evidence regarding Semantic Web technologies and DWs, with particular interest in determining the degree of improvement on the DW development process as a result of the use of Semantic Web technologies.

In this study, restrictions regarding specific outcomes were not considered. As it is necessary to obtain a wide view of the entire area investigated based mainly on the existence of references regarding the Semantic Web and DWs, not specific Semantic Web technologies, the objective is to determine which technologies, defined as part of the Semantic Web, are referenced and used by the research community. Table 1 shows the RQs, which correspond to this study's objectives and are in accordance with the standards indicated by Petersen et al. (2008).

In order to answer the RQs, data were collected from the existing literature. This task, in general terms, involves defining a search strategy, identifying data sources, selecting studies, and analyzing and synthesizing data.

3.2 | Second phase: data collection

In order to answer the RQs, data were collected from the research literature. This activity involved conducting research and the screening of papers.

3.2.1 | Conduct research

Conducting research involved developing a search strategy, and identifying data sources.

TABLE 1 Definition of the RQs (scope of the literature review)

ID	Questions	Objectives
RQ1	Where were the research studies published?	To determine the place of a study's publication (e.g., journals, conferences, others). This makes it possible to know where the publications are concentrated and, using that information, determine the maturity level of their results.
RQ2	What Semantic Web technologies are being used in DW development?	To determine the set of associated Semantic Web technologies, which are being used in DW development. For example, ontologies (as a concept) or some of its related languages, for example, representation languages such as RDF and OWL, query languages such as SPARQL, among others. The answer to this question will make it possible to determine the most commonly used technologies.
RQ3	During which stages in the DW development process are these technologies being used?	To determine in which phases of the DW development process a particular technology is being used. For example, requirements, ETL, or OLAP. This makes it possible to determine the area of greatest interest for researchers.
RQ4	What type of research study was conducted?	To determine what type of study was performed in order to show the use of Semantic Web technology. For example, experiment, case study, project description, and among others. This allows the published studies to be categorized.
RQ5	Is there evidence of improvement due to the use of Semantic Web technology?	To determine if the evidence shows that the presented results correspond to an improvement in the DW development process. This allows the impact of the results on DW development to be evaluated.

Search strategy

The search strategy was developed through the review of the data needed to respond to each RQ. Primary studies were identified using search strings on scientific databases, or searching manually through conference proceedings or publications in specialized journals. An initial set of key words was refined after a preliminary search that retrieved many results of little relevance. Several combinations of search items were used until an appropriate set of key words was reached. The search chain consisted of Boolean expressions composed of key words that described the stages of DW development. Table 2 shows the composition of the search string applied. Terms within a row are connected by the operation "OR," while the different parts of the search string are connected by the operator "AND" in order to improve result completeness. This results in the search string (("Data Sources" OR "ETL" OR "OLAP" OR "Analytics" AND ("Data Warehouse" OR "Data Warehousing") AND ("Semantic Web" OR "Linked Data"))).

Data sources

The search included important journals and related conferences with research topics related to data warehousing and the Semantic Web. The search was restricted to studies published between the years 2002 and 2017 in order to achieve wide coverage of the study area. This is due to the fact that several important problems in data storage creation, which were identified years ago, are still considered to be unresolved problems (Nguyen, Tjoa, & Trujillo, 2005).

The initial step was conducting a search using the terms previously described in Section 3.2, via digital library search engines. Publications obtained from ScienceDirect, SCOPUS, IEEE Xplore, ACM Digital Library, and Spring Tools were considered. The second step was the search of international review journals with articles considered relevant, which were published by IEEE, ACM, Elsevier, and Springer, as they are considered high-level publication editors (McGregor, 2002). Conference proceedings were also searched. Where a conference displayed its proceedings on a published website, this was also accessed. When the proceedings were not available on a conference website, a search was conducted via the DBLP Computer Science Bibliography. The search for conference proceedings and journals produced many results that had already been obtained through the digital library search. In this case, the last results were discarded and only the first results were considered, given that these had already been included in the final list. After the search was executed for conferences and journals using digital libraries and proceedings, it became possible to identify which known publications—commonly referenced by other studies in this area, such as technical reports and theses—had not been included in the resulting final list. It was therefore decided to include these as gray literature entries. Gray literature is the term used to describe materials not commercially published or not indexed by the main databases.

TABLE 2 Table of search terms

No		Terms
1		Data sources, ETL, OLAP, analytics
2		Data warehouse, data warehousing
3		Semantic Web, linked data

3.2.2 | Screening of papers

The screening of papers involved selecting studies to analyze, and data analysis and synthesis.

Study selection

The set of search strings was applied to the search engines, specifically in those mentioned in the previous section. The criteria for inclusion and exclusion were used to filter studies that were not relevant for answering the RQs. Inclusion criteria were used to select all the studies during the search stage. Afterward, the criteria of exclusion were mainly applied to the titles of studies, and then to the summaries and conclusions. Regarding the inclusion criteria, the studies were only considered if they included:

- *DW development approaches that included semantic aspects.* The summary of the study explicitly mentioned the term Semantic Web or semantics in the context of DWs. From the summary, the reviewer was able to deduce that the focus of the research study was on using a Semantic Web technology at some stage of DW development.

Studies were excluded if they met the following criteria:

- *Research focus unrelated to DWs.* The article was outside the field of DWs.
- *Research focus related to DWs but insufficient information regarding the Semantic Web.* The Semantic Web was not part of the contribution to the article and its terms were only mentioned in the introductory sentence of the summary.
- *Research focus related to DWs but not related to the Semantic Web or linked data.* The article was within the field of DWs, but the Semantic Web or LOD was not a part of the contribution to the article.
- *Duplicate studies.* When the same study was published in different articles, only the most recent was included.
- *The study had already been included from another source.*
- *The article was in a language other than English.* All studies not written in English were filtered.
- *Technical reports and theses.* Given that they neither ensure an in-depth peer review, nor are widely validated by the scientific community, technical reports, and theses were excluded. In the case of theses, it was assumed that, if they offered some contributions, these originated from other publications, explaining why they were excluded from this study.

The selection of studies involved a process of analysis composed of three filters which were intended to select the most appropriate results, as the probability of retrieving inappropriate studies would have been high. Figure 2 briefly describes what was considered in each filter. In addition, the figure presents the number of written pieces that were obtained after the application of each filter.
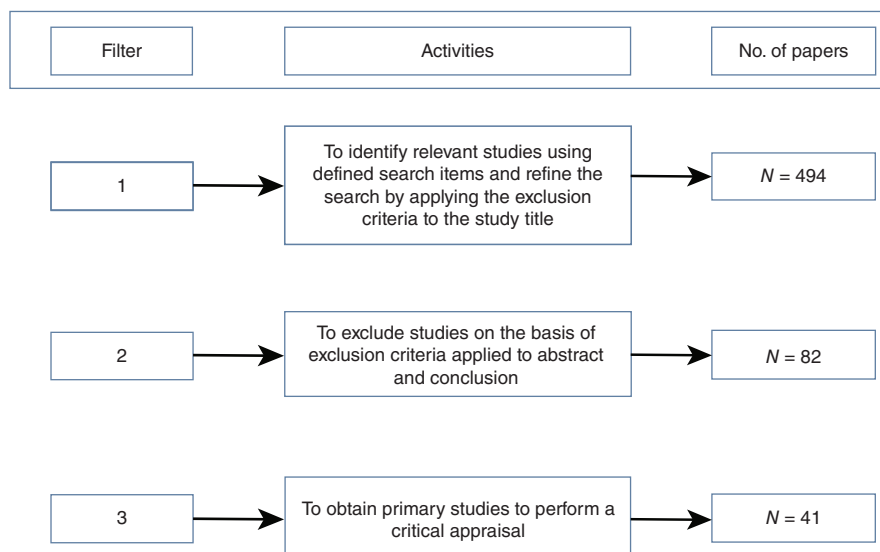
Filter	Activities	No. of papers
1	To identify relevant studies using defined search items and refine the search by applying the exclusion criteria to the study title	N = 494
2	To exclude studies on the basis of exclusion criteria applied to abstract and conclusion	N = 82
3	To obtain primary studies to perform a critical appraisal	N = 41

FIGURE 2 Stages of the article selection process

Data extraction

The method for data extraction was designed to extract all the information needed to answer the RQs. The following information was extracted from each article: *authors; source; conference/journal; publication year; summary; a brief opinion regarding its strengths and weaknesses; and the study's objectives.* It was decided that when several studies were reported in the same paper, each relevant study would be treated separately. However, this situation did not occur.

3.3 | Third phase: outcomes

This section describes the classification framework (or schema) used and the results of data extraction. Once the framework was defined, the relevant studies were ordered according to the framework. The output of this stage is the study map, which is presented at the end of this section together with the discussion.

3.3.1 | Classification schema

This study followed the systematic process shown in Figure 3. We used the idea proposed by Petersen et al. (2008) in order to categorize studies in facets. In our case, we defined two facets. One facet examined the type of research, and the other arranged the topic in terms of the RQs.

The search for key concepts is one way of reducing the time necessary to develop the classification schema, and ensure that the schema considers all articles. This search is conducted in two steps. First, the reviewers read the article summaries. Then, they look for key words and concepts that reflect the contribution of the article, and subsequently identify the research context. Once the above has been completed, a set of key words are obtained from different summaries, which are combined in order to achieve a high level of understanding with respect to the nature and contribution of the article, in addition to its reported use of Semantic Web technologies.

In this study, three aspects of interest were considered for each article: (a) the structure of the objective topic (i.e., the use of Semantic Web technologies in DW development); (b) the names of the Semantic Web technologies used, which were derived from the list of key words, and (c) the stage of DW development at which the said technology was used. Next, a search was conducted for evidence of improvement of the DW development process, which was expressed in terms of evaluations and/or validations of the results. For the first facet related to the type of research, analyzed studies were categorized according to the classification of research proposed by Wieringa, Maiden, Mead, and Rolland (2005), which was used with the intention of determining the type of research proposed. This classification identified six types of articles, defined as follows:

1. *Research evaluation.* Techniques are implemented in practice and an evaluation is conducted. This means that the way in which the technique is implemented (implementation of the solution) is presented, as are the consequences of the implementation in terms of its benefits and inconveniences (evaluation of the implementation). This also includes identifying problems within industry.
2. *Proposed solution.* A solution is proposed for the problem. The solution could be a new technique, or the extension of an already existing one. The potential benefits and the applicability of the solution are shown through a small example or a short, but effective argument.
3. *Research validation.* The techniques researched are recent and have still not been implemented in practice. The techniques used are, for example, experiments or, in other words, work performed in a laboratory.
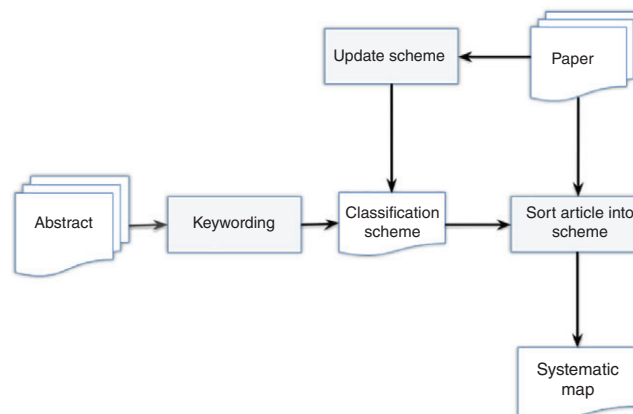

FIGURE 3 Creation of the classification schema

4. *Philosophical study.* These articles establish a new way of analyzing current topics of interest by structuring the area in a taxonomic way or through a conceptual framework.

5. *Opinion articles.* These articles express the personal opinion of someone with respect to whether a certain technique is good or bad, or in regards to how it should be performed. These opinions are not based on related studies or research methodologies.

6. *Personal experience study.* Articles based on personal experience that explains what and how something was done in practice.

3.4 | Results

In this section, we present the elements that were found through the analysis of the selected articles with the intention of answering the RQs. The evidence obtained from the data extraction process is highlighted, as are the particularities found. Prior to the above, a classification of the collected studies analyzed is included. The final list of primary studies used in this study is shown in Tables 5 and 6.

3.4.1 | Classification of the studies

A classification of the obtained studies is presented below. The classification process was intended to determine relevant aspects of the set of collected studies, particularly the primary studies that were ultimately selected. First, the distribution by data sources (e.g., journal, conference, book chapters, and others) is given. Next, the distribution of articles according to the place they were published is shown. Finally, the distribution of articles by publication year is presented. At this point, it is worth presenting a more complete view of the distribution of publications in order to better characterize what has been published regarding the interrelationship between DWs and the Semantic Web. Significant numbers of studies were published in journals, followed by significant numbers of publications that were not submitted for peer review, (e.g., technical reports and opinion articles) though they can still be found in some Web search engines. The above shows that attempt to address the topic have been based on the opinions of experts, or on studies that require more development. Indeed, one can easily find a plethora of writings authored by consultants, technological companies, and postgraduate students.[3] Given what was stipulated in Section 3.2.2 as exclusion criteria—that a research article must be published by high-level editors (McGregor, 2002)—such studies were excluded from the final list, as is illustrated in Figure 4. This presents a complete distribution of the studies and their place of publication. All of the publications initially found, along with the results of applying the defined filters, were considered.

In terms of publication year, the analyzed studies covered a range of years beginning with the year 2000 up to the year 2017. The distribution of primary studies by publication year is illustrated in Figure 5. According to the figure, the publications analyzed first began to take into account the relationship between DWs and the Semantic Web in the year 2003. It was also evident that the area of the Semantic Web was of considerable interest to researchers in 2009. Interest in the Semantic Web began in 2007 and peaked in 2009. Since that year, interest in the Semantic Web has decreased substantially. However, interest began to increase again in 2012, and subsequently declined between 2014 and 2017. In 2017, zero publications were
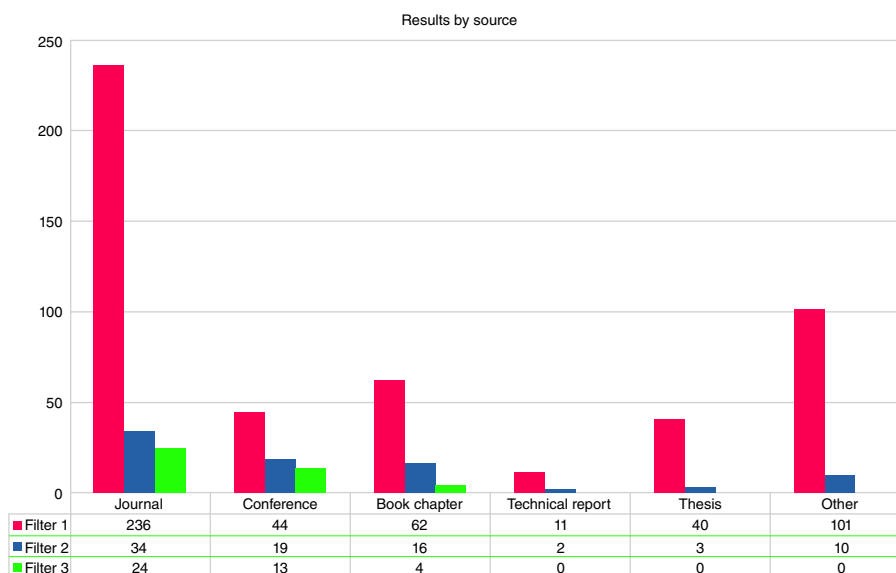

Results by source

	Journal	Conference	Book chapter	Technical report	Thesis	Other
Filter 1	236	44	62	11	40	101
Filter 2	34	19	16	2	3	10
Filter 3	24	13	4	0	0	0

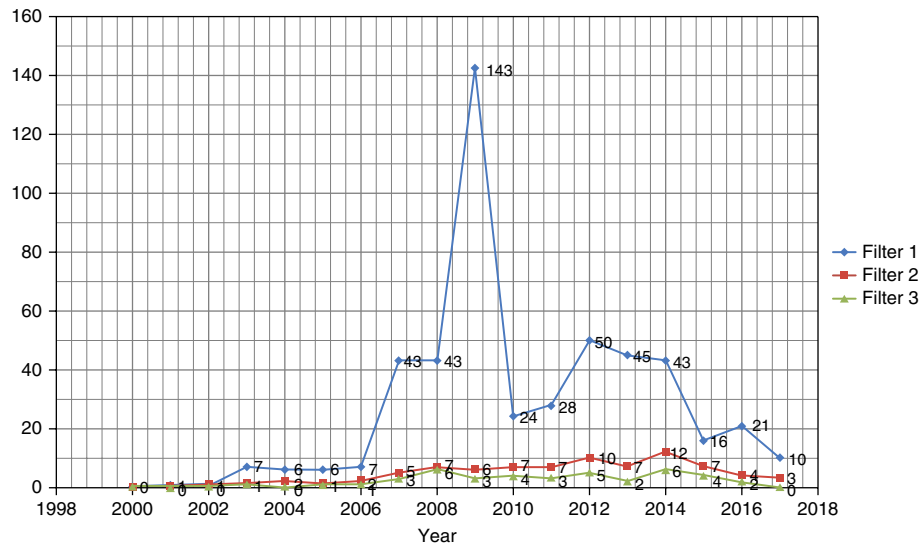**FIGURE 4** Place of study publication for all filters

FIGURE 5　Primary source distribution by publication year

presented showing relations between the process of building a DW and the Semantic Web. This is likely to be due to the fact that there is a growing migration of interest from DWs to LOD. In 2016, there was some research work showing the relationship between DWs and LOD. For instance, Ravat et al. (2016). The trend of DW development proposals employing Semantic Web technologies in the business domain has declined. Considering that the increase in the number of publications is an indicator of interest within the scientific community regarding a particular topic, this result shows that the interest of the scientific community has probably shifted to other related areas such as: LOD and Big Data, among others. In fact, LOD have become one of the most important sources of information, allowing the enhancement of business analyses, based on warehoused data, with external data. However, DWs do not directly cooperate with LOD datasets because of the differences between data models. Moreover, in the business domain, companies need to overcome the inevitable tension between the value they traditionally assign to proprietary data, and the value of opening up to be able to adopt LOD for sophisticated data analysis. Thus, an important challenge is to determine how effective a particular solution for building a DW, based on Semantic Web technologies has been, as well as evaluating its impact on the industry, if it has been implemented.

3.4.2 | Research questions

RQ1—where were the research studies published?

The 41 publications analyzed and published between the years 2002 and 2017, were distributed as follows: 24 were published in journals, 13 were published in conference proceedings, and four were published as book chapters.

Figure 6 presents a map of the distribution of publication place and year. The results revealed that the highest number of studies was published in journals, with the year 2015 having the highest number of all.

RQ2—what Semantic Web technologies are being used in DW development?

The analysis of the primary studies found two main types of Semantic Web technologies that were used in data storage development: (a) ontologies and vocabularies and (b) languages for representing ontologies.

In the primary studies examined, the individual and combined use of these Semantic Web technologies were reported. Some studies refer only to ontologies as a concept and others used ontologies along with a language such as RDF or OWL. The attention of most researchers was concentrated on the use of ontologies, and subsequently on the use of the OWL together with the ontologies. The primary studies that used ontologies included: Priebe and Pernul (2003), Sell et al. (2005), Skoutas and Simitsis (2006), Skoutas and Simitsis (2007), Niemi et al. (2007), Nazri et al. (2008), Salguero et al. (2008a), Sell et al. (2008), Diamantini and Potena (2008), Salguero et al. (2008b), Spahn et al. (2008), Skoutas et al. (2009), Niinimäki and Niemi (2009), Nebot et al. (2009), Simitsis et al. (2010), Romero and Abelló (2010), Khouri and Ladjel (2010), Jiang et al. (2010), Romero et al. (2011), Bergamaschi et al. (2011), Martin et al. (2011), Nebot and Berlanga (2012), Selma et al. (2012), Sell et al. (2012), Thenmozhi and Vivekanandan (2012), Thenmozhi and Vivekanandan (2013), Ali et al. (2013), Khouri et al. (2014), Elamin and Feki (2014), Talebzadeh et al. (2014), Ramasamy and Palanivel (2014), Samuel (2014), El Sarraj et al. (2014), Di et al. (2015), Abelló et al. (2015), Lamolle et al. (2015), and Steiner et al. (2015).
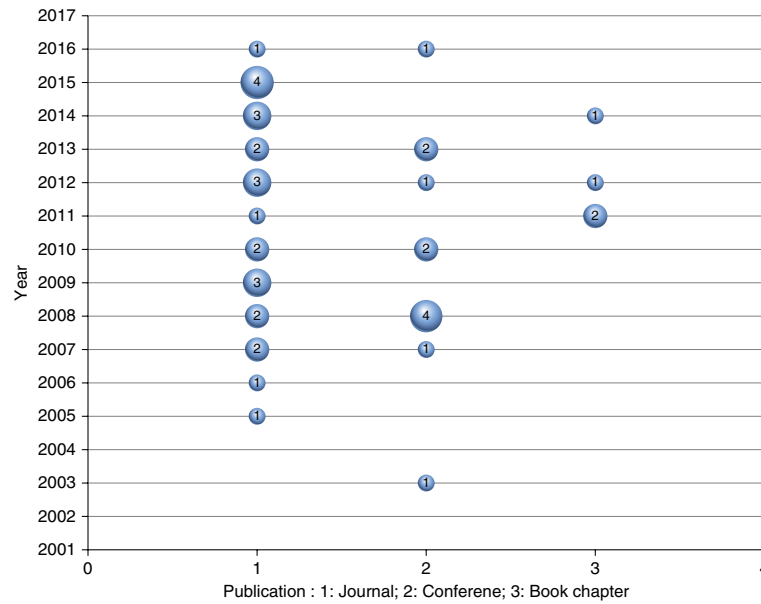
FIGURE 6   Distribution map of primary studies by publication venue and year

The main objective of the analyzed studies was to map data through an ontology in order to manage unstructured, or slightly structured data. Some approaches assumed that an ontology is easily constructed, in the event that it was not provided in advance. However, in many cases, finding an appropriate ontology for a specific domain is not a trivial task.

RQ3—during what stages in the DW development process are these technologies being used?

The analysis of the primary studies revealed four stages of DW development: (a) data acquisition; (b) transformation and loading; (c) data integration; and (d) data Access that were mentioned in the primary studies, as illustrated in Figure 7. All of the primary studies considered at least one stage of DW development. From the studies, two approaches were identified. One focused on the automation of MD design based on the use of ontologies. For example, Niinimäki and Niemi (2009) used ontologies to populate OLAP cubes. The other approach examined methods that focused on analyzing large amounts of Semantic Web data. For instance, Nebot and Berlanga (2012) proposed a semiautomated method for the extraction of semantic data, and for storing it in a MD database, for analysis using traditional OLAP techniques. Figure 7 shows that the majority of primary studies are concentrated in the stage of transformation and loading, and in the use of ontologies, rather than in the use of some language such as RDF or OWL.
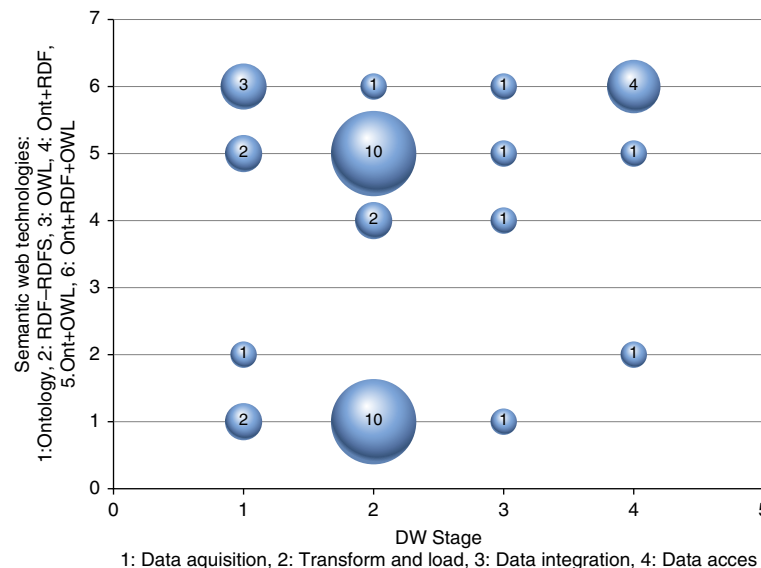

FIGURE 7   Visualization of a systematic map relating the stages of DW development to Semantic Web technologies using a bubble chart

Studies that focused on the stage of transformation and loading and the use of ontologies included: Skoutas and Simitsis (2006), Skoutas and Simitsis (2007), Nazri et al. (2008), Salguero et al. (2008a), Sell et al. (2008), Salguero et al. (2008b), Spahn et al. (2008), Skoutas et al. (2009), Niinimäki and Niemi (2009), Simitsis et al. (2010), Jiang et al. (2010), Romero et al. (2011), Bergamaschi et al. (2011), Martin et al. (2011), Thenmozhi and Vivekanandan (2012), Ali et al. (2013), Elamin and Feki (2014), Ramasamy and Palanivel (2014), Samuel (2014), Di et al. (2015), Steiner et al. (2015), Ravat et al. (2016), and Nebot and Berlanga (2016).

The following primary studies focused on the stage of transformation and loading, the use of ontologies and some language, such as RDF or OWL: Skoutas and Simitsis (2006), Skoutas and Simitsis (2007), Nazri et al. (2008), Salguero et al. (2008a), Salguero et al. (2008b), Niinimäki and Niemi (2009), Simitsis et al. (2010), Romero et al. (2011), Thenmozhi and Vivekanandan (2012) Steiner et al. (2015), Ravat et al. (2016), and Nebot and Berlanga (2016).

RQ4—what type of research study was conducted?

The classification of research approaches proposed by Wieringa et al. (2005) was used to relate the type of research approach of a particular study to Semantic Web technologies. The greatest number of articles concentrated on "solution proposal" approaches, followed by "research evaluation." The systematic search did not find any articles that documented experiences with respect to industrial applications of Semantic Web technologies. Thus, this area represents an important challenge for future research. Examples of primary studies that show the relationship between solution proposals and the use of ontologies included: Priebe and Pernul (2003), Sell et al. (2005), Skoutas and Simitsis (2007), Niemi et al. (2007), Salguero et al. (2008a), Sell et al. (2008), Diamantini and Potena (2008), Salguero et al. (2008b), Spahn et al. (2008), Skoutas et al. (2009), Niinimäki and Niemi (2009), Simitsis et al. (2010), Khouri and Ladjel (2010), Jiang et al. (2010), Bergamaschi et al. (2011), Martin et al. (2011), Selma et al. (2012), Sell et al. (2012), Thenmozhi and Vivekanandan (2012), Thenmozhi and Vivekanandan (2013), Elamin and Feki (2014), Talebzadeh et al. (2014), Ramasamy and Palanivel (2014), Samuel (2014), El Sarraj et al. (2014), and Di et al. (2015). Primary studies that showed the relationship between solution proposals, the use of ontologies and a language included: Niemi et al. (2007), Khouri and Ladjel (2010), Selma et al. (2012), Sell et al. (2012), Thenmozhi and Vivekanandan (2013) Talebzadeh et al. (2014), Ravat et al. (2016), and Nebot and Berlanga (2016). The results of the classification are presented in Figure 8, which presents a summary of the type of article alongside the use of Semantic Web technologies.

RQ5—is there evidence of improvement due to the use of Semantic Web technology?

In order to analyze the evidence of improvement presented in primary studies, the classification framework proposed by Shaw (2003) was used. This framework classifies: (a) the type of research results and (b) the validation type found in the studies. The classification by type of research results, as well as validation type, allowed the evidence of improvement resulting from the use of Semantic Web technologies to be evaluated objectively. On the one side, it can be established that the type of result
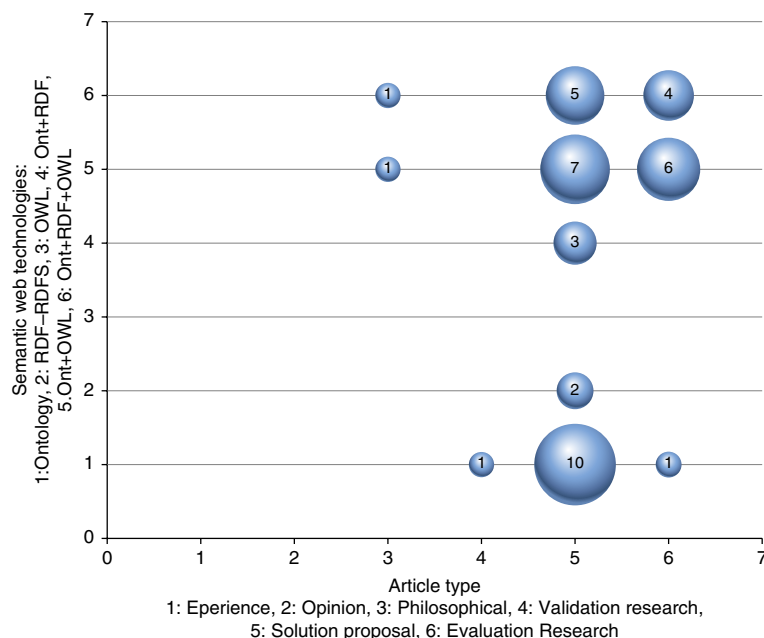

FIGURE 8 Visualization of a systematic map relating article type to the use of Semantic Web technologies using a bubble chart

TABLE 3 Types of research results

ID	Type of results	Example
RI-1	Procedure or technique	New or better ways of realizing a given task, such as design, implementation, maintenance, measurement, evaluation, or selection of alternatives.
RI-2	Qualitative or descriptive model	Taxonomic structure for a problem area; architectural style, framework or design pattern; analysis of an informal domain, checklist or well-argued generalizations.
RI-3	Empirical model	Predictive, empirical model based on observed data.
RI-4	Analytical model	Structural model that permit formal analysis or automated manipulation.
RI-5	Notation or tool	Software tool that implements a technique or a formal language to support a technique or model.
RI-6	Specific solution, prototype, response or judgment	Solution to a problem that shows the application of some principles—such as design, prototype or complete implementation and careful analysis of a system or its development, resulting in a specific analysis, evaluation, or comparison.
RI-7	Report	Interesting observations. Golden rules. It is not sufficiently general or systematic to be considered a descriptive model.

Note. Shaw (2003).

(ranging from the most formal to the least formal), and the validation type (ranging from a formally validated result to a simple declaration of improvements), both constitute evidence of a proposal's value. For example, if the proposal was a report in which only declarations of improvements were made, it was not considered to contain evidence of improvement according to this framework. In contrast, if the proposal presented was a rigorous analysis of a formal or empirical model, its results were considered to show valid evidence of improvement. In addition to the above, the text of each study when analyzed was searched for some explicit mention of the results indicating an improvement due to the use of Semantic Web technologies. Research results may include procedures or techniques for development or analysis; models that generalize examples, specific tools, or solutions for particular systems. Table 3 describes the type of research results that were reported in the studies analyzed, which was classified according to the framework proposed by Shaw (2003).

For the type of result, the greatest concentration of primary studies was found in the RI-6 category (Specific Solution). This is shown in Figure 9. This means that the majority of examined proposals responded to specific solutions, or implementations of a prototype that used some type of Semantic Web technology. On the other hand, the research studies also used different classes of evidence to support their results. Thus, it is necessary to select a form of validation that is appropriate for the type of result and the methods used to obtain the result. For example, a formal model should be supported by rigorous testing and not by examples of use. Table 4 presents the possible evidence (validation) types found using the framework proposed by Shaw (2003).

Among the primary studies analyzed, the most common validation types were RV-2 (Evaluation) and RV-4 (Example), as shown in Figure 10. In the case of validation type RV-2 (Evaluation), the authors created a proposal and then evaluated the obtained results. For validation type RV-4 (Example), the authors presented a description of how a web technology is used and then applied it to an example. The results of the systematic map on this topic are illustrated in Figure 11.

Regarding the systematic map of the relationship between the type of result and validation type, the majority of primary studies focused on specific proposals, or solution prototypes in which well-argued frameworks and generalizations were proposed. Such studies were considered evaluations of the proposal. Studies that proposed specific solutions and evaluated their proposals included: Skoutas and Simitsis (2007), Spahn et al. (2008), Niinimäki and Niemi (2009), Romero and Abelló
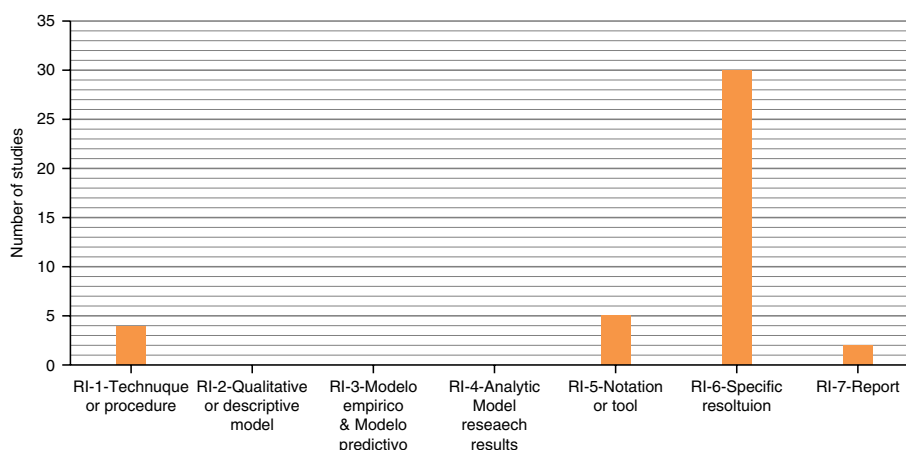

FIGURE 9 Type of research results

TABLE 4   Validation type

ID	Validation type	Example
RV-1	Analysis	Provides a rigorous analysis of a formal model, empirical model or controlled experiment with rigorous testing, data in controlled situations, or experiments with abundant and significant statistical results, respectively.
RV-2	Evaluation	Includes studies of feasibility, pilot projects for descriptive, qualitative, or empirical models.
RV-3	Experience	The results have been used in real examples and the evidence of their effectiveness, usefulness and accuracy is given by qualitative, empirical models, tools, or techniques through narrative descriptions, statistical data, or predictions that fit the current data.
RV-4	Example	Example of a given technique or procedure is used. A fragment of reality (simplified example of reality) can be convincing especially if it is accompanied by an explanation of why this simplified example retains the essence of the problem being solved. Small examples or examples found in books often fail to provide convincing validations (except for standard examples used as model problems in the area).
RV-5	Persuasion	Validation only by persuasion makes it difficult to consider an article as research study.
RV-6	Brazen assertion	No serious attempt is made to evaluate the results.

Note. Shaw (2003).

(2010), Jiang et al. (2010), Romero et al. (2011), Selma et al. (2012), Thenmozhi and Vivekanandan (2012), Etcheverry and Vaisman (2012), Samuel (2014), Lamolle et al. (2015), Nebot and Berlanga (2016), and Ravat et al. (2016). Studies that proposed specific solutions that were evaluated using an example included: Skoutas and Simitsis (2006), Niemi et al. (2007), Nazri et al. (2008), Salguero et al. (2008a), Sell et al. (2008), Salguero et al. (2008b), Bergamaschi et al. (2011), Martin et al. (2011), Sell et al. (2012), Thenmozhi and Vivekanandan (2013), Ali et al. (2013), Khouri et al. (2014), Elamin and Feki (2014), Talebzadeh et al. (2014), El Sarraj et al. (2014), Di et al. (2015), and Steiner et al. (2015). The analyzed studies did not show evidence of comparisons being made to other similar proposals. As a result, it is not possible to determine if the results obtained were better in comparison to the results covered by similar proposals.

4 | ANALYSIS OF THE RESULTS AND DISCUSSION

Of the results described in Section 3.3, the following elements stood out:

1. The most commonly used Semantic Web technologies to build DWs in the business domain were ontologies and markup languages, such as RDF or OWL. However, the use of ontologies, and the markup languages, has been limited to representing terms and improving model equivalence as well as specifying registry structures for databases, according to the analysis of the studies. There were no uses that showed greater complexity than the representation of conceptualizations that describe aspects of DW domain information. Such an approach would enable, for example, the creation of recommendations that support the DW development process.

2. A total of 58% of the primary studies analyzed were published in journals, which indicate a higher degree of maturity among these proposals. This can be explained by the fact that both DW development and the Semantic Web are areas that
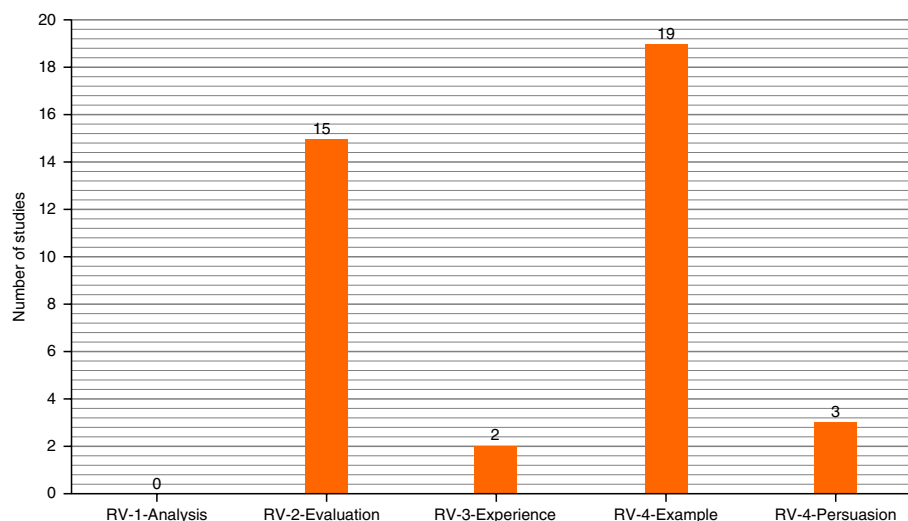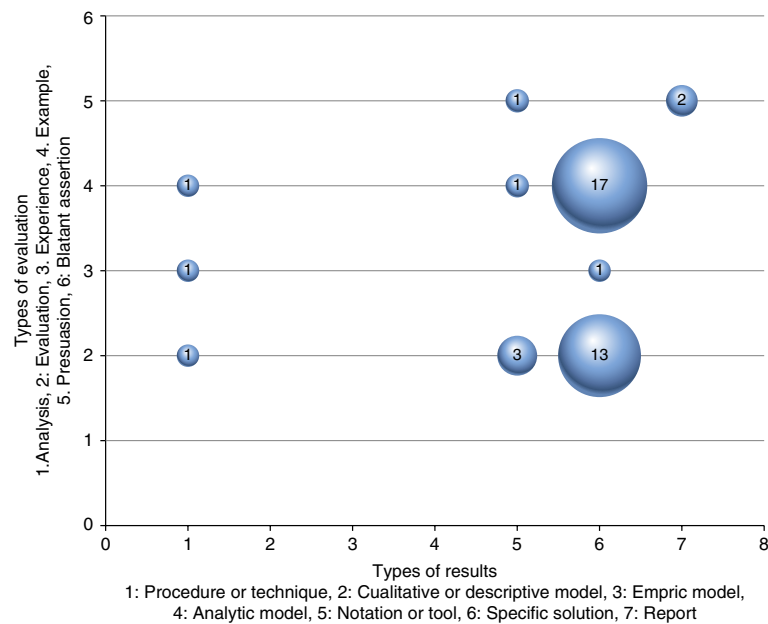

FIGURE 10   Validation type

FIGURE 11 Visualization of a systematic map relating the type of result to the validation type using a bubble chart

have shown important advances and growth in their maturity levels over the time period examined. Thus, the combination of both areas in one proposal assumes a higher maturity level.

3. Of the primary studies analyzed, there is a lack of sufficient evidence showing that the results of the studies had been transferred successfully to industry. The majority of studies analyzed proposed solutions intended to improve certain aspects of DW development. However, concrete support elements were not derived from these studies, such as standardized models or publicly available tools.

4. The results indicate an increasing lack of interest in applying Semantic Web tools to DW development. Unfortunately, there is also no evidence of mature solutions being transferred to industry. There are two possible explanations for the above: (a) the topic stopped being of interest to researchers or (b) new topics of interest are attracting the attention of researchers. We believe that the second option is the more viable, as the emergence of new topics such as: LOD, Cloud Computing, Big Data, and Cognitive Computing, among others, have caused the focus of researchers to become centered on other topics of current interest, such as: security, cloud computing, efficient processing and new web architectures, and so on. Since LOD have attracted the interest of researchers, it seems the Semantic Web community continues to develop the field (i.e., tools and technologies) by following the linked data approach.

5. Considering that some Semantic Web technologies were present (e.g., ontologies and markup languages) in the analyzed studies, it is clear that there is a need for greater development of the semantic aspects of DW development. The creation of metamodels, the inclusion of semantic rules that support DW development and the intensive use of natural language, are fundamental to improving aspects of DW creation. A combination of BI technologies with the Semantic Web will both enhance BI analysis with web data, and allow the analyzing of Semantic Web data through BI tools. In fact, business can get benefits by using Semantic Web technologies for data analytics, data integration, and knowledge discovery. However, combining BI with the Semantic Web is not a trivial task due to the scalability, complexity, and heterogeneity of Semantic Web data.

Currently, a DW is widely used as a consistent and integrated data repository in BI systems. According to Ravat et al. (2016), additional information coming from outside an organization, mostly found on the Web, should also be included in analyses to provide multiple perspectives for decision-makers, within today's highly competitive business context. LOD have been introduced as a promising paradigm, based on the Semantic Web, for opening up data which integrates published datasets, because it facilitates data integration on the Web (Bizer et al., 2009). Similar to data warehousing approaches, LOD can be prepared to enable sophisticated data analysis. In 2015, the report entitled: *Gartner's Hype Cycle for Enterprise Information Management, 2015*[4] has put linked data into the trough of disillusionment stage. According to Garner's Hype Cycle Research Methodology, the phase *Trough of Disillusionment* is one of five key phases of a technology's life cycle, which is defined as follows: "Trough of Disillusionment: Interest dies out as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters."[5] Some issues that impede its adoption in the business domain are as follows:

1. *Cost-benefits trade off.* Businesses are not clear on how to get benefits, value, and extra earnings from adopting LOD. One problem to overcome is the tension between the value they traditionally assign to proprietary data, and the value of opening up.
2. *Reliability and data quality.* Over the last few years, linked data have developed into a large number of datasets, with open access from several domains leading to the LOD cloud. Similar to other types of information such as structured data, linked data suffer from quality problems such as inconsistency, inaccuracy, being out-of-date, and incompleteness, which are frequent, and imply serious limitations to the full exploitation of such data.
3. *High-computational costs for queries.* To get information from the global Web of data, the query language SPARQL is used. The problem is that queries can require jumping from one server to another. Thus, joint operations can be more costly than in the relational data base.
4. *Complexity and heterogeneity of the technologies.* Currently, there are many different and complex semantic technologies such as frameworks, RDF, Turtle, Microdata, language query, and among others. New technologies will appear, and some others will survive. The challenge is, therefore, how to use and integrate them.

In 2017, Gartner's Hype Cycle for Analytics and BI[6] classified logical DW as a future useful technology for BI. Logical DW is defined as an architectural layer that sits at the top the usual DW store of persisting data. The logical DW complements the traditional core warehouse (and its primary function of a priori data aggregation, transformation, and persistence) with functions that fetch and transform data, in real time (or near to it), thereby instantiating nonpersistent data structures, as needed.

4.1 | Main findings of the study

Given the above, three important challenges can be identified regarding the use of Semantic Web technologies in DW development in the BI domain.

1. *New uses and applications based on ontologies.* Using ontologies to represent information aspects and not just terminological expressions, allows for greater capabilities to be included in applications by creating tools with greater capacities, such as advisers and intelligent assistance systems to be used in DW development.
2. *Measuring solution effectiveness and impact.* Another important challenge is in determining how effective Semantic Web-based technologies have been, and what their impact has been on industry. This is important, as the ultimate objective of research in this field should be aimed at advancing existing knowledge in order to be able to apply these advances to industry. Aspects of cost, performance, quality assurance, and DW design approaches have not been extensively considered.
3. *Inclusion of semantic aspects in addressing new paradigms.* Another important challenge is to consider semantic elements in remedying current difficulties in DW development and other technological challenges. For example, the minimal use of software tools for ETL opens a space for creating new tools, or improving existing ones with semantic aspects. In the same way, the emergence of Cloud Computing and Big Data involves assuming new considerations in DW design, which could be supported with tools or Semantic Web-based technologies.

4.2 | Review of industrial projects

This section reviews some industrial tools and projects within the broad topic of DWs and the Semantic Web. First, we review briefly some existing industrial tools, and then we focus on reviewing industrial projects.

Unfortunately, research on the intersection between DWs and the Semantic Web has not moved toward industrial tools (Abelló et al., 2015). Apart from software that provides functionality to derive linked data from existing databases, such as Virtuoso[7] (from OpenLink vendor) or Apache Jena[8], there are few industrial tools that fully consider the potential of the Semantic Web together with DWs. Interestingly, most of them are OLAP tools, such as TARGIT Decision Suite (Middelfart & Pedersen, 2011)[9] which uses extended semantics to extend the traditional MD model with semantic associations between measures and dimensions; or the OpenCube Toolkit (Kalampokis et al., 2014)[10] which provides facilities for analyzing RDF data as MD cubes. On the other hand, graph databases (such as GraphDB from OntoText[11]) have emerged as important tools for exploring how interlinking data could be useful in developing a new generation of DWs that unleashes the potential of unstructured data, and for inferring new knowledge. The graph paradigm differs from other database models that make it interesting for enterprises, since it naturally model real-world concepts and relations according to Bryce Merkl Sasaki[12]. Furthermore, the growing prevalence of graph databases in the industrial and marketing arenas will require considering this paradigm as a driving force for storing and analyzing data. Consequently, approaches that use graph databases in the development of DWs, and specifically, those considering enterprise knowledge graphs, are a must. In fact, a Forrester

Research survey[13], states that "51% of global data and analytics technology decision makers either are implementing, have already implemented, or are upgrading their graph databases." According to Yu Xu[14], who is the founder and CEO of Tiger-Graph, the world's first native parallel graph database, graph databases (a) are simple and provide more natural data modeling, (b) provide better, faster queries, and analytics, as well as (c) support flexibility for evolving data structures. These key benefits of graph databases might benefit the building of DWs.

Finally, we have to highlight DWs for linked data such as CM-Well (Bennett, Engelbrecht, & Landau, 2017) from Thomson-Reuters[15] that allows a company to manage large volumes of linked data and to warehouse its knowledge graph.

Importantly, some efforts have been made in research projects, for example, as funded by the European Union. We have conducted a search of projects on DWs and the Semantic Web in the search engine provided by (Community Research and Development Information Service, which is the European Commission's primary public repository and portal for disseminating information on all EU-funded research projects and their results)[16]. We have found the following projects:

- GeoKnow[17]: this project aims to make the Web a data source for geospatial knowledge, that is, moving from geographic information systems to a web of interlinked geographic knowledge systems. GeoKnow focuses on achieving the evolution of the web from a medium for information exchange, to a medium for (spatial) knowledge integration.
- SEEK[18]: this project aims to enrich semantically trajectory knowledge discovery, that is, taking advantage of data collected from personal devices such as mobile phones and other location-aware devices. SEEK focuses on extracting behavioral patterns through a knowledge discovery process, where positioning data collected from mobile devices are first transformed into semantically enriched trajectory data, and then stored in a DW and analyzed with OLAP operations that allow the summarization of the trajectory features.
- OpenCube[19]: this project focuses on processing RDF data as OLAP cubes, that is, MD data represented as RDF and structured according to the RDF Data Cube ontology.

5 | RESEARCH CHALLENGES

Designing a DW is a complex task that often forces designers to acquire wide knowledge of the domain, thus calling for a high level of expertise and so becoming a prone-to-fail task. Taking into consideration the results of our SM, in the following section, our findings and experience are combined with some proposed solutions for the issues that have been discovered in the SM. Importantly, based on our experience, we have detected a set of situations in which we believe that the use of some kind of Semantic Web technology will improve the design of DWs.

The main focus of research using Semantic Web technologies in the development of DWs is in the development of ETL processes, as stated in our SM; however, there are other important challenges in which Semantic Web technologies may formalize the use of specific-domain knowledge for data warehousing, namely:

- Reusing expert knowledge from different domains.
- Enriching specific metadata by completing definitions and annotating their semantics.
- Enabling metadata interchange among repositories.
- Populating the designed databases from public data sources.
- Empowering data integration, and analysis.
- Automatizing reasoning on metadata.
- Validating data instances and models.

Other more elaborated proposals for using semantic Web Technologies in the development of DWs are described in the following points:

- Reconciling information requirements and data sources. Due to the special idiosyncrasy of DWs, it should not only be the information requirements of decision makers that are considered for MD design, but also a second driving force: data sources, which need to be reconciled with those information requirements. Requirements analysis for MD modeling aims to elicit the information needs of decision makers in order to deploy a DW that will satisfy their expectations. Furthermore, internal and external data sources should be taken into account and matched with information requirements, since these data will populate the DW. To do so, Semantic Web technologies can be used to match formally the data sources with information requirements in the early stages of development, in order to make the not-so-obvious correspondence between information requirements and their counterparts in data sources. However, according to our SM, these topics have not been investigated so far, and more research is required.

- Incompleteness in MD models. Semantic Web technologies may enrich a MD model in aspects that have not been taken into account during requirement analysis or reconciliation of data sources. For example, "public" LOD sources can be used to complete the unsupported elements within a dimension hierarchy. Moreover, taxonomies such as the Computing Classification System (CCS[20]) can be also used as additional data sources for designing MD models. CSS provides a classification for topics on computations that can be, and usually is, used for classifying computer literature. For instance, there is a taxon for Information Systems (H) where, for example, Database Management (H.2) topics are classified. This taxonomy can also be used for completing aggregation hierarchies, whenever a DW requires a dimension on computer-related topics.
- Data types of measures of DWs. Measures of analysis described in DWs by MD models in fact do not aid either designers or end-users to gather important details, such as their units or scale. Therefore, more research in the semantic description of measures in MD models (addition of units, magnitudes, scales, etc.) should be conducted in order to aid designers to annotate the identified measures.
- Semantic-aware summarizability. DWs use statistic functions to aggregate measures to provide data analysis. The application of aggregation functions depends on which sort of measure and aggregation criteria are involved according to the compatibility given by some additivity constraints. In particular, type compatibility states the compatibility of category attributes (aggregation levels), summary attributes (measures), and statistical functions (aggregation functions). Some taxonomies of measure have been proposed to deal with this compatibility, such as those in Lenz and Shoshani (1997) where measure can be classified as either a "flow," a "stock," or a "value-per-unit." However, the proposed taxonomies for dealing with summarizability constraints lack a wider consideration of semantics. This means that designers do not have the necessary knowledge for working out what really is a given piece of information and thus, which constraints should be held. More research is required on this topic, to gain knowledge for a better understanding of the very nature of summarizability constraints, and of being able to understand why, when, and how to hold them.
- Semantically traceable models. Mapping from MD models to relational models is accomplished by structural matching. Since structure and semantics are closely related, whenever structure is manually changed, the source semantics are lost. Automatic model transformations for the deployment of DWs are managed with the manual transformations that software

TABLE 5 Summary of analyzed studies

ID	Study title	Year	Source
P01	Toward integrative enterprise knowledge portals	2003	Priebe and Pernul (2003)
P02	Adding semantics to BI	2005	Sell, Cabral, Motta, Domingue, and Pacheco (2005)
P03	Designing ETL processes using Semantic Web technologies	2006	Skoutas and Simitsis (2006)
P04	The process of metadata modeling in industrial data warehouse environments	2007	Jossen and Dittrich (2007)
P05	Ontology-based conceptual design of ETL processes for both structured and semistructured data	2007	Skoutas and Simitsis (2007)
P06	Ontologies with Semantic Web/grid in data integration for olap	2007	Niemi, Toivonen, Niinimaki, and Nummenmaa (2007)
P07	Automatic data warehouse conceptual design	2008	Nazri, Noah, and Hamid (2008)
P08	Ontology-based framework for data integration	2008	Salguero, Araque, and Delgado (2008a)
P09	SBI: a semantic framework to support BI	2008	Sell et al. (2008)
P10	Diamantini adn Potena	2008	Diamantini and Potena (2008)
P11	Spatiotemporal ontology-based model for data warehousing	2008	Salguero, Araque, and Delgado (2008b)
P12	Supporting BI by providing ontology-based end-user information self-service	2008	Spahn, Kleb, Grimm, and Scheidl (2008)
P13	Ontology-driven conceptual design of ETL processes using graph transformations	2009	Skoutas, Simitsis, and Sellis (2009)
P14	An ETL process for OLAP using RDF/OWL ontologies	2009	Niinimäki and Niemi (2009)
P15	MD integrated ontologies: a framework for designing semantic data warehouses (DWs)	2009	Nebot, Berlanga, Pérez, Aramburu, and Pedersen (2009)
P16	Representation of conceptual ETL designs in natural language using Semantic Web technology	2019	Simitsis, Skoutas, and Castellanos (2010)
P17	A framework for MD design of DWs from ontologies Romero and Abelló (2010)	2010	Simitsis et al. (2010)
P18	A methodology and tool for conceptual designing a DW from ontology-based sources	2010	Khouri and Ladjel (2010)
P19	A domain ontology approach in the ETL process of data warehousing	2010	Jiang, Cai, and Xu (2010)
20	GEM: requirement-driven generation of ETL and MD conceptual designs	2011	Romero, Simitsis, and Abelló (2011)
P21	A semantic approach to ETL technologies	2011	Bergamaschi, Guerra, Orsini, Sartori, and Vincini (2011)
P22	A framework for BI application using ontological classification	2011	Martin, Maladhy, and Venkatesan (2011)
P23	Building DWs with Semantic Web data	2012	Nebot and Berlanga (2012)

engineers make, that is, modifying the current label of a modeling element, removing some data, relating some elements, and so on. Traceability is the mechanism for propagating changes in a transformation chain, taking into account only the changes and not entire models. It solves the problem whenever automatic transformations are involved. However, modifying the label implies changing its semantics, or even worse, losing its meaning because of the misuse of semantics. This fact is easily shown in data sources where a data item may be labeled in a cryptic form (s0702) from which its semantics cannot be inferred (the sales from 2007 to 2009). Therefore, more research on this topic will define novel approaches for using Semantic Web technologies that may solve this drawback in the development of DWs by having semantically annotated MD models.

6 | LIMITATIONS OF THE STUDY

Some concerns regarding the validity of our study are presented below:

- *Consideration of Synonyms.* The terms used in the search string could have synonyms. As a result, it is possible that the search passed over some studies. On occasion, authors may only mention the term Semantics, without alluding to the Semantic Web, or use some particular technology of the Semantic Web, without making it explicit.
- *Quality of the Evaluation.* The quality of the selection and evaluation of the studies, as well as the weighting factor used to quantify each, could have failed to appropriately represent the importance of the selected sets. In order to mitigate this problem, the quality attributes were grouped into subsets in order to facilitate future classifications and improve the selections.

7 | RELATED WORK

Several studies have focused on automating MD design using Semantic Web artifacts such as existing ontologies. Some surveys have been published focusing on the above issue. For instance, Abelló et al. (2015) go on to survey the use of Semantic Web technologies for data modeling and data provisioning, including semantic data annotation and semantic-aware ETL processes. A characterization of DW/OLAP environments is presented, followed by an introduction to the relevant Semantic Web foundation concepts. The study describes the relationship of MD models and Semantic Web technologies, including the relationship between MD models and Semantic Web formalisms. Wache et al. (2001) review the use on ontologies for the integration of heterogeneous information sources. They discuss how ontologies are used to support the integration task. Chakraborty,

TABLE 6 Continuation of the table Resumen

ID	Study title	Year	Source
P24	Ontology-based structured web data warehouses (DWs) for sustainable interoperability: requirement modeling, design methodology, and tool	2012	Selma et al. (2012)
P25	Adding semantics to business intelligence: toward a smarter generation of analytic tools	2012	Sell, da Silva, Ghisi, Todesco, and Napoli (2012)
P26	An ontology-based hybrid approach to derive MD schema for DW	2012	Thenmozhi and Vivekanandan (2012)
P27	QB4OLAP: a new vocabulary for OLAP cubes on the Semantic Web	2012	Etcheverry and Vaisman (2012)
P28	A tool for DW MD schema design using ontology	2013	Thenmozhi and Vivekanandan (2013)
P29	Using schema matching in data transformation for warehousing web data	2013	Ali, Abdelrahman, and Mohamed (2013)
P30	Requirements driven DW design: we can go further	2014	Khouri, Bellatreche, Jean, and Ait-Ameur (2014)
P31	Toward an ontology based approach for data warehousing	2014	Elamin and Feki (2014)
P32	Automated creating a data warehouse from unstructured semantic data	2014	Talebzadeh, Seyyedi, and Salajegheh (2014)
P33	Triplet dependency views of university data warehousing toward decision support system	2014	Ramasamy and Palanivel (2014)
P34	Toward a DW fed with web services	2014	Samuel (2014)
P35	An ontology-driven personalization approach for data warehouse exploitation	2014	El Sarraj, Espinasse, and Libourel (2014)
P36	Academic DW design using a hybrid methodology	2015	Di, Lefons, and Tangorra (2015)
P37	Using Semantic Web technologies for exploratory OLAP: a survey	2015	Abelló et al. (2015)x
P38	Incremental checking of master data management model based on contextual graphs	2015	Lamolle, Menet, and Le Duc (2015)
P39	Judgment and analysis rules for ontology-driven comparative data analysis in DWs	2015	Steiner, Neumayr, and Schrefl (2015)
P40	Statistically driven generation of MD analytical schemas from linked data	2016	Nebot and Berlanga (2016)
P41	Designing MD cubes from warehoused data and linked open data	2016	Ravat, Song, and Teste (2016)

TABLE 7 Summary and contributions of each analyzed study

ID	Summary	Contribution
P01	This article discusses integration aspects within enterprise knowledge portals and presents an OLAP-based approach for communicating the user context (revealing the user's information need) among portlets, utilizing Semantic Web technologies.	A prototype to evaluate the approach, demonstrating how OLAP can be integrated with information retrieval.
P02	This article proposes an approach, which aims at integrating business semantics into analytical tools by providing semantic descriptions of exploratory functionalities and available services.	An architecture for BI, which uses Semantic Web technologies based on IRS-III. In addition, a prototype tool based on this architecture.
P03	This article argues that ontologies constitute a suitable model for the purpose of establishing the appropriate mappings between the attributes of the data sources and the attributes of the data warehouse (DW) tables, which is critical in specifying the required transformations in an ETL workflow. The authors show how the usage of ontologies can enable a high degree of automation regarding the construction of an ETL design.	There are several contributions of This article: (a) construction of a domain vocabulary; (b) annotation of the data stores; (c) generation of a domain ontology; and (d) generation of conceptual ETL design.
P04	This article discusses the process of identifying metadata model requirements, defining a new metadata model, and finally implementing it in a metadata schema. This article describes the implementation of the metadata model based on the metadata standards resource description framework (RDF), and RDF schema (RDFS).	A new metadata model using RDF and RDFS for a DW environment that fulfills the needs of an industrial setting. In addition, the implementation of the model in a real-world scenario.
P05	This article proposes an ontology-based approach to facilitate the conceptual design of the integration layer of a DW. The proposed approach is based on the use of Semantic Web technologies to semantically annotate the data sources and the DW, so that mappings between them can be inferred, thereby resolving the issue of heterogeneity in data integration.	A graph-based representation, which is used as a conceptual model for the datastores. In addition, a suitable application ontology is created and used to annotate the data stores.
P06	This article presents an approach for data source integration in OLAP scenarios, based on RDF and other Semantic Web technologies.	An OWL/RDF ontology for OLAP data sources and OLAP cubes. In addition, they provide a prototype for testing the approach.

Padki, and Bansal (2017) present a survey of research in the area of data integration using ETL approaches including some semantic issues. The objective of this survey paper is to deliver a review of current approaches to data integration, various ETL tools that are already in use, linked data generation techniques, semantic-based approaches to ETL and data integration.

Although the above studies present a survey of research, they do not identify, appraise, or synthesize research evidence from individual studies based on a strict protocol in order to summarize their results.

TABLE 8 Continuation of the table summary (I)

ID	Summary	Contribution
P07	This article presents a methodology to develop a data warehouse (DW) (based on MD modeling) by integrating all three development approaches such as supply driven, goal driven, and demand driven.	An ontology embedded with specific knowledge domain. The ontology aims to identify the semantics of the domain. In addition, a prototype is developed.
P08	This article describes a framework which encompasses the entire data integration process. They deal with the problem of assisting in the process of designing information systems based on DW and the inclusion of new data sources, since it may involve re-designing the scheme of DW.	An ontology-based model to support the data integration process.
P09	This article presents a framework called SBI (semantic business intelligence) in which applied ontologies are used for the description of business rules and concepts in order to support semantic-analytical functionalities that extend traditional OLAP operations.	An analytical tool, called Extracta, which relies on SBI ontologies and modules.
P10	This article proposes a novel model for semantic annotation of a DW schema that takes into account domain ontologies as well as a mathematical ontology.	An ontology, which describes mathematical formulas underlying elements of the data cube schema, including the semantics of operators as the basis for OLAP analysis.
P11	This article describes a spatiotemporal extension of an ontology language which facilitates the generation of the scheme of a DW as well as the design of the processes which extract, transform, and load (ETL) the data from the sources in the DW according to the temporal characteristics of the data sources.	An ontolgy-based model to deal with the data sources schemes integration problem.
P12	This article presents an ontology-based architecture and end-user tool, enabling easy data access and query creation for business users. The approach is based on a semantic middleware integrating data from heterogeneous information systems and providing a comprehensible data model in the form of a business level ontology (BO).	An ontology-based architecture and an end-user tool: semantic query designer (SQD).
P13	This article proposes a customizable and extensible ontology-driven approach for the conceptual design of ETL processes.	A graph-based representation for the source and target data stores, and a method for devising flows of ETL operations by means of graph transformations.

TABLE 9 Continuation of the table summary (II)

ID	Summary	Contribution
P14	This article presents a method for on-demand construction of OLAP cubes based on relational algebra. The authors combine Semantic Web and OLAP, resulting in a method of creating a fully functional tool for data analysis.	A method and an ontology-based tool that works as a user interface for the system, from design to actual analysis.
P15	This article proposes the semantic data warehouse to be a repository of ontologies and semantically annotated data resources. In addition, an ontology-driven framework to design MD analysis models for semantic data warehouses is proposed.	An ontology-driven framework to design MD analysis models for semantic data warehouses.
P16	This article provides a method for the representation of a conceptual ETL design as a narrative, which is the most natural means of communication and does not require particular technical skills or familiarity with any specific model.	A flexible and customizable template-based mechanism for the representation of the ETL design as a narrative.
P17	This article presents a user-centered approach to support the end-user requirements elicitation and the DW MD design tasks. This proposal is based on a reengineering process that derives the MD schema from a conceptual formalization of the domain.	A user-centered approach to support the end-user requirements elicitation and the DW MD design tasks.
P18	This article proposes a conceptual design methodology of DWs from data residing in various ontology-based databases that takes into account sources and decision maker requirements.	A methodology and a case tool, called S2RWC, supporting this proposal.
P19	This article introduces an ontology model to integrate semantic heterogeneous data through ETL process for DW development. A domain ontology is introduced into ETL process of finding the data sources, defining the rules of data transformation, and eliminating the heterogeneity.	An ontology-based method to eliminate data heterogeneity so as to support the development of data warehouses.

8 | CONCLUSIONS AND FUTURE WORK

The main motivation for this study was to investigate the current status of Semantic Web technology use in DW development. It was conducted using an SM process, a useful technique for identifying the areas where there is sufficient information to describe the current state of research, as well as those areas that require more information. Through this work, we determined the existence of two lines of study. The first focused on the automation of MD design based on the use of ontologies, and the other examined methods that focus on analyzing large amounts of Semantic Web data. In addition, we identified that the proposals for the use of Semantic Web technologies addressed all stages of DW development. However, some important aspects were not considered and, when they were considered, only a brief review was provided. For example, aspects such as cost and performance were barely discussed. In relation to the experiences of implementation in industry, we found that such experiences were very limited. Many of the cases discussed in the examined studies describe small projects, with results obtained in specific applications, which makes their reproduction in other contexts impractical due to a lack of details. This underscores the necessity of experimenting with more approaches that use Semantic Web technologies in industry, rather than in academia.

TABLE 10 Continuation of the table summary (III)

ID	Summary	Contribution
20	This article presents a system called GEM (generating ETL and MD designs). GEM starts with a set of source data stores and business requirements (e.g., business queries, service level agreements [SLAs]) and based on these, it produces an MD design for the target data stores, along with a set of ETL operations required for the population of the target data warehouse (DW).	A system that facilitates the production of ETL and MD designs, starting from a set of business requirements and source data stores. In addition, some novel algorithms finding and validating an ontology subset as an MD schema, and identifying ETL operators at the same time.
P21	This article provides support for ETL by proposing a tool that: (a) allows the semiautomatic definition of interattribute semantic mappings, by identifying the parts of the data source schemas which are related to the DW schema, thus supporting the extraction process; and (b) groups the attribute values semantically related thus defining a transformation function for populating with homogeneous values the DW.	An ETL tool that couples and extends the functionalities of two previously developed systems: the MOMIS integration system and the RELEVANT data analysis system.
P22	This article proposes to consider the web as an information repository. Since retrieving specific information from the web is challenging, an ontological model is developed to capture specific information by using web semantics. From the ontology model, the relations between the data are mined using decision trees.	An ontology model and an architecture for BI.
P23	This article presents an approach for efficiently analyzing and exploring large amounts of semantic data by combining the inference power from the annotation semantics with the analysis capabilities provided by OLAP-style aggregations, navigation, and reporting.	A semiautomatic method to dynamically build MD models from semantic data guided by the user requirements. In addition, two novel algorithms are presented to dynamically create dimension hierarchies complying with OLAP properties from the taxonomic relations of domain ontologies for MD models generated with the previous method.
P24	This article presents a methodology for designing data warehousing applications from various sources by using ontologies. Each source has its local ontology referencing (and specialize/extend) a global one.	A new classification of DW design methods showing their convergence to ontology-based conceptualization. In addition, a methodology for designing DWs from ontology-based databases and users' requirements supported by a case tool.

TABLE 11 Continuation of the table summary (IV)

ID	Summary	Contribution
P25	This article presents an integrated semantic business intelligence (SBI) architecture for analytical tools. The architecture is supported by business semantics that, in turn, are applied to contextualize the organizations' resources (i.e., logic, data sources, and services).	An ontology for BI, which supplies the business terminology used to enable data sources annotation. In addition, a BI architecture, which is supported by a tool.
P26	This article proposes a framework that uses an ontology for the design of MD models. The framework uses a hybrid approach where the reconciliation of requirements and data source are done at the early stage of design. An ontology reasoning is adopted in order to automatically derive MD elements such as facts and dimensions.	A comprehensive framework using a hybrid methodology to derive an MD model from multiple ontology sources based on requirements. In addition, a set of ontology matching algorithms to map requirements with sources.
P27	This article proposes a new vocabulary, denoted QB4OLAP, which extends QB vocabulary (RDF data cube vocabulary from W3) to fully support OLAP models and operators. This provides algorithms that build the QB4OLAP structures needed to analyze observations already published using QB, and vice versa.	A new vocabulary, which extends QB to fully support OLAP models and operators. In addition, some algorithms that build the structures that allow performing both kinds of analysis, and show compatibility between QB and QB4OLA.
P28	This article presents an ontology-driven tool which helps to automatically derive the conceptual model and logical model for the data warehouse (DW) from a data source and business requirements.	Main contributions are: (a) representation of requirements formally using data warehousing requirement ontology; (b) automatically deriving MD elements present in the data source ontology; and (c) ontology-based DW schema design tool development.
P29	This article proposes a semiautomatic approach to support the schema mapping transformation process. This approach is based on the use a XML schema representation of web data and the existing warehouse schema.	A warehousing web data extraction framework. In addition, some rules to transform web data into a DW.
P30	This article proposes an explicit requirements engineering phase in DW development, which is goal oriented.	A goal-oriented approach to develop DWs.

Research on the intersection between DWs and the Semantic Web has not moved toward industrial tools. Some concerns regarding the validity of our study are presented below:

1. *Quality of the evaluation.* The quality of the selection and evaluation of the studies, as well as the weighting factor used to quantify each, may have failed to represent appropriately the importance of the selected sets. In order to mitigate this problem, the quality attributes were grouped into subsets to facilitate future classifications, and improve the selections.

TABLE 12 Continuation of the table summary (V)

ID	Summary	Contribution
P31	This article presents an ontology-based hybrid approach for data warehouse (DW) design consisting in three main steps: (a) building MD models based on business requirements; (b) building MD models from an operational data source; and (c) matching the two sets of schemas to produce a DW model that agrees to user and data source simultaneously.	A comparative study involving some research works tackling the DW design process. In addition, an ontology-based hybrid approach for the DW design and a framework to produce a DW model.
P32	This article proposes a solution for automated creating a DW from unstructured semantic data using ontology. An algorithm for creating a DW from semantic data by using ontology is presented.	An algorithm can create the desired DW automatically and store the available data from an ontology OWL or RDF.
P33	This article presents three-tier perspectives of the educational data warehousing to abstract ontologies from structural metadata of relational databases which are the resultant of ETL processing as well as generating reports from the data re-fined from the ETL processing.	An ontology-based approach to build an educational DW.
P34	This article presents a prototype named DaWeS (DW fed with web services) and explores how ETL using the mediation approach benefits this trade-off for enterprises with complex data warehousing requirements. The goal of this study is to investigate and devise an approach to address the trade-off between scalability and adaptability in large scale integration with numerous ever-evolving web services.	A prototype to use DaWeS in the business analytics subject. In addition, some optimization heuristics.
P35	This article presents an ontology-driven personalization system (OPS) based on three connected ontologies: domain ontology, DW ontology and resources ontology. OPS return a set of personalized resources search based on users' domain and his recurring interests.	An architecture of the "ontology-driven personalization system" and the use-cases supported by this system. In addition, a methodology used to develop the knowledge base of the personalization system.
P36	This article presents the architecture of a BI system for academic organizations. Then, it is illustrated the design process of the DW devoted to the analysis of the main factors affecting the importance and the quality level of every university, such as the evaluation of the research and the education. The design process is based on a hybrid methodology that is largely automatic and relies on an ontological approach for the integration of the different data sources.	A hybrid methodology. In addition the design of an academic DW and a research data mart.

TABLE 13   Continuation of the table summary (VI)

ID	Summary	Contribution
P37	The paper first presents a characterization of OLAP environments, followed by an introduction to the relevant Semantic Web foundation concepts. Then, it describes the relationship of MD models and Semantic Web technologies. Next, the paper goes on to survey the use of Semantic Web technologies for data modeling and data provisioning, including semantic data annotation and semantic-aware ETL processes. Finally, all the findings are discussed and a number of directions for future research are outlined.	A survey to categorize how Semantic Web technologies have been applied to solve the new requirements of exploratory OLAP systems. In addition, future challenges are identified.
P38	An incremental validation method based on model driven engineering is used to develop a master data management system represented by XML schema models. To do so, authors define an abstraction layer using UML class diagrams. The validation method aims to minimize the model errors and to optimize the process of model checking.	A validation model developed in ArgoUML IDE.
P39	This article presents a conceptual modeling of judgment and analysis rules together with their organization and MD contextualization in rule families, explain different rule evaluation strategies, and briefly report on the implementation of the approach in order to detect MD elements and OLAP operations.	A metamodel for the representation of comparative MD ontologies, including comparative scores, comparative concepts, and comparative cubes. In addition, a modeling of judgment and analysis rules, their organization in rule families, their specialization along subsumption hierarchies of comparative concepts, and their contextualized evaluation according to different rule evaluation strategies.
P40	This article presents an automatic approach to generate useful MD model analytical patterns from linked data sources. MD patterns are based on the semantics of the data.	Mappings from MD models to a statistical layer on top of linked data sources. In addition, a statistical framework to discover MD patterns in an automatic way.
P41	This article provides a new MD model, named unified cube, which offers a generic representation for both warehoused data and linked open data at the conceptual level. A two-stage process is proposed to build a unified cube according to decision-makers' needs. As a first step, schemas published with specific modeling languages are transformed into a common conceptual representation. The second step is to associate together related data to form a unified cube containing all useful information about an analysis subject.	A generic modeling solution, named unified cube, for both warehoused data and linked open data. In addition, a process to build a unified curve. Moreover, a high-level declarative language is provided to enable non-expert users to define the relevance between data according to their analysis needs.

2. *Consideration of synonyms.* The terms used in the search string may have synonyms. As a result, it is possible that the search passed over some studies. On occasion, authors may only mention the term Semantics, without alluding to the Semantic Web, or use some particular technology of the Semantic Web, without making it explicit.

This study identified that the interest in Semantic Web technologies in the development of DWs in the business domain steadily decreased after the year 2012, and that the year of greatest interest was 2009. However, interest began to increase again in 2012, and subsequently declined from 2014 to 2017. This is likely to be due to the fact that there is a growing migration of interest from DWs to LOD and/or logical DW. This systematic map highlights some aspects that require greater research, such as aspects regarding cost, performance, quality assurance, and DW design approaches that consider new technologies. Also, since the use of graph databases in the industrial and marketing arenas is increasing, it is reasonable to think that this technology might benefit the building of DWs in several key issues not only for data modeling but also for other complex issues, such as the consideration of knowledge graphs for evolving DW design or analytic graphs for using advanced data analysis algebra. In our future work, we plan to combine the evidence identified in this study and industrial projects, in order to define hypotheses and theories that will form the basis for the design of new methods, processes and tools to integrate Semantic Web technologies into the DW development process.

9 | BEYOND SM

This section provides a description of each research article analyzed in the study. Tables 5 and 6 present the list of analyzed studies. Tables 7–13 present the summary and the contributions of each research article analyzed in the study.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

ENDNOTES

[1] http://linkeddata.org.

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

[3] (e.g., http://www.dataversity.net/down-with-the-data-warehouse-long-live-the-semantic-data-warehouse/).

[4] https://www.gartner.com/doc/3096424/hype-cycle-enterprise-information-management.

[5] https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp.

[6] https://www.gartner.com/doc/3096424/hype-cycle-enterprise-information-management.

[7] https://virtuoso.openlinksw.com/.

[8] https://jena.apache.org.

[9] https://www.targit.com/en/software/decision-suite/.

[10] http://opencube-toolkit.eu/.

[11] https://ontotext.com.

[12] https://neo4j.com/blog/why-graph-databases-are-the-future/.

[13] Forrester Research, Forrester Vendor Landscape: Graph Databases, Yuhanna, 6 Oct. 2017.

[14] https://www.datanami.com/2017/11/30/look-graph-database-landscape/.

[15] https://github.com/thomsonreuters/CM-Well.

[16] https://cordis.europa.eu/projects.

[17] http://geoknow.eu.

[18] http://www.seek-project.eu.

[19] http://opencube-project.eu/.

[20] http://portal.acm.org/ccs.cfm?part=author&coll=portal&dl=GUIDE.

RELATED WIREs ARTICLES

Alignment and dataset identification of linked data in Semantic Web

Automatic construction of lexicons, taxonomies, ontologies, and other knowledge structures

ORCID

Ricardo Gacitua https://orcid.org/0000-0002-3900-2770

REFERENCES

Abelló, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazón, J.-N., Naumann, F., … Vossen, G. (2013). Fusion cubes: Towards self-service business intelligence. *The International Journal of Data Warehousing and Mining*, *9*, 66–88. https://doi.org/10.4018/jdwm.2013040104

Abelló, A., Romero, O., Bach Pedersen, T., Berlanga, R., Nebot, V., Aramburu, M. J., & Simitsis, A. (2015). Using Semantic Web technologies for exploratory olap: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *27*, 571–588.

Abramowicz, W., Auer, S., & Heath, T. (2016). Linked data in business. *Business & Information Systems Engineering*, *58*, 323–326. https://doi.org/10.1007/s12599-016-0446-0

Ali, A. A., Abdelrahman, T. A., & Mohamed, W. M. (2013). Using schema matching in data transformation for warehousing web data. *International Journal Information Technologies & Knowledge*, 7.

Bennett, D., Engelbrecht, J., & Landau, D. (2017). Cm-well: A data warehouse for linked data. In: Nadeschda N, Dezhao S, Achille F, Peter H, eds. *Proceedings of the ISWC 2017 posters & demonstrations and industry tracks co-located with 16th international Semantic Web conference (ISWC 2017), Vienna, Austria, October 23rd to 25th 2017*. Germany: CEUR-WS.org. http://ceur-ws.org/Vol-1963/paper518.pdf

Bergamaschi, S., Guerra, F., Orsini, M., Sartori, C., & Vincini, M. (2011). A semantic approach to etl technologies. *Data & Knowledge Engineering*, *70*, 717–731.

Berlanga, R., Aramburu, M. J., Llidó, D. M., & García-Moya, L. (2014). Towards a semantic data infrastructure for social business intelligence. In B. Catania, T. Cerquitelli, S. Chiusano, G. Guerrini, M. Kämpf, A. Kemper, et al. (Eds.), *New trends in databases and information systems* (pp. 319–327). Cham, Switzerland: Springer International Publishing.

Bizer, C., Heath, T. and Berners-Lee, T. (2009) *Linked data—The story so far*. PA: IGI Global. https://eprints.soton.ac.uk/271285/

Chakraborty, J., Padki, A., & Bansal, S. K. (2017). Semantic etl—State-of-the-art and open research challenges. In *2017 I.E. 11th international conference on semantic computing (ICSC)* (Vol. 00, pp. 413–418). NJ: IEEE. https://doi.org/doi.ieeecomputersociety.org/10.1109/ICSC.2017.94

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). *Business intelligence and analytics: From big data to big impact* (Vol. 36, pp. 1165–1188). MN: Management Information Systems Research Center. http://dl.acm.org/citation.cfm?id=2481674.2481683

Coral, C., Francisco, R., & Mario, P. (2006). *Ontologies for software engineering and software technology*. Berlin, Heidelberg: Springer-Verlag.

da Mota Silveira Neto, P. A., Carmo Machado, I. D., McGregor, J. D., de Almeida, E. S., & de Lemos Meira, S. R. (2011). A systematic mapping study of software product lines testing. *Information and Software Technology*, *53*, 407–423. https://doi.org/10.1016/j.infsof.2010.12.003

Di, T. F., Lefons, E., & Tangorra, F. (2015). Academic data warehouse design using a hybrid methodology. *Computer Science and Information Systems*, *12*, 135–160.

Diamantini, C., & Potena, D. (2008). Semantic enrichment of strategic datacubes. In: Il-Yeol S, Alberto A, eds. *Proceedings of the ACM 11th international workshop on data warehousing and OLAP* (pp. 81–88). NY: ACM.

El Sarraj, L., Espinasse, B., & Libourel, T. (2014). An ontology-driven personalization approach for data warehouse exploitation. *International Journal on Advances in Software*, *7*, 253–265.

Elamin, E., & Feki, J. (2014). Toward and ontology based approach for data warehousing. In *The international Arab conference on information technology (ACIT2014)* (pp. 170–179). Jordan, SA: ACIT.

Etcheverry, L., & Vaisman, A. (2012). Qb4olap: A new vocabulary for olap cubes on the Semantic Web. In *Proceedings of COLD*. achen, Germany: CEUR-WS.org

Golfarelli, M., Rizzi, S., & Cella, I. (2004). Beyond data warehousing: What's next in business intelligence? In *Proceedings of the 7th ACM international workshop on data warehousing and OLAP, DOLAP '04* (pp. 1–6). New York, NY: ACM http://doi.acm.org/10.1145/1031763.1031765

Hendler, J. (2001). Agents and the Semantic Web. *IEEE Intelligent Systems*, *16*, 30–37. https://doi.org/10.1109/5254.920597

Henschen, D. (2015) *Analytics, bi, data management trends for 2015*. CA: InformationWeek. http://www.informationweek.com/big-data/big-data-analytics/5-analytics-bi-data-management-trends-for-2015/a/d-id/1318551

Inmon, W. H. (1992). *Building the data warehouse*. New York, NY: John Wiley & Sons, Inc.

Jiang, L., Cai, H., & Xu, B. (2010). A domain ontology approach in the etl process of data warehousing. In *2010 I.E. 7th international conference on e-business engineering (ICEBE)* (pp. 30–35). NJ: IEEE.

Jossen, C., & Dittrich, K. R. (2007). The process of metadata modeling in industrial data warehouse environments. In *BTW workshops* (pp. 16–27). Germany: Verlagshaus Mainz, Aachen.

Kalampokis, E., Nikolov, A., Haase, P., Cyganiak, R., Stasiewicz, A., Karamanou, A., Tarabanis, K. A. (2014). Exploiting linked data cubes with opencube toolkit. In *International Semantic Web conference (Posters & Demos)* (Vol. 1272, pp. 137–140). Aachen, Germany: CEUR-WS.org.

Khouri, S., Bellatreche, L., Jean, S., & Ait-Ameur, Y. (2014). Requirements driven data warehouse design: We can go further. In *Leveraging applications of formal methods, verification and validation. Specialized techniques and applications* (pp. 588–603). New York, NY: Springer.

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* (pp. 19–24). New York, NY: ACM.

Kimball, R., & Ross, M. (2002). *The data warehouse toolkit: The complete guide to dimensional modeling* (2nd ed.). New York, NY: John Wiley & Sons, Inc.

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering—A tertiary study. *Information and Software Technology*, *52*, 792–805. https://doi.org/10.1016/j.infsof.2010.03.006

Kitchenham, B. A. (2012). Systematic review in software engineering: Where we are and where we should be going. In *Proceedings of the 2nd international workshop on evidential assessment of software technologies, EAST '12* (pp. 1–2). New York, NY: ACM. http://doi.acm.org/10.1145/2372233.2372235

Kitchenham, B. A., Budgen, D., & Brereton, O. P. (2011). Using mapping studies as the basis for further research—A participant-observer case study. *Information and Software Technology*, *53*, 638–651.

Klyne, G. and Carroll, J. J. (2004) *Resource description framework (RDF): Concepts and abstract syntax. World wide web consortium, recommendation REC-rdf-concepts-20040210*. Cambridge, MA: W3C.

Lamolle, M., Menet, L., & Le Duc, C. (2015). Incremental checking of master data management model based on contextual graphs. *Enterprise Information Systems*, *9*, 681–708.

Lather, S. (2012). *Business intelligence and data warehousing*. India: Narosa Publishing House.

Lenz, H.-J., & Shoshani, A. (1997). Summarizability in olap and statistical data bases. In *Proceedings of 9th international conference on Scientific and statistical database management* (pp. 132–143). NJ: IEEE.

Martin, A., Maladhy, D., & Venkatesan, V. P. (2011). A framework for business intelligence application using ontological classification. *International Journal of Engineering Science and Technology*, *3*(2), 1213–1221 arXiv preprint arXiv:1109.1088.

McGregor, J. D. (2002). Building reusable test assets for a product line. In C. Gacek (Ed.), *Software reuse: Methods, techniques, and tools* (pp. 345–346). Berlin, Heidelberg: Springer Berlin Heidelberg.

Middelfart, M., & Pedersen, T. B. (2011). The meta-morphing model used in targit bi suite. In *International conference on conceptual modeling* (pp. 364–370). New York, NY: Springer.

Nazri, M. N. M., Noah, S. A. M., & Hamid, Z. (2008). Automatic data warehouse conceptual design. In *International Symposium on Information Technology, 2008. ITSim 2008* (Vol. 2, pp. 1–7). NJ: IEEE.

Nebot, V., & Berlanga, R. (2012). Building data warehouses with Semantic Web data. *Decision Support Systems*, *52*, 853–868.

Nebot, V., & Berlanga, R. (2016). Statistically-driven generation of multidimensional analytical schemas from linked data. *Knowledge-Based Systems*, *110*, 15–29 http://www.sciencedirect.com/science/article/pii/S0950705116302143

Nebot, V., Berlanga, R., Pérez, J. M., Aramburu, M. J., & Pedersen, T. B. (2009). Multidimensional integrated ontologies: a framework for designing semantic data warehouses. In *Journal on data semantics XIII* (pp. 1–36). New York, NY: Springer.

Nguyen, T. M., Tjoa, A. M., & Trujillo, J. (2005). Data warehousing and knowledge discovery: A chronological view of research challenges. In A. M. Tjoa & J. Trujillo (Eds.), *Data warehousing and knowledge discovery* (pp. 530–535). Berlin, Heidelberg: Springer Berlin Heidelberg.

Niemi, T., Toivonen, S., Niinimaki, M., & Nummenmaa, J. (2007). Ontologies with Semantic Web/grid in data integration for olap. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *3*, 25–49.

Niinimäki, M., & Niemi, T. (2009). An ETL process for OLAP using RDF/OWL ontologies. In S. Spaccapietra, E. Zimányi, & I.-Y. Song (Eds.), *Journal on data semantics XIII* (pp. 97–119). Berlin, Heidelberg: Springer-Verlag http://dl.acm.org/citation.cfm?id=2172259.2172263

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic mapping studies in software engineering. In *Proceedings of the 12th international conference on evaluation and assessment in software engineering, EASE'08* (pp. 68–77). Swinton, UK, UK: British Computer Society http://dl.acm.org/citation.cfm?id=2227115.2227123

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18 http://www.sciencedirect.com/science/article/pii/S0950584915000646

Priebe, T., & Pernul, G. (2003). Towards integrative enterprise knowledge portals. In *Proceedings of the twelfth international conference on information and knowledge management* (pp. 216–223). New York, NY: ACM.

Ramasamy, V., & Palanivel, K. (2014). Triplet dependency views of university data warehousing towards decision support system. *International Journal of Advanced Technology & Engineering Research (IJATER)*, *1*, 1–7.

Ravat, F., Song, J., & Teste, O. (2016). Designing multidimensional cubes from warehoused data and linked open data. In *10th international IEEE conference on research challenges in information science (RCIS 2016) co-located witht the 34th French conference INFORSID* (pp. 199–200). Grenoble, FR: INformatique des ORganisations et Systèmes d'Information et de Décision (INFORSID) http://oatao.univ-toulouse.fr/17167/

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

Romero, O., Simitsis, A., & Abelló, A. (2011). Gem: Requirement-driven generation of etl and multidimensional conceptual designs. In *Data warehousing and knowledge discovery* (pp. 80–95). New York, NY: Springer.

Saha, G. K. (2007). Web ontology language (owl) and Semantic Web. *Ubiquity*, *2007*, 1:1–1:1 http://doi.acm.org/10.1145/1295289.1295290

Salguero, A., Araque, F., & Delgado, C. (2008a). Ontology based framework for data integration. *WSEAS Transactions on Information Science and Applications*, *5*, 953–962.

Salguero, A., Araque, F., & Delgado, C. (2008b). Spatio-temporal ontology based model for data warehousing. In *Proceedings of the 7th WSEAS international conference on telecommunications and informatics, TELE-INFO'08* (pp. 125–130). Stevens Point, WI: World Scientific and Engineering Academy and Society (WSEAS) http://dl.acm.org/citation.cfm?id=1404049.1404072

Samuel, J. (2014). Towards a data warehouse fed with web services. In *The Semantic Web: Trends and challenges* (pp. 874–884). New York, NY: Springer.

Sell, D., Cabral, L., Motta, E., Domingue, J., & Pacheco, R. (2005). Adding semantics to business intelligence. In *Proceedings of sixteenth international workshop on database and expert systems applications* (pp. 543–547). NJ: IEEE.

Sell, D., da Silva, D. C., Beppler, F. D., Napoli, M., Ghisi, F. B., Pacheco, R., & Todesco, J. L. (2008). Sbi: A semantic framework to support business intelligence. In *Proceedings of the first international workshop on ontology-supported business intelligence* (p. 11). New York, NY: ACM.

Sell, D., da Silva, D. C., Ghisi, F. B., Todesco, J. L., & Napoli, M. (2012). *Adding semantics to business intelligence: Towards a smarter generation of analytical tools*. London: IntechOpen Limited.

Selma, K., Ilyès, B., Ladjel, B., Eric, S., Stéphane, J., & Michael, B. (2012). Ontology-based structured web data warehouses for sustainable interoperability: Requirement modeling, design methodology and tool. *Computers in Industry*, *63*, 799–812.

Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The Semantic Web revisited. *IEEE Intelligent Systems*, *21*, 96–101. https://doi.org/10.1109/MIS.2006.62

Shaw, M. (2003). Writing good software engineering research papers: Minitutorial. In *Proceedings of the 25th international conference on software engineering, ICSE '03* (pp. 726–736). Washington, DC: IEEE Computer Society http://dl.acm.org/citation.cfm?id=776816.776925

Simitsis, A., Skoutas, D., & Castellanos, M. (2010). Representation of conceptual etl designs in natural language using Semantic Web technology. *Data & Knowledge Engineering*, *69*, 96–115.

Skoutas, D., & Simitsis, A. (2006). Designing etl processes using Semantic Web technologies. In *Proceedings of the 9th ACM international workshop on data warehousing and OLAP* (pp. 67–74). New York, NY: ACM.

Skoutas, D., & Simitsis, A. (2007). Ontology-based conceptual design of etl processes for both structured and semi-structured data. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *3*, 1–24.

Skoutas, D., Simitsis, A., & Sellis, T. (2009). Ontology-driven conceptual design of etl processes using graph transformations. In *Journal on data semantics XIII* (pp. 120–146). New York, NY: Springer.

Spahn, M., Kleb, J., Grimm, S., & Scheidl, S. (2008). Supporting business intelligence by providing ontology-based end-user information self-service. In *Proceedings of the first international workshop on ontology-supported Business intelligence* (Vol. 10). New York, NY: ACM.

Steiner, D., Neumayr, B., & Schrefl, M. (2015). Judgement and analysis rules for ontology-driven comparative data analysis in data warehouses. In *Proceedings of the 11th Asia-Pacific conference on conceptual modelling (APCCM 2015)* (Vol. 27, p. 30). Sydney: Australian Computer Society Inc.

Talebzadeh, S., Seyyedi, M. A., & Salajegheh, A. (2014). Automated creating a data warehouse from unstructured semantic data. *International Journal of Computer Applications*, *88*, 19–25.

Thenmozhi, M., & Vivekanandan, K. (2012). An ontology based hybrid approach to derive multidimensional schema for data warehouse. *International Journal of Computer Applications*, *54*, 36–42.

Thenmozhi, M., & Vivekanandan, K. (2013). A tool for data warehouse multidimensional schema design using ontology. *International Journal of Computer Science Issues (IJCSI)*, *10*, 161–168.

Trujillo, J., & Maté, A. (2012). Business intelligence 2.0: A general overview. In *Business intelligence: First European summer school, eBISS 2011, Paris, France, July 3-8, 2011, Tutorial Lectures* (pp. 98–116). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-27358-2_5

Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., & Hübner, S. (2001). *Ontology-based integration of information—A survey of existing approaches* (pp. 108–117). CA: IJCAI.

Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2005). Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. *Requirements Engineering*, *11*, 102–107. https://doi.org/10.1007/s00766-005-0021-6