Requirements Engineering for Data Warehouses

Azadeh Nasiri*,**, [1] Esteban Zimányi*, Robert Wrembel** [2]

*Department of Computer and Decision Engineering (CoDE)
Université Libre de Bruxelles
{azadeh.nasiri, ezimanyi}@ulb.ac.be
**Institute of Computing Science,
Poznan University of Technology
{azadeh.nasiri, Robert.Wrembel}@cs.put.poznan.pl

Abstract. Data Warehouses (DWs) aim at supporting the decision-making process of an organization. In the Requirements Engineering (RE) domain, several methods were proposed for the development of DWs, most of them based on the Goal-Oriented Requirements Engineering (GORE) approach. However, there is not yet a comprehensive and unified perspective of the various methods proposed. In this paper, a coherent view of the GORE approach for the development of DWs is presented, by classifying existing methods according to the decision-making process, integrating them together, and linking modeling and analysis techniques throughout the overall process. The result of our study is an integrated GORE-based method for the development of DWs. We illustrate the method with a concrete example from the health care sector.

1 Introduction

Business Intelligence (BI) systems deliver the capability of data analysis in order to contribute to the decision-making process. A Data warehouse (DW) is a fundamental component of BI systems. A DW aggregates data from different data sources and specifically structures them to be used in BI systems. To develop a transaction-oriented system, we often need to take into account the requirements of how to perform automatically the repetitive operations of an organisation. However, the analytical requirements supporting the decision-making process in an organisation need to be captured in order to develop a DW. Such requirements are not easy to elicit and specify, so that DWs are sometimes developed based on an incomplete and inconsistent set of requirements, causing many of BI projects to fail. Therefore, the success of the system under the development can be strongly affected by Requirements Engineering (RE) (Prakash and Gosain, 2008). RE is defined as the process of discovering the needs of involved

[1] This research is funded by the Erasmus Mundus Joint Doctorate IT4BI-DC.
[2] Polish National Science Center (NCN), grant No. 2011/01/B/ST6/05169.

stakeholders and supporting those needs by modeling and documenting them in a form that is analysable and communicable.

In order to support RE for DW systems, multiple methods have been proposed in the literature, which are based mainly on the Goal-Oriented Requirements Engineering (GORE) approach. The GORE approach uses goals for eliciting, modeling, analyzing, negotiating, and modifying requirements. However, GORE-based methods proposed for RE of DWs, are not unified; they have been developed based on different principles targeting various RE problems in the DW domain. That made it difficult to give a comprehensive GORE perspective in the DW domain, in which a complete and consistent set of the DW requirements are taken into account. To address this problem, we propose a GORE-based method for DWs in this paper.

Our methodology to develop this method is based on the method engineering approach. Method engineering is the discipline to construct a new method from existing ones. The components of the new method are obtained from existing methods. They are combined into a consistent structure as an activity model (comprising activities, activity structure, and activity sequence) with certain modeling techniques and types of documents (Winter and Strauch, 2004). Our methodology includes the following steps:

- The existing GORE-based methods for BI are classified based on their contributions to each phase of the decision-making process. Then, the components of the proposed method are adapted from the classified GORE-based methods.
- The basis to construct our method is the decision-making process containing three phases. Therefore, the appropriate components are selected, among the classified methods (mentioned in the previous item), based on the best coverage of capturing the requirements of the decision-making process of each phase.
- For each phase, guidelines are proposed including necessary activities for abstracting the decision-making process, supported by proper goal modeling and analysis techniques.
- Guidelines containing activities, goal modeling and analysis technique, mentioned in the previous item, are integrated all across the decision-making phases.

The advantages of the proposed method are as follows:

- We involve the decision-making process in the early phase of the system development where the requirements are captured. It contributes to the problem addressed in the BI literature, saying that the capabilities offered by DW systems are not matched with the decision-making process.
- According to our knowledge, there is no method in the RE literature for DWs that covers all phases of the decision-making process. Our approach covers them all. The majority of proposed methods focus mainly in the last phase of the decision-making process, where the information requirements of the DW are obtained from the goal model. However, our proposed method covering all the decision-making phases, pays also attention to the analysis of the organisational context, where different decision alternatives are constructed.
- Using the method engineering gives us the opportunity to take advantage of the contribution of existing research works in the RE for BI systems.

The paper is organised as follows. Section 2 outlines the GORE approach according to the decision-making process using a running example from the health care sector. Section 3 concludes the paper and points to future works.

2 GORE and the Decision-Making Process

The GORE approach uses goals for eliciting, modeling, analyzing, negotiating, and modifying requirements. In this approach, a goal is an objective, which the system under development should achieve. Goal identification is not necessarily an easy task. Goals may be formulated at different abstraction levels, ranging from high-level, strategic concerns to low-level, technical ones. This suggests a structuring mechanism for identifying goals. Goal refinement trees provide such a mechanism. A goal refinement tree or a goal model is a graphic representation of the reduction of goals to sub-goals that elaborate how the goal is achieved. Goal modeling supports heuristic, qualitative or formal reasoning schemes during RE.

According to Herbert (1960), the decision-making process consists of three phases: (a) searching for conditions that call for decision-making, (b) analysing possible courses of actions, and (c) selecting a proper course of action from available options. In the following subsections, we explain how the GORE approach can reflect these phases of the decision-making process. Each subsection contains a phase of the decision-making process including: the main objective of the phase, the main existing approaches proposed in the literature, and their application to a running example. We use the following case study from the health care domain as running example.

In the pharmaceutical industry, Adverse Events (AEs) are crucial in the assessment of drug, and consequently patient safety. According to World Health Organisation (WHO), an AE is an unexpected, or potentially harmful reaction resulting from the use of a prescribed medication. Many AEs are not discovered through limited pre-marketing clinical trials. Instead, they are only seen in long term, post-marketing surveillance of drug usage. In light of this, pharmaceutical companies typically need to collect related data and carry out analytics over these data to make well-informed decisions in order to avoid AEs. In this regard, we aim to analyse requirements of an analytical system through which a pharmaceutical company is able to answer the question of what happens regarding AEs and why it is happening, to drive decision-making. The target system should help decision-makers gain some descriptive insight regarding AEs occurrences. Briefly, the analytical system under development should enable the company to follow up its strategic goals relevant to AEs, to observe the current status regarding goals, and to find the factors influencing them. The company decided to develop a DW to meet the aforementioned requirements.

2.1 Phase 1: Conditions for Decision Making

Objectives In the context of BI, requirements are initiated by senior managers and executives through high-level strategic goals. The RE initial challenge is to find out what decisions needs to be made to fulfill a particular strategy. Deriving decisions from strategic goals is not a straightforward process. This needs understanding the cause-effect nature of decisions related to a certain strategic goal. Then, the strategic goal needs to be analysed in its business context to clarify these cause-effect relationships. For example, in a pharmaceutical company, the senior manager defines "Increase patient safety" as a high-level strategic goal. It is too general and needs to analyse the context of pharmaceutical industry in order to clarify what are the decisions needs to be taken to increase the probability of accomplishing this goal. The first step is to understand what can affects this strategy. In the context of the pharmaceutical industry, AEs occurrence is shown to affect patient safety. We still need to clarify what causes the occurrence of AEs in order to achieve the "Decrease AEs" goal. The lack of patient's knowledge of how to use a drug or low-quality drugs are some causes.

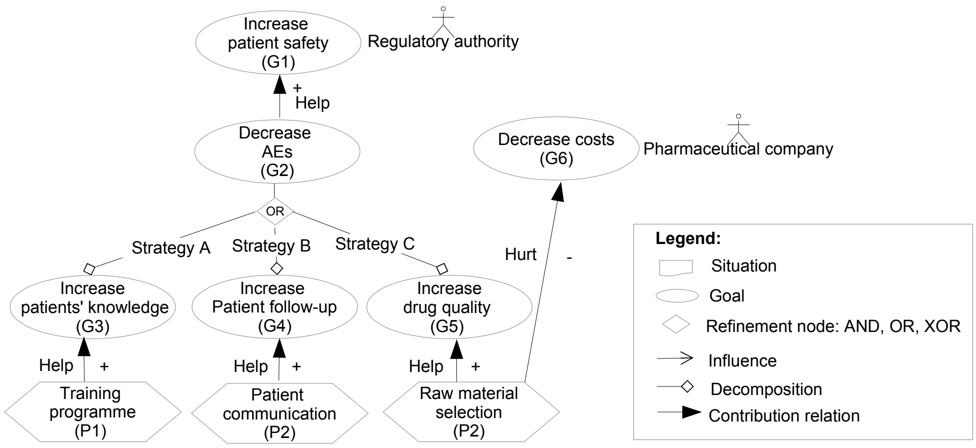

FIG. 1: Goal model representing a strategic goals and associated tasks

The strategic goal is translated into decisions that need to be made. The manager should now decide either to invest in patient's knowledge or drug quality, in order to decrease AEs and subsequently, increase patients' safety.

Proposed Approches RE aims to provide formal modeling foundations and proper representations of variables and the relationship among them, which are important for decisions required to be made in a business. Therefore, the final objective is to illustrate the context for the decision itself. To this end, a conceptual model of the business is constructed to give decision makers a comprehensive view of what they need to take into account to achieve strategic goals. This model captures business objectives, business processes, performance indicators, or any other business related concepts to drive strategic decision-making. Different approaches were proposed. For example, in (Pourshahid et al., 2009) the model is developed based on the Business Process Improvement (BPI) concept. The proposal in (Jiang et al., 2011) was inspired by SWOT analysis. The other proposal uses (Babar et al., 2010) a strategy map for for this purpose. However, it is important to note that proposing a detailed methodology of an appropriate business model is beyond the scope of RE. How to formally model and represent business concepts to abstract the context for the decision is the RE challenge.

The basis for all proposals is the goal model, proposed by one of the well-established GORE frameworks. The notion of a goal is used to represent the strategy. Goal refinement trees are used to abstract possible alternatives to satisfy strategic goals.

Running Example Inspired by (Pourshahid et al., 2009, 2011), we show how the cause-effect nature of the decision is modeled and evaluated. Adapted from the User Requirements Notation

(URN) framework, various business concepts like strategic goals, decision makers, and tasks (or processes), are used to represent the business model for the pharmaceutical company. Business concepts are linked by defined AND/OR decomposition and contributions links. They help to simulate cause-effect relationships in the decision context. The elements connected by the links, influence each other through contributions which are qualitative positive (make, help, some positive) and negative (break, hurt, some negative).

Fig. 1 illustrates some of the business concepts related to the pharmaceutical company example. Regulatory authority wants the pharmaceutical company to "Increase patient safety". This goal has been defined as a strategic goal for the company and has been refined to a lower-level goal of "Decrease AEs". This goal is broken down into two non-mutually exclusive goals of "Increase patient's knowledge" and "Increase drug quality". The task of "Training programme" can help to satisfy "Increase patient's knowledge". The "Raw material selection" task can satisfy "Increase drug quality" goal. However, the latter task can hurt the another strategic goal of the pharmaceutical company of "Decrease costs".

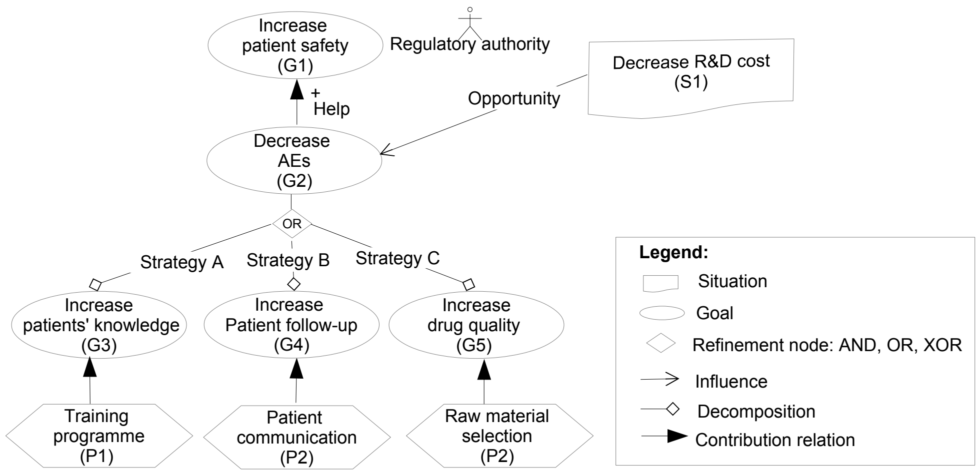

FIG. 2: Goal model of Fig. 1 incorporating SWOT analysis

Depending on various approaches taken to model the business, some new concepts might be added to the goal model. For example, in (Barone et al., 2011a,b; Jiang et al., 2011; Maté et al., 2012) new concepts called situation-based on the SWOT (strengths, weaknesses, opportunities, and threat) analysis, are involved in the goal model. As it was mentioned earlier, detailed methodology of the SWOT analysis is not our RE challenge. How to represent and analyse the new concept of situation (or any other business concepts) in the goal model is the challenge.

Fig. 2 illustrates how the concept of situation adapted from the SWOT analysis is incorporated with the goal model. For example, "Decrease R & D cost" is a potential opportunity for a pharmaceutical company. To model this situation, representative symbols are required. Analogously to the contribution relationships, there is an "influence" arrow representing the relationship between a situation and any other element in the goal model. Also, analogous to satisfaction levels for goals, there are occurrence levels for situations, that define the degree to which a situation occurs in the current circumstances.

2.2 Phase 2: Analysing Possible Courses of Actions

Objectives Business strategic goals were translated into decisions, and alternatives of any decision were determined by clarifying the cause-effect relationships of business concepts (Section 2.2). Now, RE aims to evaluate decision alternatives by quantifying or qualifying the effect of each alternative on the goal satisfaction. For example, in the pharmaceutical context, it was clarified that the manager has to decide between two alternatives: investing in patients' knowledge or investing in drug quality in order to decrease AEs and subsequently increase patient safety. Now, the RE aims to clarify which one of those alternatives has more effect on patient safety. To do that, RE needs to evaluate the satisfaction level of the strategic goal for a given alternative. Qualitative and quantitative evaluations are both covered.

Proposed Approches RE aims to provide a proper analysis foundation to evaluate the goal satisfaction. The analysis is conducted over the business conceptual model described in the previous part (see Section 2.2). This satisfaction value of goals can be presented through defining some criteria as well. In the two following subsections, we discuss how to conduct the analysis to evaluate the satisfaction level of goals, and then we show how this analysis differs when a criterion is defined to measure the satisfaction level.

2.2.1 Evaluating Goal Satisfaction

Goal models are the basis for this analysis, as the business conceptual model is built on the goal model (Section 2.2). As Fig. 2 illustrates, each branch in the goal refinement tree represents an alternative to satisfy the strategic goal. To see which alternative is the most effective, we need to evaluate the cause-effect relationships drawn in each branch of the goal refinement tree. Different proposals exist in the literature. Some papers use conventional goal evaluation methods based on the GORE framework. For example, (Pourshahid et al., 2009, 2011) used the URN conventional evaluation method. There are also proposals applying some decision making mechanisms for evaluation. For example, (Jiang et al., 2011), makes the influence diagram with goals and uses probability-based statistics to evaluate the satisfaction level of a goal.

Running Example Inspired by the approach in (Pourshahid et al., 2009, 2011) and based on "GRL evaluation strategies" defined in URN, a "what-if" analysis is conducted to evaluate the impact of an alternative solution on a high-level goal. GRL evaluation strategies enable modelers to assign initial satisfaction values to some of the elements of the goal model and propagate this information to the other elements through links. Links are assigned a contribution level by experts, defined as qualitative positive (make, help, some positive) or qualitative negative (break, hurt, some negative). Also, quantitative contribution levels on a scale going from -100 to +100 might be assigned.

Fig. 3 illustrates the result of an analysis for our example where the “Raw material selection” is picked to fulfill the strategic goal. The satisfaction value of this task is initialised to 100 (initialised elements are displayed with dashed contours). This eventually leads to a satisfied value of +25 for the goal of “Increase safety” and to an unsatisfied level of −25 for “Decrease cost” (−25 and +25 are the values resulted from the initiated 100 for the “Raw material selection”, which is propagated through links with quantitative contribution levels defined by experts). Colour-coding is used to highlight satisfaction levels, red for denied, yellow for neutral, and green for satisfied, which improves intuitive understanding.

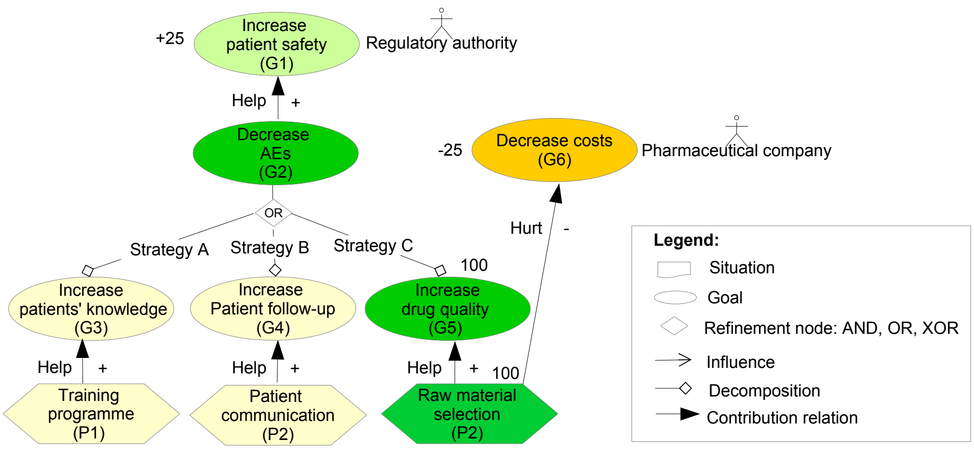

FIG. 3: Goal model with evaluation strategy

The resulting satisfaction values in top level goals should be used to compare different strategies and find suitable trade-offs rather than to be interpreted as the certain percentage value for the satisfaction of that top-level goal.

2.2.2 Evaluating Goal Satisfaction with Criteria

The goal satisfaction value can be defined using some criteria. The RE literature has used the concept of a Key Performance Indicator (KPI) for this purpose. A KPI is defined as a metric that evaluates performance with respect to certain objectives. The subject of a KPI is a particular feature or quality of an element in the business environment. In the RE context, a KPI evaluates the progress or degree of fulfillment of goals. For example, the number of AEs experienced by patients indicates the satisfaction level of “Increase AEs” as a goal. KPIs are modeled using the conventional GORE frameworks. Standard goal models of a GORE framework is enriched with a new element representing a KPI.

Now, it is necessary to discuss techniques that propagate values of KPIs from a lower level to an upper level in a hierarchy. Based on how KPIs are calculated, they are called atomic or composite (Maté et al., 2012; Barone et al., 2011a). Atomic KPIs are those whose value is obtained from data sources. Composite KPIs are those whose value is obtained from other KPIs (called component KPIs in this paper). For example, to evaluate “Increase patient safety”, we need to know the value of "Number of well-treated patient after drug usage". This is a composite KPI because this value depends on "Number of AEs" which is defined as a component KPI. Notice that component KPIs are not necessarily atomic KPIs. For example, "Number of AEs" itself is a composite KPI which takes its current value from other component KPIs, such as "Number of training hours". The latter KPI obtains its value from available data sources, making it an atomic KPI. In this section, we outline techniques that propagate values of composite KPIs. In (Barone et al., 2011b) some techniques have been proposed to derive the value of a composite KPI such as: conversion factors, normalisation, and qualitative techniques.

Conversion Factor Technique This technique is used when a composite KPI does not share the same unit of measure with its components and a suitable conversion factor is available.

Running Example Considering the composite KPI "Number of AEs" and component KPI "Number of training hours", we need to convert the current value of "Number of training hours" into the number of avoidable AEs experienced by patients.

One possible conversion factor is to use the average of AEs experienced by patients due to their lack of knowledge. Assume this average has been 400 patients per year. Training doctors and pharmacists to increase the knowledge of patients has a considerable effect on this average. The average training time per year is 350 hours. These averages come from the previous data or the experience of experts. Therefore, for each training hour, we have 400 $\div$ 350 = 1.1 AEs (rounded to 1). We can consider this number as the conversion factor meaning for each training hour we would have one avoidable AE (a conversion factor typically is estimated based on a previous experience).

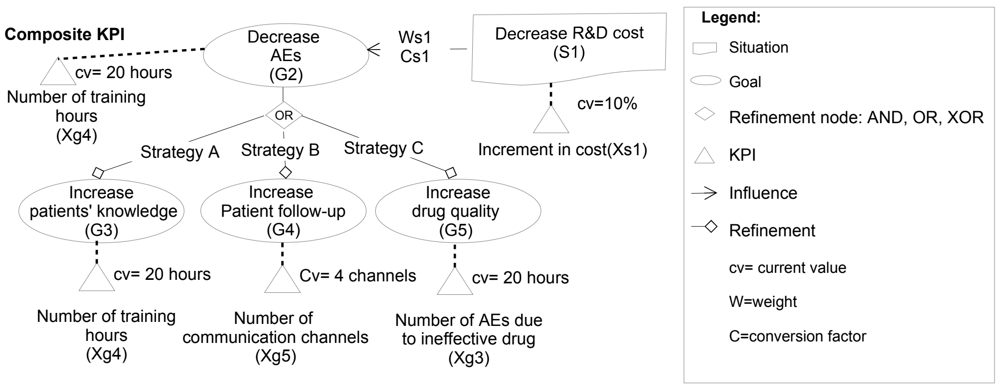

FIG. 4: Conversion factor technique for composite KPIs

When conversion factors are available, a formula containing the following components can calculate the value of the composite KPI: composite KPI: X^c_{gi}, component KPI: X_{gi}, situation-related component KPI: X_{si}, influence strength: W_{gi}, Situation-related influence strength: W_{si}, conversion factor: C_{gi}, situation-related conversion factor: C_{si}, expected value of a composite KPI: X^e_{gi}. The formula for the example in Fig. 4 is as follows:

$$X_{g2} = X^e_{g2} + W_{s1} \cdot C_{s1} \cdot X_{s1} + \sum_{i=3}^{i=5} W_{gi} \cdot C_{gi} \cdot X_{gi}$$

The formula comprises two major components which is influencing the value of the composite KPI: the first components refers to KPIs of situation (related to the SWOT analysis, see Section 2.2), and the second component is related to KPIs of sub-goals. X^e_{g2} is the expected value of AEs experienced by patients obtained from historical data or experts.

Normalisation Technique This approach is especially useful when there is no suitable conversion factor. In this technique, values of KPIs, are typically represented between a range of -100 to +100. Each KPI has a current value, which is evaluated against a set of parameters: target, threshold, and worst. The result of such evaluation is normalised, ranging between -100 to +100, called performance level. When the current value is between the target and the threshold value, the performance level is calculated with the normalisation function of (threshold $-$ current)/(threshold $-$ target) $*$ 100. When the current value is between the threshold and the worst value, the normalisation function becomes (threshold $-$ current)/(worst $-$ threshold) $*$ (-100), which results in a negative value. If the result is higher than 100, then it becomes 100 (if it is lower than -100, then it becomes -100).

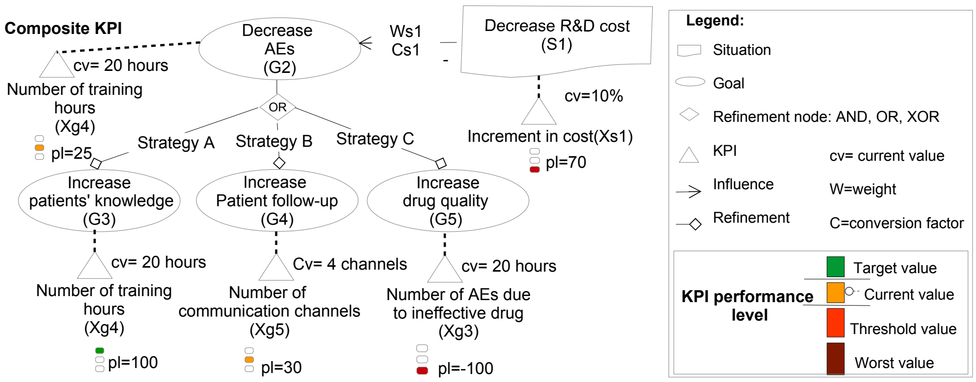

FIG. 5: Normalisation technique for composite KPIs

Running Example Fig. 5 illustrates how the performance level of a KPI is calculated using its current value and target, threshold, and worst parameters. A formula containing the following components can calculate the value of the composite KPI: performance level of composite KPI: PL(X^c_{gi}), performance level of component KPI: PL(X_{gi}), performance level of situation-related component KPI: PL(X_{si}), influence strength: W_{gi}, situation-related influence strength: W_{si}. The formula for the example in Fig. 5 is:

$$X_{g2} = CX_{s1} * PL(X_{s1} + max[PL(X_{g3}), PL(X_{g4}), PL(X_{g5})].$$

The KPI takes the maximum performance level among the sub-nodes, and sums such a level to the result obtained from multiplying the influence strength by the performance level of the component KPI "Increment R & D cost".

Qualitative Technique Unlike propagating performance levels in the normalisation technique, labels assigned to KPIs are propagated in a quantitative technique. Each KPI is associated with two variables: positive performance (per^+) and negative performance (per^-) evidence. They range from {F, P, N} (full, partial, none), such that F > P > N. For example, for KPI "Number of training hours" we have a current value of 4000. The associated target and the threshold values are 2000 and 6000, respectively. Therefore, by following the rule ($M \leq cv < t$), we can conclude that for the "Number of training hours" performance evidences are per^+="partial" and per^-="none".

A major difference between this technique with the normalisation and conversion factor techniques is that conflicts are allowed. It means that a KPI can have, at the same time, evidence of both positive and negative performance. To assign a current evidence to a KPI, the mapping rules, described in Table 1, are used.

$cv \geq t$	$M \leq cv < t$	$th \leq cv < M$	$w \leq cv < th$
per^+="Full" per^-="None"	per^+="Partial" per^-="None"	per^+="None" per^-="Partial"	per^+="None" per^-="Full"

TAB. 1: Mapping rules in qualitative techniques for composite KPIs (Barone et al., 2011b)

Required components to derive the qualitative value of a composite KPI from related component KPIs are as follows: composite KPI: X^c, component KPI: X^a, positive performance: per^+, negative performance: per^-, F="full", P="partial", N="none", F > P > N, influence strength-satisfied: (or), (+D), (-D), (++D), (- -D), influence strength-denied: (+S), (-S), (++S), (- -S).

$(X_i^a, X_j^a) \stackrel{and}{\longmapsto} X^c$	$X_i^a \stackrel{+S}{\longmapsto} X^c$	$X_i^a \stackrel{-S}{\longmapsto} X^c$	$X_i^a \stackrel{++S}{\longmapsto} X^c$	$X_i^a \stackrel{--S}{\longmapsto} X^c$
$\min\begin{cases} per^+(X_i^a) \\ per^+(X_j^a) \end{cases}$	$\min\begin{cases} per^+(X_i^a) \\ P \end{cases}$	N	$per^+(X_i^a)$	N
$\max\begin{cases} per^-(X_i^a) \\ per^-(X_j^a) \end{cases}$	N	$\min\begin{cases} per^+(X_i^a) \\ P \end{cases}$	N	$per^+(X_i^a)$

TAB. 2: Mapping rules in qualitative techniques for composite KPIs (Barone et al., 2011b)

Table 2 (adapted from qualitative reasoning with goal models (Giorgini et al., 2003)) helps to propagate the qualitative evidence for composite KPIs. The qualitative evidences are assigned to each component KPI using the mapping rule in Table 1. The next step is to rely on the propagation rules described in Table 2 to evaluate qualitatively the composite KPI (the second row in the table represents the (per^+) value and the third raw the (per^-) value of a composite KPI).

Running Example For "Number of AEs", we start to propagate three KPIs associated to the corresponding sub-goals by relying on the rules in the first column of Table 2. Therefore, for the per^+ and per^- variables, we need to choose the maximum and minimum respectively, among values of the component KPIs. The result is:

$$\begin{cases} per^+(\text{from-subnodes}) = \max[per^+(X_{g3}), per^+(X_{g4}), per^+(X_{g5})] = \text{Full} \\ per^-(\text{from-subnodes}) = \min[per^-(X_{g3}), per^-(X_{g4}), per^-(X_{g5})] = \text{None} \end{cases}$$

We also need to propagate the component KPI which refers to situation "Decrease R & D cost" related to goal "Increase AEs". To calculate its impact on the composite KPI, we rely on the second column rule in Table 2.

$$per^{+}(\text{from-influencer}) = \min \begin{cases} per^{+}(X_{s_1}) = \text{None} \\ P \end{cases} = \text{None}$$

$$per^{-}(\text{from-influencer}) = \min \begin{cases} per^{-}(X_{s_1}) = \text{Full} \\ P \end{cases} = \text{Full}$$

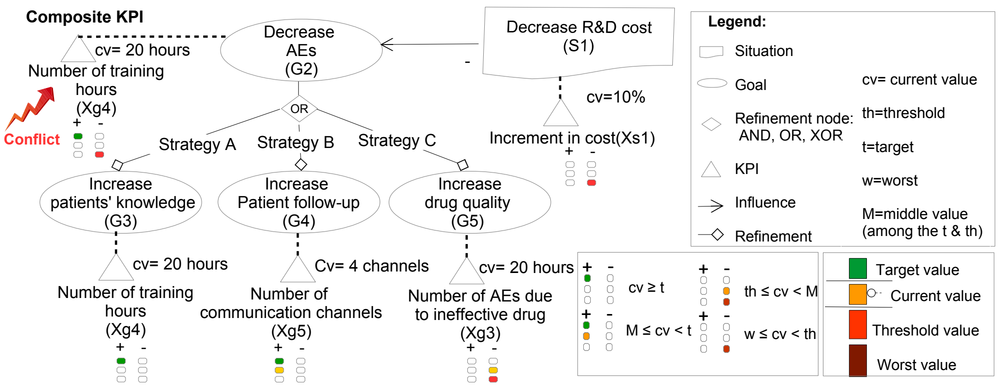

FIG. 6: Qualitative technique for composite KPIs

At the end, results obtained from component KPIs of sub-goals need to be combined with results obtained from influencers. The final result is illustrated in Fig. 6. A conflict is discovered for the composite KPI "Number of AEs" when values are propagated.

2.3 Phase 3: Selecting a Proper Course of Action

Objectives After discussing the evaluation mechanism of goal satisfactions, it is necessary to discuss information requirements in order to evaluate the satisfaction of goals. Generally speaking, RE discusses what data in which form is of particular interest for decision makers to store in DWs. There are various drivers to obtain the information including: business processes proposed by (Chowdhary et al., 2006; Frendi and Salinesi, 2003; Kimball, 1998), decision process proposed by (Winter and Strauch, 2004) as well as goals and objectives of an organisation proposed by (Bonifati et al., 2001; Gallardo et al., 2009; Ghezzi et al., 2008; Malinowski and Zimányi, 2006; Mazón et al., 2007a, 2005, 2007b; Prakash and Bhardwaj, 2012; Prakash and Gosain, 2008; Silva et al., 2012).

Proposed Approches Organisational goals are the most popular drivers to obtain the information requirements of a DW. Information is derived mainly using GORE frameworks. Goal models are adapted to eventually represent the information in the Multidimensional (MD) schema with elements of facts (center of analysis) and dimensions (context of analysis). We aim to derive this information from KPIs adapting guidelines provided by Mazón et al. (2005).

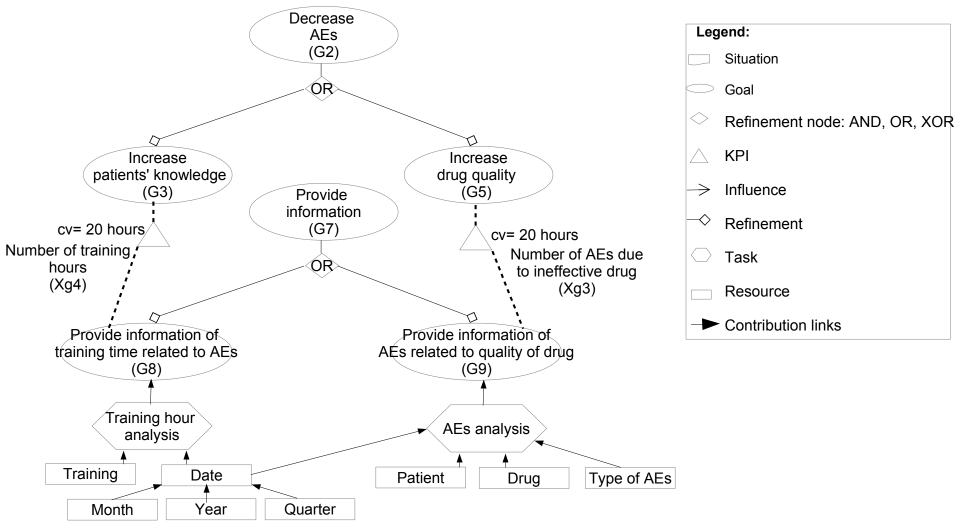

FIG. 7: Deriving information requirements from KPIs

The first step is to incorporate DW in the goal model as an actor. The goal of the "Data warehouse" actor is to provide information requirements. For each atomic KPI in the model, a goal is represented stating the information target of a KPI (typically aggregated measures). Measures need to be refined into tasks that calculate measures in some contexts. Tasks are associated with goals using contribution links. It is necessary to include the context as a new element in the model too. They are represented as sources attached to the tasks by decomposition links.

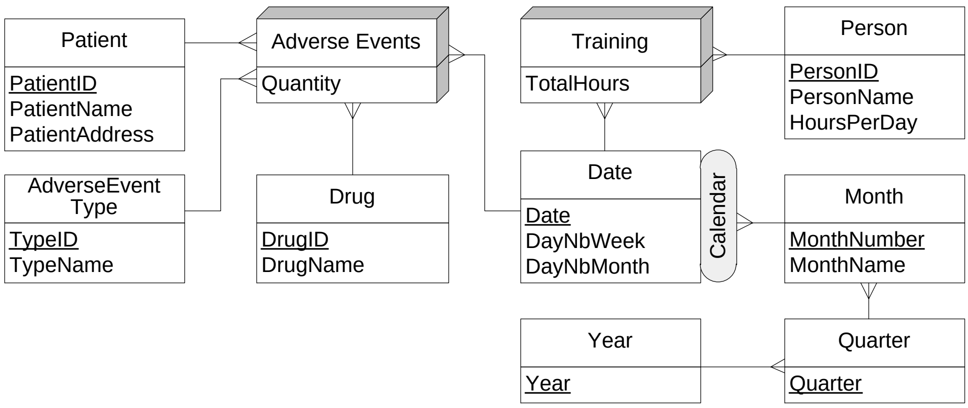

FIG. 8: Data warehouse schema

Running Example Given the KPI "Number of AEs due to lack of drug quality", a goal "Provide information of AEs related to quality of drug" is defined for the "Data warehouse" actor.

The goal is refined to the task of "AEs analysis", connected to the issue of drug quality. This means that we need to take into account only aggregation of AEs where a certain drug's quality caused AEs occurrence. Therefore, the context of analysis of AEs includes drug, patient, and type of AEs. These contexts are represented in the model as resources of information, and attached to the tasks by contribution links (see Fig. 7).

For KPI "Number of training hours", we need to see how training of pharmacists, doctors, and hospital personnel for increasing patients' knowledge, is related to AEs occurrence. A goal of "Provide information of training hours related to AEs" is defined and then refined to two tasks of "AEs analysis" and "Training hours analysis". The first task is the same task introduced for the KPI explained in the previous paragraph but, this time, we look for AEs caused by patients' lack of knowledge.

The second task looks for total training hours in a certain period so that the training and the date of training are considered as a context. These contexts can also be linked to "Number of AEs" task as they should be counted in a certain period. To conclude, resources are: "Drug", "Date", "Training", "Type of AEs", and "Patient". Within a context of analysis, there are several levels of aggregation, which are also represented as resources. For example, "Date" can be aggregated by month, quarter, and year (see Fig. 7). Now we can derive the multidimensional schema from the goal model in Fig. 7. For each task of the DW, a fact is created. Resources defining the context of analysis are dimensions in MD model. Finally, the levels of aggregation attached to resources are represented as hierarchies of dimension in the DW schema (see Fig. 8).

3 Conclusions

This paper studies the RE phase for DW development. To support RE for DWs, several methods were proposed in the literature. Authors of these methods have argued for the usefulness of the GORE approach. In this paper, we provide an overview of the existing GORE-based methods in the BI domain and proposed our method using method engineering concept. The components of the proposed method are adapted from existing GORE-based methods for BI systems. The appropriate components are selected based on the best coverage of abstracting the decision-making process. For each phase of the decision-making process, we proposed guidelines including necessary activities supported by proper goal modeling and analysis techniques in an integrated manner. Our method involves all phases of the decision-making process in the early stage of DW projects where the requirements are captured, as the final objective of DW projects is to help the decision-making process in the organisation. Besides, the proposed method takes advantages of the contribution of major existing research works in this domain.

We already covered RE for the static part of the DWs, where MD model is obtained. We plan to extend our method with the dynamic part of the DW, where the requirements of operations on the DW are captured. In this part the interaction of users with the DW are discussed. We also plan to evaluate the proposed method by applying it to a real case study.

Acknowledgements We would like to thank Dr. Miroslaw Ochodek from Poznan University of Technology for valuable discussions on literature overview in the area of requirements engineering.

References

Babar, A., D. Zowghi, and E. Chew (2010). Using goals to model strategy map for business IT alignment. In *Proc. of Int. Workshop on Business/IT Alignment and Interoperability*, pp. 16–30.

Barone, D., L. Jiang, D. Amyot, and J. Mylopoulos (2011a). Composite indicators for business intelligence. In *Proc. of Int. Conf. on Conceptual Modeling*, pp. 448–458. LNCS 6998.

Barone, D., L. Jiang, D. Amyot, and J. Mylopoulos (2011b). Reasoning with key performance indicators. In *Proc. of Conf. on the Practice of Enterprise Modeling*, pp. 82–96. LNCS 68.

Bonifati, A., F. Cattaneo, S. Ceri, A. Fuggetta, and S. Paraboschi (2001). Designing data marts for data warehouses. *ACM Transactions on Software Engineering and Methodology 10*(4), 452–483.

Chowdhary, P., G. Mihaila, and H. Lei (2006). Model-driven data warehousing for business performance management. In *Proc. of IEEE Int. Conf. on e-Business Engineering*, pp. 483–487.

Frendi, M. and C. Salinesi (2003). Requirements engineering for data warehousing. In *Proc. of Workshop on Requirements Engineering*, pp. 75–82.

Gallardo, J., G. Giacaman, C. Meneses, and Ó. Marbán (2009). Framework for decisional business modeling and requirements modeling in data mining projects. In *Proc. of Int. Conf. on Intelligent Data Engineering and Automated Learning*, pp. 268–275. LNCS 5788.

Ghezzi, C., M. Jazayeri, and D. Mandrioli (2008). GRAnD: a goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems 45*(1), 4–21.

Giorgini, P., J. Mylopoulos, E. Nicchiarelli, and R. Sebastiani (2003). Reasoning with goal models. In *Proc. of Int. Conf. on Conceptual Modeling*, pp. 167–181. LNCS 2503.

Herbert, A. (1960). *The new science of management decision*. Harper & Brothers.

Jiang, L., D. Barone, D. Amyot, and J. Mylopoulos (2011). Strategic models for business intelligence. In *Proc. of Int. Conf. on Conceptual Modeling*, pp. 429–439. LNCS 6998.

Kimball, R. (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. John Wiley & Sons.

Malinowski, E. and E. Zimányi (2006). Requirements specification and conceptual modeling for spatial data warehouses. In *Proc. of OTM Int. Workshop on the Move to Meaningful Internet Systems*, pp. 1616–1625. LNCS 4278.

Maté, A., J. Trujillo, and J. Mylopoulos (2012). Conceptualizing and specifying key performance indicators in business strategy models. In *Proc. of Conf. on the Center for Advanced Studies on Collaborative Research*, pp. 102–115. IBM Corp.

Mazón, J., J. Pardillo, and J. Trujillo (2007a). A model-driven goal-oriented requirement engineering approach for data warehouses. In *Proc. of Workshop on Advances in Conceptual*

Modeling–Foundations and Applications, pp. 255–264. LNCS 4802.

Mazón, J., J. Trujillo, and J. Lechtenbörger (2007b). Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms. *Data & Knowledge Engineering 63*(3), 725–751.

Mazón, J., J. Trujillo, M. Serrano, and M. Piattini (2005). Designing data warehouses: from business requirement analysis to multidimensional modeling. In *Proc. of Int. Conf. on Requirements Engineering for Business Need and IT Alignment*, pp. 44–53.

Pourshahid, A., D. Amyot, L. Peyton, S. Ghanavati, P. Chen, M. Weiss, and A. Forster (2009). Business process management with the user requirements notation. *Electronic Commerce Research 9*(4), 269–316.

Pourshahid, A., G. Richards, and D. Amyot (2011). Toward a goal-oriented, business intelligence decision-making framework. In *Proc. of Int. Conf. on E-Technologies: Transformation in a Connected World*, pp. 100–115. LNBIP 78.

Prakash, N. and H. Bhardwaj (2012). Early information requirements engineering for target driven data warehouse development. In *Proc. of Conf. on the Practice of Enterprise Modeling*, pp. 188–202. LNBIP 134.

Prakash, N. and A. Gosain (2008). An approach to engineering the requirements of data warehouses. *Requirements Engineering 13*(1), 49–72.

Silva, V., J. Mazón, I. Garrigós, J. Trujillo, and J. Mylopoulos (2012). Monitoring strategic goals in data warehouses with awareness requirements. In *Proc. of ACM Symposium on Applied Computing*, pp. 1075–1082. ACM.

Winter, R. and B. Strauch (2004). Information requirements engineering for data warehouse systems. In *Proc. of ACM Symposium on Applied Computing*, pp. 1359–1365.

Résumé

Les entrepôts de données (EDs) visent à soutenir le processus de prise de décision d'une organisation. Dans le domaine de l'Ingénierie des Exigences (IE), plusieurs méthodes ont été proposées pour le développement des EDs, la plupart d'entre elles basées sur l'approche appelé Goal-Oriented Requirements Engineering (GORE). Cependant, il n'existe pas encore une perspective globale et unifiée des différentes méthodes proposées. Dans cet article, une vision cohérente de l'approche GORE pour le développement d'EDs est présentée, en classant les méthodes existantes selon le processus de prise de décision, les intégrant ensemble, et reliant les techniques de modélisation et d'analyse tout au long du processus de prise de décision. Le résultat de notre étude est une méthode intégrée pour le développement d'EDs basée sur l'approche GORE. Nous illustrons la méthode avec un exemple concret du secteur des soins de santé.