



HAL
open science

Requirement Analysis in Data Warehouses to Support External Information

Mohamed Lamine Chouder, Rachid Chalal, Waffa Setra

► **To cite this version:**

Mohamed Lamine Chouder, Rachid Chalal, Waffa Setra. Requirement Analysis in Data Warehouses to Support External Information. 5th International Conference on Computer Science and Its Applications (CIIA), May 2015, Saida, Algeria. pp.243-253, 10.1007/978-3-319-19578-0_20 . hal-01789941

HAL Id: hal-01789941

<https://inria.hal.science/hal-01789941>

Submitted on 11 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Requirement Analysis in Data Warehouses to Support External Information

Mohamed Lamine Chouder¹, Rachid Chalal¹, and Waffa Setra¹

¹LMCS (Laboratoire de Methodes de Conception de Systemes),
ESI (Ecole nationale Supérieure d'Informatique), Algiers, Algeria
{m_chouder, r_chalal, w_setra}@esi.dz

Abstract. In strategic decision-making, the decision maker needs to exploit the strategic information provided by decision support systems (DSS) and the strategic external information emanating from the enterprise business environment. The data warehouse (DW) is the main component of a data-driven DSS. In the field of DW design, many approaches exist but ignore external information and focus only on internal information coming from the operational sources. The existing approaches do not provide any instrument to take into account external information. In this paper, our objective is to introduce two models that will be employed in our approach: the requirement model and the environment model. These models are the basis of our DW design approach that supports external information. To evaluate the requirement model, we will illustrate with an example how to obtain external information useful for decision-making.

Keywords: Data warehouse; design; requirement analysis; external information; business environment.

1 Introduction

A data warehouse (DW) that supports external information is a knowledge source for strategic decision-making. It provides historical information about the enterprise business environment. External information is the strategic information useful for decision-making, about competitors, customers, markets, suppliers, products... Unfortunately, this type of information is informal, heterogeneous and unstructured, which makes the process of developing a DW that satisfies decision-makers needs a difficult and a complex task. For many years, it is widely accepted that the basis for designing a DW is multidimensional (MD) modeling [1, 2]. Today, the MD form is natural to decision makers, by means of its structure composed of analysis measures and dimensions that represent the context for analysis.

In the literature, two different categories of DW design approaches exist: data-driven and requirement-driven. The former starts from operational sources to define the MD model of the DW [3, 4]. The latter tries to identify the requirements to build the DW and define its contents [5]. These approaches collect requirements through different orientations: users, processes, and goals and using different techniques (See section 2). In this work, our interest is concentrated on requirement-driven, mostly

goal-oriented, DW design approaches for two reasons: (i) the strategic goals of the organization in the business environment are considered the main resource to identify external information requirements; (ii) the structure of the external information source, which will aliment the DW, is not defined, unlike in operational sources, so it must be defined. We argue that in the field of DW design, the existing approaches focus only on internal information coming from operational sources. These approaches raise the importance of external information, but ignore it and do not provide any instrument to support it.

To answer this, we propose a goal-oriented requirement analysis approach in DWs to support external information. This approach can be used to build a DW that contains strategic information useful for decision-making. In this paper, the models that will be employed in our approach are detailed: the requirement model and the environment model. The former is an improved requirement model from the model proposed in [6]. This model aims to identify the strategic decision-making needs for external information. The latter represents useful information for strategic decision-making about some environment elements (competitor, customer, market, product, and supplier). In a future work, a set of processes will be defined to show how to use the models described in this paper for defining external information requirements and the underlying MD model. To evaluate our proposal, the requirement model will be illustrated using an example: the strategic goal "Increase market share". To do this, a set of guidelines are defined to show how to obtain external information useful for decision-making.

The remainder of this paper is organized as follows. Section 2 discusses related work. In Section 3, the models employed in our approach are described. Section 4 represents the illustration of the requirement model with an example. Finally, Section 5 points out the conclusion and future work.

2 Related Work

In the last decade, various DW design approaches were proposed to define requirements using different techniques. For example: Business process models in [7], Goal-Question-Metric approach in [8], use cases in [9], best practices in [10], traditional requirements engineering (RE) process in [11], Decision processes analysis in [5], Map goal model in [12], GDI model in [13], extended Tropos in [14], extended i* framework in [15], etc... Due to lack of space, in this section we give a brief description of the most relevant requirement-driven, mostly goal-oriented, DW design approaches.

Starting with [13], DW requirements are determined in the broader context of the goals and objectives of an organization. At first, in an organizational perspective, requirements are grouped into several levels of abstraction using the Goal-Decision-Information (GDI) model. It starts by determining goals, then the decisions that influence the satisfaction of these goals. Finally, the information needed to make decisions is identified. At second in a technical point of view, information scenarios are applied for each decision, to define DW contents and their proprieties. This approach shares

similarities with ours in abstraction levels and because it deals with decisional goals, however, it does not consider external information.

[14] presented GRAND, a goal-oriented approach, which has extended the early phase of Tropos [16] to the requirements engineering of DWs. Tropos is an agent-oriented method, which is a variant of i^* [17]. In requirement analysis, the stakeholder's dependencies are represented in an actor diagram. Then, two perspectives are adopted, organizational and decisional. In the former, facts and attributes are identified and associated with goals of different actors. In the latter, each fact is related to their dimensions and a set of measures is found out and associated with facts. In conceptual modeling, this approach can be either employed, within requirement-driven or mixed requirement/data-driven, to specify the conceptual MD model. This approach share similarities with ours in goal reasoning. However, it does not consider decisional goals unlike [13] and does not provide any instrument to take into account external information.

In [15] another goal-oriented approach is proposed based on the i^* framework [17]. At first, the strategic, decisional and informational goals are identified through interviews. The information requirements (tasks and resources) are obtained from the informational goals of different actors using two i^* models: Strategic Dependency (SD) model and Strategic Rational (SR) model. This latter is applied for the DW actor to define the rational model, which will give rise to the design of a conceptual MD model using a UML profile [18]. In this approach, organizational modeling unlike [14] and external information are not supported.

In recent years, many researchers focus on understanding the business context in which the DW will be implemented. In [19] an extended version of the work in [15] was proposed to align DW requirements with the business strategy. This approach considered the business strategy using VMOST (vision, mission, objective, strategy, and tactics) and the business motivation model (BMM) to align DW goals and the organization strategy. In the same direction, another effort has been made by [20] to align the i^* concepts for requirement analysis in DWs with the business strategy model proposed in the business intelligence model (BIM) [21].

BIM is a business modeling language that offers many diagrams to help business users make sense of data manipulated in business intelligence systems. Different reasoning techniques about goals, goal influences, situations and indicators are used to define a complete business strategy plan [22]. The interesting thing for us in this model is that external and internal situations, that influence the fulfillment of a goal, are identified. Then, one or many external indicators are associated with an external situation (e.g. number of competitors). However, the authors do not indicate how to identify these indicators and how to represent them in a data perspective (dimensions and measures). In addition, this model is used to shape the organizational strategy, not to build a DW.

Although, the formalisms used in the presented approaches for requirement analysis step are different, their expressivity is very close, and show that a core of common information has been identified [23]. In the next section, we will present an outline of our approach for requirement analysis in DWs to support external information, which is not supported in the above-presented approaches.

3 Requirement Analysis in DWs to Support External Information

In strategic decision-making, the decision maker needs to exploit mainly the information emanating from the enterprise business environment. This information is qualified as external information. Since DW design approaches try to provide organizations with information to support decision-making. A DW design approach should be defined to support external information useful for decision-making. Therefore, better decisions will be taken and strategic goals are achieved.

In our work, the strategic objectives that the organization must achieve are considered the strategic goals to identify external information. This latter is the information required to support decision-making that must be provided by the DW. A number of proposals for requirement analysis in DWs have been made as seen in section 2 with focus on information coming from operational sources. Many efforts also have been made to define the system development lifecycle (SDLC) for DW development. Some approaches take the ER diagram or the database scheme of operational sources as an input to their DW requirement analysis stage [4]. In [13], the authors argue that DW development must be rooted by the set of goals and decisions interesting the organization rather than the schema of operational sources.

In our approach, we propose a hybrid SDLC (See table 1) where the DW schema and the source schema¹ are defined at the same time. In requirement analysis, two models are adopted: a requirement model and an environment model. The former is used to identify external information required for decision-making. The latter is a UML class diagram, which will be used in requirement analysis to derive information that must be provided by the DW represented in the form of dimensions and measures. In addition, in conceptual design, this model will give rise to the design of the DW multidimensional schema and the schema of the data source that will aliment the DW.

Table 1. The proposed SDLC

Stage	Output
Requirement analysis	Requirement diagram, Specific environment model
Conceptual design	DW multidimensional schema, Source conceptual schema
Logical design	DW logical schema, Source logical schema
Physical design	DW physical schema, Source physical schema

In the following subsections, the models that will be employed in our approach are described in detail. The manner of using the described models together in requirement analysis is not addressed in this paper. This aspect will be extended in a future work.

¹ The source schema is the structure of the data source that will aliment the DW in contrast with operational sources

3.1 The Requirement Model

The requirement model is an improvement to the strategic goal model proposed in a previous work [6]. In our model, a strategic goal is considered as an objective that is to be met by the organization at the strategic level. From the implementation process of a strategic goal, a set of goals are derived [6]. Once a goal is defined, it either needs means to concretize it or decisions to realize it.

As shown in Fig.1 a strategic goal consists of a set of goals, which makes a goal hierarchy. A goal can be either qualified as an operational goal or a decisional goal. The latter is a long-term goal in the strategic/tactical level of the organization that needs decisions to realize it. Whereas an operational goal is an objective that can be met by a transactional information system, which is concretized by realization and control means.

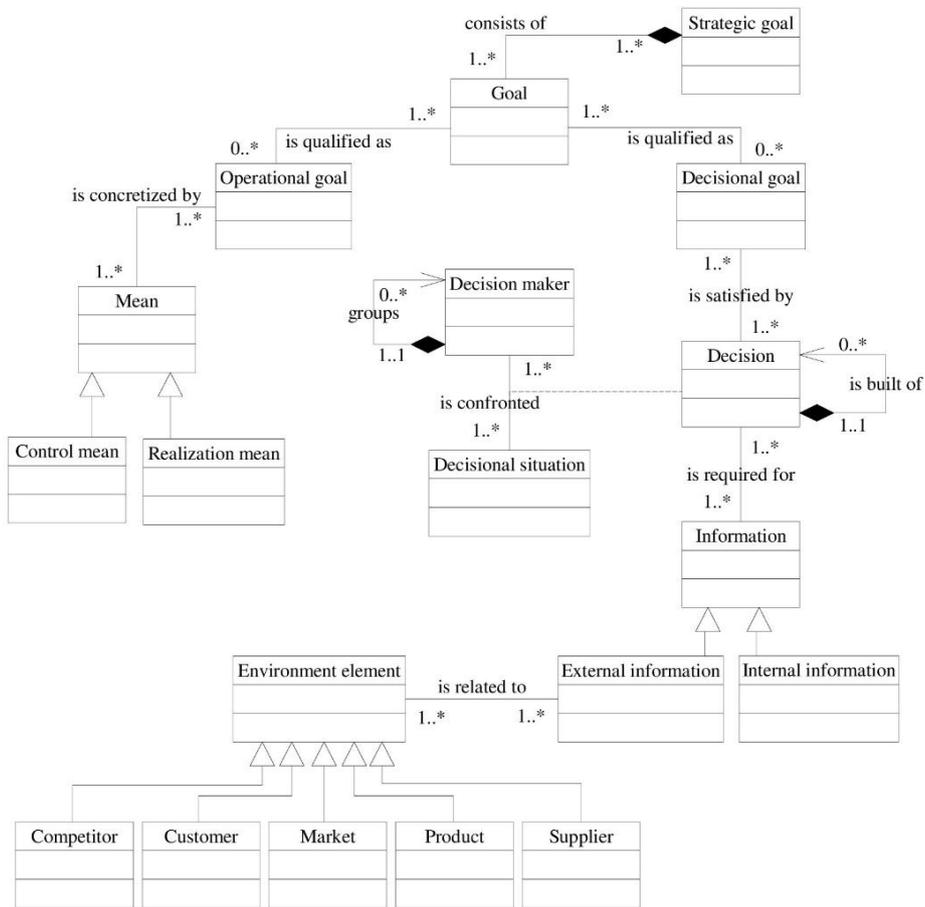


Fig. 1. The requirement model

A decision is the intention to perform the actions that cause its implementation to fulfill decisional goals [13]. Decision-making is an activity that results in the selection of the decision to be implemented. While performing this activity, the decision maker is in a decisional situation where he/she requires the appropriate information to select the right decision among alternative decisions. As shown in Fig.1, a decision can be built out of other decisions as in [13]. A decision maker, which can be an individual or a group, is confronted to a decisional situation where decisions have to be taken. The association ‘is satisfied by’ between decisional goal and decision identifies the decisions which, when taken can lead to decisional goals satisfaction.

Knowledge necessary to take decisions is represented by the form of information. Fig.1 shows that there is an association ‘is required for’ between information and decision. This association identifies the information required to take a decision. The information can be external or internal. The internal information branch is out the scope of our interest. External information is the specification of data that will be stored in the DW. It is the information about environment elements (competitor, customer, market, product, and supplier). As shown in Fig.1, the association ‘is related to’ between external information and environment element identifies the information about the environment.

To offer a graphical support for requirement modeling, the notation of some elements from the i* framework [17] will be used. The graphical extended notation is summarized in Fig.2.

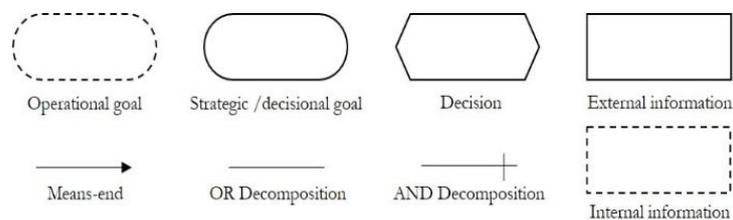


Fig. 2. The notation used in requirement modeling

3.2 The Environment Model

As seen in the previous subsection, information about the environment is necessary in strategic decision-making, which has an impact on the enterprise competitiveness. The business environment can be defined as the external factors that influence directly or immediately the enterprise. It is composed of two categories: (i) the macro-environment, which is the general environment that integrates political, economic, social, technological, legal... aspects; (ii) the micro-environment, that is our concern, is the close environment constituted with factors like customers, competitors, markets, products, and suppliers...

One of the largest used models in analyzing the environment is the Porter’s five forces model [24]. It identifies five forces influencing the enterprise in a competitive environment: rivalry between competitors, threat of potential entrants, bargaining power of suppliers, bargaining power of customers, and threat of substitute prod-

ucts/services. Taking into account these forces, we can assume that the five major elements of the enterprise environment are: competitor, customer, market, product, and supplier. The literature on environment analysis only proposes outlines and does not provide any structuration of information about the enterprise business environment [25]. Therefore, the proposal of an environment model, which will offer a generic view of this environment.

The first model produced by our research team is the competitor model [25] (See Fig.3). In this paper, only this model will be described in detail. Nevertheless, the other parts of the environment model have the same definition, as the competitor model, described in the next paragraph. The competitor model assembles informational bricks serving for the acquisition of information about the competitor. This model has been modeled based on the UML class diagram for many reasons: (i) it is largely used today, so it is familiar to designers; (ii) it permits to represent different points of view. In addition, to facilitate its navigation and use, the model has been conducted using the meta-modeling principle. Thus, it consists of two levels of modeling: meta-class level and class level.

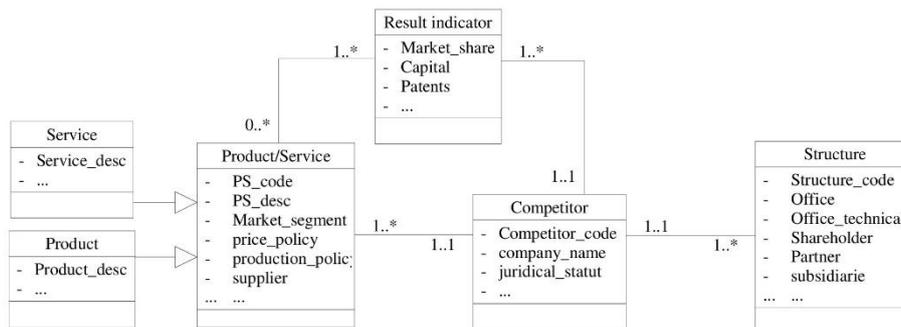


Fig. 3. Competitor meta-model [25]

As illustrated in Fig.3, the competitor meta-model regroups different meta-classes that are related to the competitor:

- Result indicator meta-class regroups financial, statistical, and commercial results about the competitor like capital, market share, patents...
- Structure meta-class represents: (i) information about the competitor identity, structural and organizational aspects, shareholders, partners...; (ii) information about the competitor-implemented strategies, techniques used for each enterprise domain: commercialization, distribution, projects funding, provision, research and development...
- Product meta-class is the main component in this model. It describes information about the competitor activity (products and services) and policies applied to products...

Fig.4 shows the product meta-model where the product class has many associations with other classes: campaign, customer opinion, market, supplier, price-policy, production-policy, promotion...

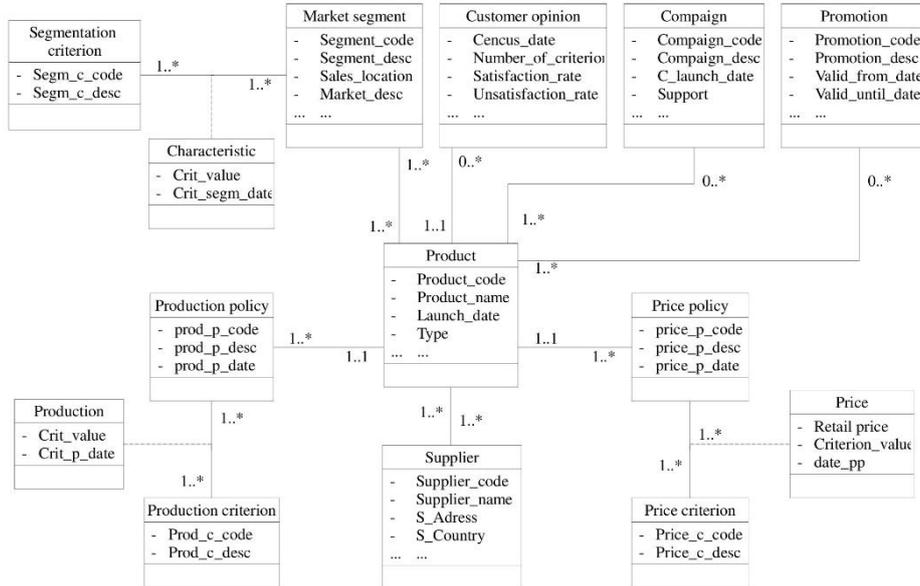


Fig. 4. The product meta-class of the competitor model [25]

In a future work, the environment model will be used in requirement analysis with the requirement model to shape the external information required in the form of dimensions and measures. In addition, it will be used in conceptual design to define the MD model and the schema of the data source that will aliment the DW. In the next section, the requirement model will be illustrated with an example and a set of guidelines are defined to show how to use it.

4 Sample Application of the Requirement Model

In this section, we propose a set of guidelines to show how to use the requirement model. These guidelines are used to demonstrate how to derive external information useful for decision-making from the strategic goal “Increase market share”. The outcome of requirement modeling is a requirement diagram.

Guideline 1. The process of implementing a strategic goal must be described. It will be analyzed to identify a set of goals relative to a strategic goal. See [6] for more details on how to do this task semi-automatically.

Guideline 2. The goals identified in the previous step are qualified as decisional goals or operational goals. Operational goals are excluded. The resulted hierarchy is modeled as a goal hierarchy using the goal notation and decomposition links.

Guideline 3. A set of decisions is identified for each decisional goal and linked to their respective goals with a means-end link. After that, complex decisions are

decomposed. The decision hierarchy is modeled using the decision notation and decomposition links.

Guideline 4. For each decision, a decisional situation is identified. A decisional situation is the description of the decisional problem, which identifies the internal and external variables that influence the decision-maker when taking a decision. Therefore, it could define the nature of the information needed.

Guideline 5. From the decisions and decisional situations identified in the previous step, environment elements are identified. The external information needed to take a decision is defined in the form “information about environment elements”. This external information is associated with a means-end link to each decision in the requirement diagram.

Example. Fig.5 shows the resulted requirement diagram of requirement modeling, after analyzing the strategic goal “Increase market share”. This goal is built out of two decisional goals: “Increase sales” and “Retain customer loyalty”. For the goal “Increase sales”, two decisions are identified “Launch a new product” and “Open new sales channels”. Moreover, for the goal “Retain customer loyalty”, the decision “Improve quality of existing products” is identified.

For the decision “launch a new product”, two environment elements are identified product and competitor. Therefore, the external information “information about competitors’ products” is associated to the decision “launch a new product”. Identically “information about competitors’ sales channels” is associated with the decision “open new sales channels”. Furthermore, “information about customer preferences” is associated to the decision “improve quality of existing products”.

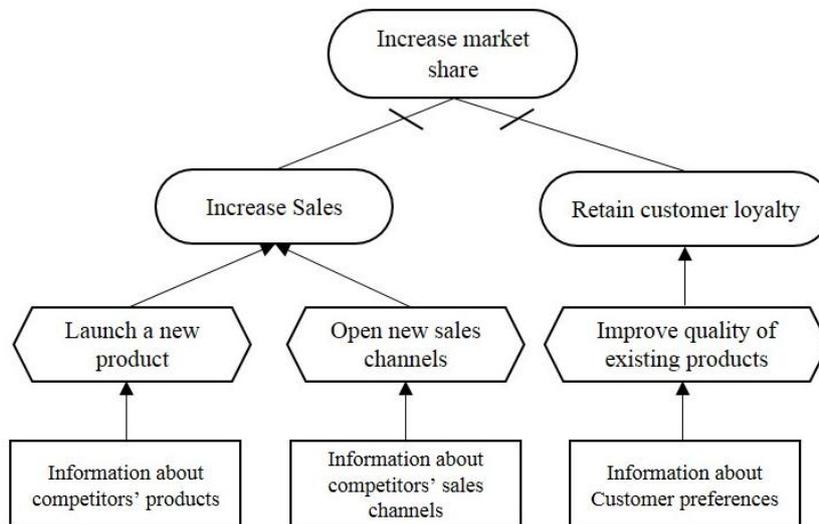


Fig. 5. Requirement diagram for the strategic goal “Increase market share”

5 Conclusion

In this paper, we have presented the first step of a goal-oriented approach for requirement analysis in DWs to support external information. As existing approaches mainly focus on internal information coming from operational sources, our approach provides the means to build a DW that contains strategic information about the enterprise business environment. We were limited to the description of the models that will be employed in our approach: the requirement model and the environment model. The former is used to identify external information required for strategic decision-making. The latter represents information useful for decision-making about the enterprise business environment. Immediate planned future work involves defining a set of processes to show how to use these models to shape the external information required in the form of dimensions and measures. Then to obtain the underlying MD model and the structure of the data source that will aliment the DW.

References

1. Kimball, R., Ross, M.: The data warehouse lifecycle toolkit. John Wiley & Sons (2002)
2. Inmon, W.H.: Building the data warehouse. John Wiley & Sons (1996)
3. Golfarelli, M., Maio, D., Rizzi, S.: The dimensional fact model: a conceptual model for data warehouses. *International Journal of Cooperative Information Systems* 7, 215-247 (1998)
4. Hüsemann, B., Lechtenböcker, J., Vossen, G.: Conceptual data warehouse design. Universität Münster. *Angewandte Mathematik und Informatik* (2000)
5. Winter, R., Strauch, B.: A method for demand-driven information requirements analysis in data warehousing projects. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pp. 9 pp. IEEE, (2003)
6. Boukrara, A., Chalal, R.: Specification of useful information for the strategic decision support: risks-based approach. *International Journal of Decision Sciences, Risk and Management* 4, 276-293 (2012)
7. Böhnlein, M., Ulbrich-vom Ende, A.: Business process oriented development of data warehouse structures. *Data Warehousing*, pp. 3-21. Springer (2000)
8. Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., Paraboschi, S.: Designing data marts for data warehouses. *ACM transactions on software engineering and methodology* 10, 452-483 (2001)
9. Bruckner, R., List, B., Scheifer, J.: Developing requirements for data warehouse systems with use cases. In: *AMCIS Proceedings*, pp. 66. (2001)
10. Schiefer, J., List, B., Bruckner, R.: A holistic approach for managing requirements of data warehouse systems. In: *AMCIS Proceedings*, pp. 13. (2002)
11. Paim, F.R.S., de Castro, J.F.B.: DWARF: An approach for requirements definition and management of data warehouse systems. In: *Proceedings of 11th IEEE International Requirements Engineering Conference* pp. 75-84. IEEE, (2003)
12. Gam, I., Salinesi, C.: A requirement-driven approach for designing data warehouses. In: *Proceedings of Requirements Engineering: Foundation for Software Quality (REFSQ)*. (2006)

13. Prakash, N., Gosain, A.: An approach to engineering the requirements of data warehouses. *Requirements Engineering* 13, 49-72 (2008)
14. Giorgini, P., Rizzi, S., Garzetti, M.: GRAnD: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems* 45, 4-21 (2008)
15. Mazón, J.-N., Pardillo, J., Trujillo, J.: A model-driven goal-oriented requirement engineering approach for data warehouses. In: *Advances in Conceptual Modeling—Foundations and Applications*, pp. 255-264. Springer, (2007)
16. Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J.: Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems* 8, 203-236 (2004)
17. Yu, E.S.-K.: *Modelling strategic relationships for process reengineering*. University of Toronto (1995)
18. Luján-Mora, S., Trujillo, J., Song, I.-Y.: A UML profile for multidimensional modeling in data warehouses. *Data & Knowledge Engineering* 59, 725-769 (2006)
19. Cravero Leal, A., Mazón, J.N., Trujillo, J.: A business-oriented approach to data warehouse development. *Ingeniería e Investigación* 33, 59-65 (2013)
20. Maté, A., Trujillo, J., Eric, S.: Aligning Data Warehouse Requirements with Business Goals. In: *iStar*, pp. 67-72. (2013)
21. Barone, D., Yu, E., Won, J., Jiang, L., Mylopoulos, J.: Enterprise modeling for business intelligence. In: *Lecture Notes in Business Information Processing*, pp. 31-45. Springer Berlin Heidelberg, (2010)
22. Horkoff, J., Barone, D., Jiang, L., Yu, E., Amyot, D., Borgida, A., Mylopoulos, J.: Strategic business modeling: representation and reasoning. *Softw Syst Model* 13, 1015-1041 (2014)
23. Golfarelli, M.: From User Requirements to Conceptual Design in Data Warehouse Design. In: *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction*. (2010)
24. Porter, M.E.: *Competitive Strategy: Techniques for Analyzing Industries and Companies*. Free Press (1980)
25. Chalal, R., Boukrara, A., Saddok, M., Guiri, S.: A model for the acquisition of competitor information for strategic decision support. In: *ISKO-France*. (2013)