# A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses*

Jose-Norberto Mazón, Jesús Pardillo, and Juan Trujillo

Dept. of Software and Computing Systems
University of Alicante, Spain
{jnmazon,jesuspv,jtrujillo}@dlsi.ua.es

**Abstract.** The development of a data warehouse has been traditionally guided by an in-depth analysis of the underlying operational data sources, thus overlooking an explicit development phase in which information requirements of decision makers are addressed. This scenario has prompted that the deployed data warehouse often fails in delivering the expected support of the decision making process. To overcome this problem, we propose to use the *i** modeling framework and the *model driven architecture* (MDA) in order to describe (i) how to model goals and information requirements for data warehouses, and (ii) how to derive a conceptual multidimensional model that provides the required information to support the decision making process.

## 1 Introduction

*Data warehouse* (DW) systems provide decision makers with information related to a business process. This information is useful for decision makers to fulfil their goals in order to improve the business process. Both practitioners and researchers agree that the development of these systems must be based on a conceptual *multidimensional* (MD) model [1] that allows designers to easily structure information into facts (which contain interesting measures of a business process) and dimensions (which represent the context for analyzing the measures of a business process). Since the DW integrates several operational data sources, the development of conceptual MD models has been traditionally guided by their detailed analysis [2]. However, several studies [3,4] have pointed out that most of these conceptual MD models fail in addressing the required information as a result of a poor communication between DW developers and decision makers. Actually, information needs cannot be understood by only analyzing the operational data sources, and a requirement analysis stage is needed in order to model the information requirements of decision makers and derive a suitable conceptual MD model [5,6]. Furthermore, this stage should be based on a *goal-oriented*

* This work has been partially supported by the METASIGN (TIN2004-00779) project and by the FPU grants AP2005-1360 and AP2006-00332 from the Spanish Ministry of Education and Science, and by the DADS (PBC-05-012-2) project from the Castilla-La Mancha Ministry of Education and Science (Spain).

J.-L. Hainaut et al. (Eds.): ER Workshops 2007, LNCS 4802, pp. 255–264, 2007.

*requirement engineering* (GORE) framework since (i) the DW aims at providing adequate information to support the decision making process, thus helping to fulfil goals of an organization [1], and (ii) information requirements for DWs are difficult to specify from scratch [4], since decision makers often only express general expectations about which goals the DW should support.

Therefore, in this paper, we propose a GORE approach for modeling organizational goals that the DW supports and relating them to information requirements. To this aim, we have used the profiling mechanism of the *unified modeling language* (UML) [7] to adapt the *i** modeling framework [8,9] to requirement analysis in DWs according to our previous work [10]. On the other hand, the success of a DW project highly depends on the acceptance of the DW as a valuable resource for the organization [11,12]. To assure this acceptance, the derived conceptual MD model should represent the MD elements that satisfy information requirements and goals, in such a way that the decision makers understand the purpose of the DW. To accomplish this, our GORE approach is integrated into a *model driven architecture* (MDA) [13] framework for the development of DWs that has been described in [14]. This framework is based on defining a *computation independent model* (CIM) which addresses goals and information requirements, a *platform independent model* (PIM) to specify MD properties at the conceptual level, and a *platform specific model* (PSM) tailored to a specific database technology. Following these considerations, our GORE approach for DWs is used to define a CIM, while a PIM for MD modeling is derived by establishing a formal transformation between these models via the *query/view/transformation* (QVT) language [15]. The main advantage is that the conceptual MD model, represented in a PIM, meets every goal and information requirement defined in the CIM, since links between elements of both models are implicitly created with the execution of each QVT transformation. Therefore, QVT provides mechanisms to assure traceability in such a way that changes into a CIM are propagated to a PIM by re-executing the transformation.

The remainder of this paper is structured as follows: related work is described in the next section. Our approach for requirement analysis for DWs is presented in section 3. How to specify a conceptual MD model for DWs is defined in section 4. Section 5 describes an example of applying our approach. Finally, in section 6, we present our conclusions and sketch some future work.

## 2 Related Work

Only few approaches have considered requirement analysis as a crucial task in early stages of the DW development. In [4], a method is proposed in order to both determine information requirements of DW users and match these requirements with the available data sources. The work in [16] presents the *data warehouse requirements definition* (DWARF) approach that adapts traditional requirements engineering process for requirements definition and management of DWs. The approach described in [5] focuses on describing a requirement elicitation process for DWs by identifying goals of the decision makers and the required information that supports the decision making process. Finally, in [17], the authors present a goal-oriented framework to model requirements for DWs, thus obtaining a conceptual MD model from them by using a set of guidelines.

Unfortunately, these approaches present one main drawback, since they do not provide formal mechanisms to assure the traceability between the requirement model and the conceptual MD model, which is a desirable property just as stated in [18]. At most, these approaches only provide a set of informal guidelines to derive a conceptual MD model from requirements. To overcome this problem, we propose to use a GORE approach within our MDA framework for DWs.

## 3 Requirement Analysis for Data Warehouses

A requirement analysis stage for DWs aims at obtaining informational requirements of decision makers [5], which are related to interesting measures of business processes and the context for analyzing these measures [19]. However, decision makers often ignore how to suitably describe information requirements, since they are rather concerned about goals which the DW helps to fulfil. Therefore, a requirement analysis phase for DWs should start discovering goals of decision makers. Afterwards, the information requirements will be easier discovered from these goals. Finally, information requirements will be related to the required MD concepts, i.e. the measures of the business process or the context for analyzing these measures. Within our MDA approach for the development of DWs, both goals and information requirements have to be modeled in a CIM by using a UML profile for the $i*$ modeling framework[1]. From the defined CIM, a conceptual MD model which contains the necessary elements to achieve information requirements can be derived in a PIM.

### 3.1 Goals and Information Requirements for Data Warehouses

GORE is concerned about modeling goals, thus obtaining user requirements by following a refinement process [20]. To ease the task of discovering and eliciting goals and requirements for DWs, we have defined a classification of the different kind of goals that decision makers expect to fulfil with the envisaged DW. We consider three kind of goals, depending on their level of abstraction:

- **Strategic goals** represent the highest level of abstraction. They are main objectives of the business process. They are thought as changes from a current situation into a better one. For example: "increase sales", "increase number of customers", "decrease cost", etc. Their fulfilment causes an immediate benefit for the organization.
- **Decision goals** represent the medium level of abstraction. They try to answer the question: "how can a strategic goal be achieved?", and they aim to take the appropriate actions to fulfil a strategic goal. For example

[1] The $i*$ framework is used because the DW implies many organizational units and it can often involve dependencies among many kind of users with different goals.

"determine some kind of promotion" or "open new stores". Their fulfilment only causes a benefit for the organization if it helps to reach strategic goals, since decision goals only take place within the context of a strategic goal.
– **Information goals** represent the lowest level of abstraction. They try to answer the question: "how can decision goals be achieved in terms of information required?", and they are related to the information required by a decision goal to be achieved. For example "analyze customer purchases" or "examine stocks". Their fulfilment helps to achieve decision goals and they only happen within the context of a decision goal.

These goals form a hierarchy that can de defined via two well-known strategies [20]: discovering goals by refinement or discovering goals by abstraction. The former consists on asking "how" questions over goals already identified ("how can this goal be satisfied?"), while the latter is driven by "why" questions ("why is this goal useful?"). Once this hierarchy of goals is defined, information requirements can be directly obtained from the information goals. Later, the different MD elements, such as *facts* or *dimensions* will be discovered from these information requirements in order to specify the corresponding conceptual MD model of the DW.

### 3.2 Designing a CIM for Data Warehouses

A CIM is a view of a system from a computation independent viewpoint [13], thus taking into account the business environment of the system in order to improve the communication between decision makers and DW developers. Goals and information requirements for the DW are modeled in a CIM by using an adaptation of the $i^*$ modeling framework [8,9]. This adaptation is based on two extensions of UML [7] (see Fig. 1): (i) a profile for $i^*$, in order to integrate it within our MDA framework; and (ii) a profile which adapts $i^*$ to the DW domain. Due to space constraints, we only show an overview of the designed UML profiles.

The $i^*$ modeling framework [8,9] provides mechanisms to represent actors, their dependencies, and structuring the business goals that the organization

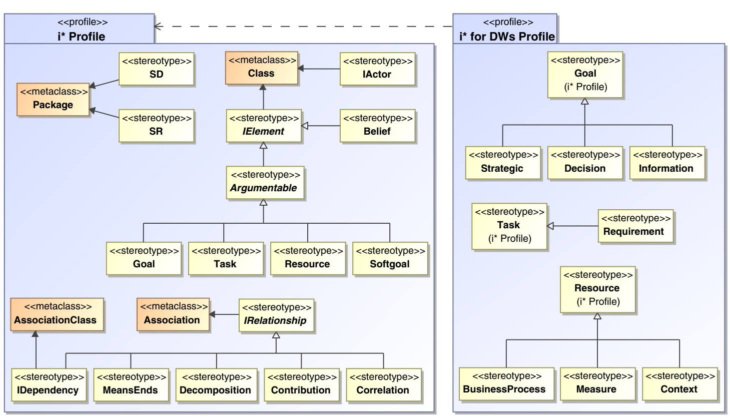


**Fig. 1.** Overview of the profiles for $i^*$ modeling in the DW domain

pretends to achieve. This framework establishes two models: the *strategic dependency* (SD) model for describing the dependency relationships among various actors in an organizational context, and the *strategic rationale* (SR) model, used to describe actor interests and concerns, and how they might be addressed. From now on, we focus on describing the SR models to specify goals and information requirements of decision makers[2].

The SR model (modeled with the *SR* stereotype and represented as ) provides a detailed way of modeling internal intentional elements and relationships of each actor (*IActor*, ). In our CIM, we use intentional elements such as goals (*Goal*, ), tasks (*Task*, ), resources (*Resource*, ); and intentional relationships such as means-end (*MeansEnd*, ) representing alternative ways for fulfilling goals, or task-decomposition (*Decomposition*, ) representing the necessary elements for a task to perform. Due to space constraints, we refer reader to [8,9] for a further explanation of *i**.

In order to define SR models for DWs, goals, tasks, and resources are represented as intentional elements for each decision maker. In a CIM, goals of decision makers are defined by using the *Strategic*, *Decision*, and *Information* stereotypes by specializing the previously defined *Goal* stereotype; and intentional means-end relationships (*MeansEnds*, ) between them. From information goals, information requirements (*Requirement*) are derived and represented as tasks. All of the described modeling elements are designed in the *i* Profile* (sketched in Fig. 1).

Furthermore, the requirement analysis for DWs needs some MD concepts to be added (in the sense of [17]). Therefore, the following concepts are added as resources in the CIM: business processes related to the goals of decision makers (*BusinessProcess* stereotype), relevant measures related to information requirements of decision makers (*Measure*), and contexts needed for analyzing these measures (*Context*). Additionally, foreseen relations between context of analysis are modeled. For instance, the city and the country contexts are related because cities can be aggregated in countries. For modeling these relationships, we use the (shared) aggregation relationship of UML (*Association* UML metaclass, represented as ).

In summary, several steps must be followed to properly define a CIM: (i) discovering the intentional actors (i.e. decision makers), thus defining SR models for each one, (ii) discovering their goals according to the classification described in Sect. 3.1, (iii) deriving information requirements from information goals, and (iv) obtaining the MD concepts related to the information requirements.

## 4 Obtaining a Conceptual MD Model

Once goals and information requirements are specified in a CIM, a conceptual MD model that supports them must be derived in a PIM. Within our MDA

[2] Dependencies among different decision makers would be defined in a SD model, but for the sake of clarity this issue is out of the scope of this paper.

approach we apply several QVT transformation rules to perform this task, thus assuring traceability between information requirements and the necessary MD elements related to them.

### 4.1 A PIM for Data Warehouses

A PIM describes a system hiding the necessary details related to a particular technology [13]. This point of view corresponds to the representation of MD elements at the conceptual level, independently from the platform in which the DW will be implemented. Our PIM is based on our UML profile for MD modeling presented in [21]. Although in this section we focus on describing a subset of this UML profile, an overview of the whole profile (showing stereotypes and extended metaclasses) is given in Fig. 2.

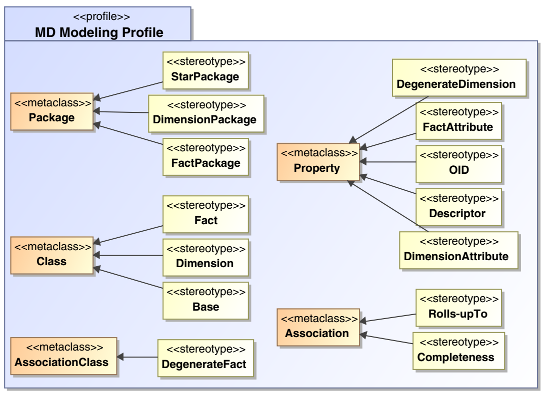


**Fig. 2.** Overview of the UML profile for MD modeling of DWs

This profile contains the necessary stereotypes in order to elegantly represent main MD properties at the conceptual level by means of a UML class diagram in which the information is clearly organized into facts and dimensions. These facts and dimensions are modeled by *Fact* (represented as ) and *Dimension* ( ) stereotypes, respectively. Facts and dimensions are related by shared aggregation relationships (the *Association* UML metaclass). While a fact is composed of measures or fact attributes (*FactAttribute*, FA), with respect to dimensions, each aggregation level of a hierarchy is specified by classes stereotyped as *Base* (B). Every *Base* class can contain several dimension attributes (*DimensionAttribute*, DA) and must also contain a *Descriptor* attribute (*Descriptor*, D). An association stereotyped as *Rolls-UpTo* between *Base* classes specifies the relationship between two levels of a classification hierarchy. Within it, role *R* represents the direction in which the hierarchy rolls up, whereas role *D* represents the direction in which the hierarchy drills down.

### 4.2 From CIM to PIM

A QVT transformation has been developed to derive the corresponding PIM from the CIM. A QVT transformation is composed of several relations that must hold between model elements of a set of candidate models (source and target models). Each of these relations is defined by using the elements described in Fig. 3.

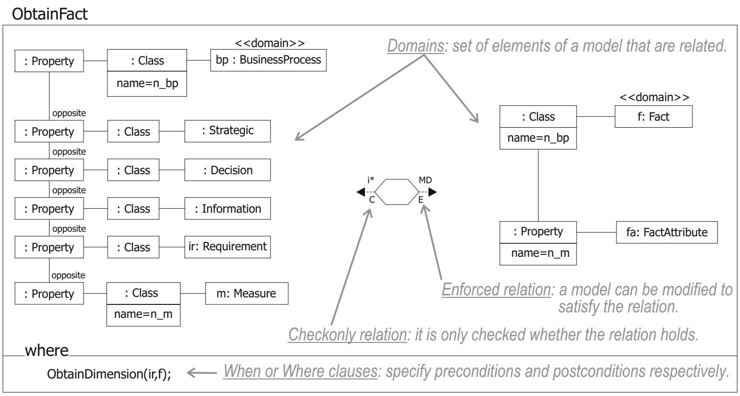


**Fig. 3.** QVT relation to obtain facts and their attributes

Due to space constraints, we only show one of the designed QVT relations (*ObtainFact*, Fig. 3). In this relation, the source domain is a set of elements of the CIM that represents the hierarchy of strategic, decision, and information goals of decision makers. This hierarchy is build on *Property* and *Class* (stereotyped as *Strategic*, *Decision*, or *Information*) UML metaclasses. A information requirement is also represented by using the *Requirement* stereotype. Finally, a strategic goal is related to a business process (*BusinessProcess*), while the information requirement is related to measures (*Measure*). The *ObtainFact* relation enforces the following set of elements in the PIM: a fact (*Fact* stereotype on *Class* instances) containing one fact attribute (*FactAttribute* on *Property* ones). Once this relation holds, the *ObtainDimension* relation must be carried out (according to the *where* clause) in order to obtain every required dimension and their hierarchies. After applying the QVT transformation, we obtain a PIM that provides the required information to support the decision making process.

## 5 Sample Application of Our Approach

In this section, we provide an example of our approach inspired from the case study presented in [21]. In this case study, a company sells automobiles across several countries. Therefore, we focus on the automobile sales business process which is related to one main actor, the sales manager, via the strategic goal *"increase automobile sales"*. From this strategic goal, three different decision goals are derived: *"decrease sale price"*, *"determine promotion according to a country"*, and *"give incentive to salespersons"*. From each of these decision goals the following information goals have been obtained: *"analyze automobile price"*, *"analyze automobile sales"*, and *"study salesperson sales"*. The derived information requirements are as follows: *"analyze automobile sale price"*, *"analyze the amount of sold automobile by customer age"*, *"analyze total amount by customer city and country"*, and *"analyze total amount by salesperson and date"*.

Each of these elements are defined in a SR model that represents a CIM according to our UML profile for $i*$ (see Fig. 4). In this CIM, strategic, decision, and information goals are represented as goals, and information requirements correspond to tasks. Furthermore, the necessary measures and contexts of analysis are associated to the information requirements as resources. Specifically, the measures are *"quantity"*, *"price"*, and *"total"*, and the elements that represent the context of analysis are *"salesperson"*, *"date"*, and *"automobile"*. *"Customer"*, *"age"*, *"city"*, and *"country"* also represent the context of analysis, and they are related each other, since they are useful for aggregating the *"customer"* data.

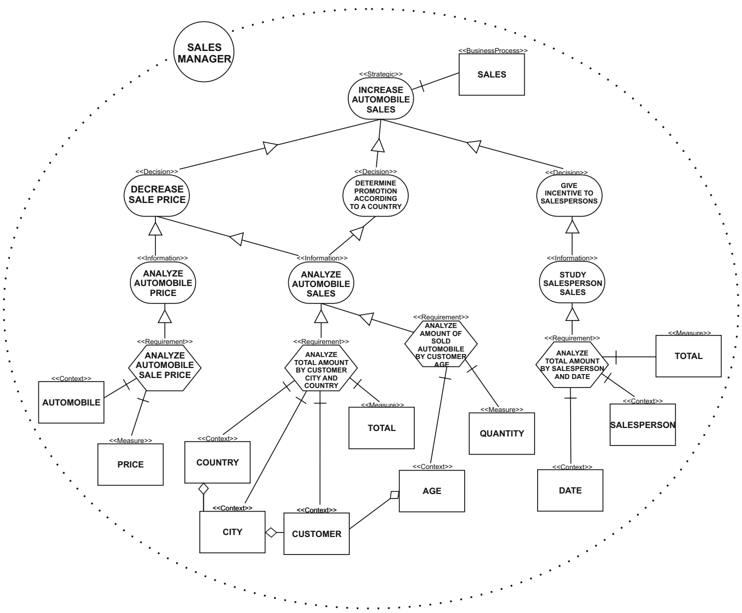


**Fig. 4.** Defined CIM for our example

From the defined CIM, the QVT transformation is applied to obtain a PIM (i.e. the conceptual MD model, Fig. 5) that delivers enough information in a suitable way to accomplish the information requirements and goals of decision makers. Table 1 overviews which MD elements are created in the PIM and their relations with information requirements. For instance, when the *ObtainFact* relation (Fig. 3) is executed, it takes the previously defined CIM as an input to create certain MD elements which conform the PIM. This relation would create a *Sales* fact related with the *"sales"* business process, and following the hierarchy of goals *"increase automobile sales"*– *"determine promotion according to a country"*– *"analyze automobile sales"*– *"analyze total amount by customer city and country"*, the relation would find the *"total"* measure and would create the corresponding fact attribute associated with the *Sales* fact. Additionally, the *where* clause of the relation would be executed, then dimensions and their hierarchies would be also created from the related contexts and aggregations modeled in the CIM. The resulting PIM is shown in Fig. 5.

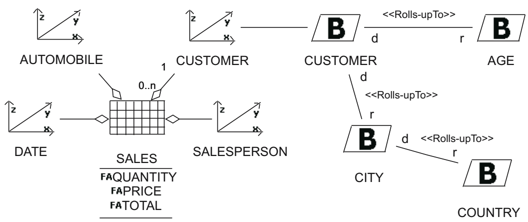


**Fig. 5.** Obtained PIM for our example

**Table 1.** Information requirements and their corresponding MD elements

| Information requirement | Fact attribute | Base or dimension |
|---|---|---|
| Analyze automobile sale price | Price | Automobile |
| Analyze the amount of sold automobile by customer age | Quantity | Age |
| Analyze total amount by customer city and country | Total | Customer, city, country |
| Analyze total amount by salesperson and date | Total | Date, salesperson |

## 6 Conclusions and Future Work

Modeling goals and requirements in DW projects is crucial for deploying a successful DW that properly supports the decision making process. In this paper, we have proposed a GORE approach within our MDA framework for the development of DWs. First, a CIM is specified by using the *i** modeling framework in order to model goals and information requirements for DWs. Then, the conceptual MD model of the DW is derived from this CIM into a PIM by using some QVT transformation rules. This PIM supports the goals that the decision makers plan to achieve with the DW. Our immediate future work comprises several tasks, such as adding quality measures to *i** models to analyze the understandability of *i** diagrams for DW developers, or adding softgoals in order to gather non-functional requirements apart from information requirements.

## References


1. Kimball, R., Ross, M.: The Data Warehouse Toolkit. Wiley & Sons, Chichester (2002)
2. Inmon, W.: Building the Data Warehouse. Wiley & Sons, Chichester (2002)
3. Schiefer, J., List, B., Bruckner, R.: A holistic approach for managing requirements of data warehouse systems. In: Americas Conf. on Information Systems, pp. 77–87 (2002)
4. Winter, R., Strauch, B.: A method for demand-driven information requirements analysis in data warehousing projects. In: HICSS (2003)
5. Prakash, N., Singh, Y., Gosain, A.: Informational scenarios for data warehouse requirements elicitation. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 205–216. Springer, Heidelberg (2004)
6. Rizzi, S., Abelló, A., Lechtenbörger, J., Trujillo, J.: Research in data warehouse modeling and design: dead or alive? In: DOLAP, pp. 3–10 (2006)
7. OMG: Unified Modeling Language Specification 2.0, `http://www.omg.org/cgi-bin/doc?formal/05-07-04`
8. Yu, E.: Modelling Strategic Relationships for Process Reenginering. PhD thesis, University of Toronto, Canada (1995)
9. Yu, E.: Towards modeling and reasoning support for early-phase requirements engineering. In: RE 1997, pp. 226–235 (1997)
10. Mazón, J.N., Trujillo, J., Serrano, M., Piattini, M.: Designing data warehouses: from business requirement analysis to multidimensional modeling. In: REBNITA (2005)
11. Chenoweth, T., Corral, K., Demirkan, H.: Seven key interventions for data warehouse success. Commun. ACM 49(1), 114–119 (2006)
12. Gam, I., Salinesi, C.: A requirement-driven approach for designing data warehouses. In: REFSQ (2006)
13. OMG: MDA Guide 1.0.1, `http://www.omg.org/cgi-bin/doc?omg/03-06-01`
14. Mazón, J.N., Trujillo, J.: An MDA approach for the development of data warehouses. Decision Support Systems doi:10.1016/j.dss.2006.12.003
15. OMG: MOF 2.0 Query/View/Transformation, `http://www.omg.org/cgi-bin/doc?ptc/2005-11-01`
16. Paim, F.R.S., Castro, J.: DWARF: An approach for requirements definition and management of data warehouse systems. In: RE 2003, pp. 75–84 (2003)
17. Giorgini, P., Rizzi, S., Garzetti, M.: Goal-oriented requirement analysis for data warehouse design. In: DOLAP 2005, pp. 47–56 (2005)
18. Stefanov, V., List, B.: Business metadata for the datawarehouse - weaving enterprise goals and multidimensional models. In: EDOC 2006 Workshops (2006)
19. Romero, O., Abelló, A.: Multidimensional design by examples. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2006. LNCS, vol. 4081, pp. 85–94. Springer, Heidelberg (2006)
20. van Lamsweerde, A.: Goal-oriented requirements enginering: A roundtrip from research to practice. In: RE 2004, pp. 4–7 (2004)
21. Luján-Mora, S., Trujillo, J., Song, I.Y.: A UML profile for multidimensional modeling in data warehouses. Data Knowl. Eng. 59(3), 725–769 (2006)