# The new era of Business Intelligence Applications: building from a collaborative point of view

#### Miguel A. Teruel, PhD

LoUISE Research Group Computing Systems Department University of Castilla - La Mancha Avda. de España s/n 02071 – Albacete (Albacete, Spain) miguel@dsi.uclm.es

#### Alejandro Maté, PhD

Lucentia Research Group Department of Software and Computing Systems University of Alicante Carr. San Vicente del Raspeig, s/n 03690 – Sant Vicent del Raspeig (Alicante, Spain) amate@dlsi.ua.es

#### Elena Navarro, PhD<sup>1</sup>

LoUISE Research Group Computing Systems Department University of Castilla - La Mancha Avda. de España s/n 02071 – Albacete (Albacete, Spain) Elena.Navarro@uclm.es

#### Pascual González, PhD

LoUISE Research Group Computing Systems Department University of Castilla - La Mancha Avda. de España s/n 02071 – Albacete (Albacete, Spain) Pascual.Gonzalez@uclm.es

#### Juan Trujillo, PhD

Lucentia Research Group Department of Software and Computing Systems University of Alicante Carr. San Vicente del Raspeig, s/n 03690 – Sant Vicent del Raspeig (Alicante, Spain) jtrujillo@dlsi.ua.es

<sup>&</sup>lt;sup>1</sup> Corresponding author

## The new era of Business Intelligence Applications: Building from a Collaborative Point of View

#### Abstract.

*Problem definition*: Collaborative Business Intelligence (BI) is being widely embraced for enterprises as a way to make the most of their business processes. However, decision makers usually work in isolation without the knowledge or the time needed to obtain and analyze all the available information for making decisions.

*Relevance*: Unfortunately, collaborative BI is currently based on interchanging e-mails and documents between participants. As result, information may be lost, participants may become disoriented, and the decision-making task many not yield the needed results.

*Methodology*: We are proposing a modeling language aimed at modeling and eliciting the goals and information needs of participants of collaborative BI systems. This approach is based on innovative methods to elicitate and model collaborative systems and BI requirements. A controlled experiment was performed to validate this language, assessing its understandability, scalability, efficiency and user satisfaction by analyzing two collaborative BI systems.

*Results*: By using the framework proposed in this work, a clear view can be provided regarding: (i) collaborative tasks, (ii) their participants, and (iii) the information to be shared among them.

*Managerial implications*: By using our approach to design collaborative BI systems, practitioners may easily trace every element needed in the decision processes, avoiding the loss of information and facilitating the collaboration of the stakeholders of such processes.

**Keywords:** Collaborative systems; Business intelligence; Goal-oriented requirements; CASE support; *I-Star*; Controlled experiment

#### 1 Introduction

In recent years, Business Intelligence (BI) has focused on providing better and more useful information to decision makers aimed at improving the decision-making process. However, a decision maker working in isolation can only make precise and informed decisions within his / her expertise field and time frame. In order to make a decision that requires going beyond one's expertise or one's available time, decision makers should collaborate with those able to cover their weaknesses. As an example, let us consider a CEO having to develop a statistical model to analyze customers' habits in order to explain a recent rise or drop in the sales of the company. This would require the CEO to have sound technical and statistical abilities and enough time to be taken from his other tasks. Unfortunately, this scenario seems unlikely.

Current practices for collaborative scenarios in BI are very simple: decision makers communicate and share data with others by sending and receiving e-mails, spreadsheets, and so on (Berthold et al. 2010). Therefore, information may be lost, participants may become disoriented and, consequently, the decision-making process would yield poor results. Indeed, in order to address these problems, BI tools can benefit from the advantages that Computer Supported Cooperative Work systems provide (Schmidt and Bannon 1992), such as the remote collaboration of different stakeholders who will be able to provide knowledge in real-time to foster the decision-making process. Moreover, the introduction of awareness information (Gutwin and Greenberg 2002; Gross 2013) in this shared workspace will enrich the collaborative decision-making process, thus making stakeholders aware of who has the required knowledge to make a decision, or their availability for real-time discussions on decision-making.

Recent research has focused on developing novel support for collaborative decision making within BI platforms. From sharing data (Rizzi 2012) to creating virtual rooms (Berthold et al. 2010), different approaches try to provide tools that enable decision makers to share the required information in a more ordered and easier manner. Nevertheless, up to now, no approach has been proposed that provides designers with facilities to model collaborative tasks as well as the information needed to carry them out, thus providing adequate support for these tasks.

In this context, the first goal of this paper is to propose an extension of CSRML4BI, a goal-oriented framework that enables BI designers to model and elicit both the user's and participants' requirements in individual and collaborative BI tasks. This revised version will thus address the shortcomings of the already existing version (*Reference removed for blind review*), by adding the following features:

- *New modeling elements*: several elements and relationships have been added to the language in order to make a more complete the specification of collaborative BI systems.
- *New awareness model*: the previous awareness model has been replaced by a far more comprehensive one which will not only enable us to model more present and past awareness requirements, bus will also include elements related to the future or social aspects.
- Model organization in diagrams: this extension of CSRML4BI now enables us to divide the collaborative BI system specification into 5 different diagrams which will improve the understandability and readability of the generated models.
- *Comprehensive formalization*: the CSRML4BI metamodel has been revised and extended, thus including the new modeling elements, as well as the multi-diagram support.
- *CASE Support*: together with this extension of CSRML4BI, a CASE tool has also been released in order to facilitate its usage.
- *Empirical evaluation*: CSRML4BI was empirically evaluated by means of a controlled experiment in order to assess its suitability.

Therefore, thanks to our approach, designers can (i) accurately identify which participants need to communicate with each other, (ii) why, and (iii) what information they need to share. Therefore, we can plan how the system will support collaboration in the Requirements Engineering stage of the Software Development Process (Pressman 2009; Pohl 2010). Our framework is based on recent approaches proposed for modeling both business intelligence requirements (*Reference removed for blind review*) and collaborative systems (*Reference removed for blind review*).

The second goal of this work consisted of evaluating the proposed CSRML4BI extension. For this, we performed a controlled experiment where our participants analyzed two collaborative BI systems modeled with CSRML4BI and *i*\* (Yu 1997)

The rest of this paper is organized as follows: Section 2 shows the related work in collaborative BI and decision-making, BI requirements collaborative modeling. Section 3 presents our framework, depicting the metamodel and introducing new elements. Section 4 describes the controlled experiment that was performed to evaluate our proposal. Our conclusions and future work are outlined in Section 5. Appendix A describes a collaborative BI system modeled by CSRML4BI.

#### 2 Related Work

The amount of information available has been continually increasing in recent years. Social Media analysis (Asur and Huberman 2010; Oh et al. 2013), Big Data (Zikopoulos et al. 2011; Embley and Liddle 2013) and Open Data initiatives (Lakomaa and Kallberg 2013; Oh et al. 2013; Lindman et al. 2013) have driven the increased interest in collaborative BI (de Moor 1999; Kaufmann and Chamoni 2014) as isolated decision makers and analysts no longer have sufficient knowledge to make a decision with confidence.

In a survey of the area, covering several approaches on collaborative BI, we can establish three well differentiated groups (Kaufmann and Chamoni 2014). Most existing approaches (Dayal et al. 2008; Berthold et al. 2010; Devlin 2012) understand Collaborative BI as a technology-driven development, i.e. the enrichment of existing BI systems by communication tools. These are classified as *internal communication* (IC) approaches. A relevant example is presented by Berthold et al. (Berthold et al. 2010), who propose an architecture for a BI platform that supports collaboration, including collaboration rooms where decision makers can jointly analyze dashboards and charts while they are aware of the presence and actions of other decision makers. On the other hand, collaborative BI systems focus on *partnership in data* (PD), where external partners are involved in the process of data provision. The approach proposed by Golfarelli et al. (Golfarelli et al. 2012; Rizzi 2012), presents the *Business Intelligence Networks* (BIN)

concept. In this scenario, every network participant can share and query information from other participants in the network by using mappings between information schemata. The third group revolves around *partnership in analysis* (PA) (Mettler and Raber 2011; Liu and Daniels 2012), i.e. the collaborators work together in the data analysis process. Mettler and Raber propose an architecture based on a central data warehouse, where suppliers and manufacturers can collaborate to manage the purchasing of supplies and the manufacturing process (Mettler and Raber 2011).

All these technical advances facilitate the application of collaborative BI. However, there is a need for an approach that helps BI system designers to elicit and model the requirements of BI systems whose final users must collaborate in order to achieve the system's goals.

As far as collaborative system requirements modeling is concerned, CSRML (Collaborative Systems Requirements Modeling Language) was presented in (*Reference removed for blind review*). CSRML is a language that expands the expressive capabilities of *i*\* (*Reference removed for blind review*) in order for analysts to specify the requirements of collaborative systems. Another language, based on XML, for User Interfaces (UI) design has been also proposed (Figueroa-Martinez et al. 2013) and extended to support collaborative information requirements. Both languages were designed for all-purpose collaborative systems and thus lack the detail that data warehouse and BI requirements modeling approaches, such as Giorgini et al.'s (Giorgini et al. 2008) and (*Reference removed for blind review*) do provide. These latter approaches are also based on *i*\*.

On the other hand, BI requirements modeling approaches (Giorgini et al. 2008) (*reference removed for blind review*) enable us to capture the rationale of individual decision makers, including the information that BI systems must store to support the decision-making process. However, they are unable to describe collaborative tasks and their characteristics, including the kind of collaboration and workspace awareness required for each task. As a result, although these approaches are more suitable for modeling BI systems, they lack adequate constructs for modeling collaborative BI requirements.

It is noteworthy that modeling approaches can benefit from several goal-reasoning techniques (Giorgini et al. 2003; Giorgini et al. 2005; Horkoff and Yu 2010). Decision makers may exploit goal-

reasoning to ask important questions about the system related to their models, such as (i) goal priorities, (ii) implementation alternatives, or (iii) which stakeholder needs can be satisfied. In order to provide answers to such questions, goal-reasoning techniques require that modeling languages are complete with respect to the concepts of the domain. Therefore, before obtaining interesting knowledge, our modeling language must be able to adequately represent all the relevant elements involved in a collaborative decision process, thereby reinforcing the need for a dedicated language that captures all the relevant aspects of collaborative BI.

In short, current advances in collaborative BI would benefit greatly from a modeling language that system designers could use to capture the requirements of this type of information system, including their collaborative aspects. However, due to the idiosyncrasy of collaborative BI, current requirements modeling proposals fall short of achieving this end.

## 3 CSRML4BI: An Improved Modeling Language for Collaborative Business Intelligence

Throughout the literature, one can find applications of goal-oriented requirements modeling for different domains, such as adaptive applications (Vitali et al. 2015) or data warehouses (*Reference removed for blind review*). However, when dealing with systems that entail the specification of collaboration, current goal-oriented approaches lack the expressive power needed to deal with this feature properly (*Reference removed for blind review*). Because of this, CSRML (*Reference removed for blind review*) was chosen as the foundation of our proposal, since it is used to specify collaborative tasks as well as groups of actors, while specifying the system from a decision viewpoint. CSRML also supports awareness modeling characteristics, enabling the system designer to specify what the stakeholders must be aware of to collaborate properly. In this regard, CSRML enables us to model awareness requirements such as where the participants are working, what they are doing, who has a certain piece of strategic information, or when a decision was taken.

In this scenario, CSRML4BI is a Goal-Oriented Requirements Engineering language designed to model collaborative BI requirements, in the form of an evolution of its previous definition (*Reference re-* *moved for blind review*). It also introduces new constructs that may be used by designers to perform the elicitation and specification of information requirements requiring the collaboration of single or multiple decision makers. These new constructs defined in this new version of the CSRML4BI metamodel are highlighted in red in Figure 1 to Figure 3.

It is worth noting that CSRML4BI involves a top-down modeling approach (see Figure 4). For this reason, the system specification begins with the identification of its participants, after which the main BI goals of the system are specified. Next, the conditions for participants (actors) to play certain roles are defined, as well as their responsibilities. The system task definitions, the cornerstone of the BI system specification, are obtained by defining how the different participants must collaborate, as well as what they have to be aware of to collaborate. Finally, the necessary quality factors are defined.

Not only new modeling elements have been added to the language, but also the previous awareness model in CSRML4BI (Gutwin and Greenberg 2002) has been extended to also consider future and social awareness needs (*Reference removed for blind review*) (see Section 3.4). A CASE tool was also developed to facilitate the modeling process with this new language (see Section 3.6)



Figure 1. CSRML4BI metamodel (elements)







Figure 3. CSRML4BI metamodel (diagrams)



#### Figure 4. Purpose and relationships among CSRML4BI diagrams

As in CSRML, CSRML4BI promotes the specification of the requirements of collaborative BI systems by means of 5 different types of diagram to improve the readability and understandability of the specification. These diagrams will be described in Subsections 3.1 to 3.5, along with the elements and relationships shown in the above metamodel. A running example will be used to show how to specify a collaborative BI system with CSRML4BI in Appendix A, and will include concepts from Kelly et al. (Kelly et al. 2004).

## 3.1 Group Hierarchy Diagram (GHD)

The Group Hierarchy Diagram (GHD) (Figure 5) depicts the different stakeholders (and their groups) involved in the BI system. Considering that CSRML4BI has user collaboration as one of its main cornerstones, *Actor, GroupActor* and *Participation Link* have been described as follows:

• *Actor*: these can be either users, programs, or entities with certain acquired capabilities. They can play a role while executing some action, using devices or being responsible for actions. Actors can play one or more *roles* regarding the information system that is being specified.

 • *GroupActor*: it designates a group formed by one or more actors who aim at achieving one or several goals. Therefore, by using these elements, groups of users can be specified. Unlike what happens with the *Role* concept (Section 3.3), a GroupActor is static. That means that an actor will always be part of the same GroupActor.

• *Participation Link*: this relationship is used to assign an actor to a groupactor (see Figure 6). Its cardinality established how many actors constitute each groupactor.



Figure 6. Graphical representation of GHD's elements

GroupActor

Actor

## 3.2 System Goals Diagram (SGD)

A System Goals Diagram (SGD) (Figure 7) is used to identify the goals of a BI system. Each of these goals will be assigned to the *actors* and *groupactors* necessary for their achievement. The actors participating in the system-to-be can have one or more *Goals*. In BI requirements modeling, informational goals (*Reference removed for blind review*) are refined into different kinds of BI goals, as opposed to traditional goals (*Reference removed for blind review*), which result in *Strategic Goals, Decision Goals and Collaborative Decision Goal*. Taking into account the high number of goals, the new version of CSRML4BI simplifies the diagram by specifying only those *Actors* who are responsible for fulfilling each *BI Goal*. The same actor is also responsible for the remaining non-collaborative informational goals derived from it. Therefore, the most relevant elements and relationships in an SGD are the following:

- *Business Process*: it represents business activities that decision makers need to analyze to improve results and thus business performance.
- *Goal*: answering "why?" questions, this element describes a state that an actor wants to achieve. Nevertheless, goals do not specify how this state should be achieved. As aforesaid, there are 4 specializations for BI goals:
  - Strategic Goal (S): goals associated with one or more decision makers aiming at improving a Business Process within the enterprise. It represents the highest level of abstraction in informational goals and gives an immediate benefit to the organization when achieved.
  - Decision Goal (D): it tries to answer "how can a strategic goal be achieved?", and represents decisions for taking actions that contribute to achieving a strategic goal. Decision goals are only specified in relation to a strategic goal and do not provide any profit to the organization on their own. They can be specialized into *Collaborative Decision Goals*.
  - Collaborative Decision Goal (CD): Decision Goals requiring the Collaboration or Coordination of several Actors for their achievement. They represent decisions that must be made by a group instead of a single person.
  - Information Goal (I): it answers the question "how can a decision goal be achieved in terms of the required information?". Its fulfillment helps to achieve one or more decision goals. Such information goals take place within the context of a decision goal.
- *Business Intelligence link*: it represents non-*i*\*-standard decompositions among BI goals (see Figure 8)
- *Participation link*: it is used to specify which actors are involved in the accomplishment of the main goals of the system. The number of occurrences of each actor or group actor is denoted by the cardinal-ity field (see Figure 8).



Several elements in the metamodel illustrated in Figure 7 (*Task* and *Task Decomposition Link*) will be described in Section 3.4.

### 3.3 Responsibility Diagram (RD)

Each Responsibility Diagram (RD) (Figure 9) represents one of the tasks identified in the SGD. The RD specifies the roles played by the actors (under certain guard conditions) and the tasks that the actors are responsible for. Such actors can play either one or more *Roles* while interacting with the system so that the same actor can be tackled in a different manner depending on the role played. The following elements can be found in RDs:

• *Role*: it designates a set of correlated tasks to be performed by an actor. Hence, when an actor plays a role, he / she may participate in both individual and collaborative tasks (by means of *participation links*) and may assume the responsibility to achieve a goal (by means of *responsibility links*). Roles can

change dynamically (unlike GroupActors, whose Actors are always the same). An Actor can play different roles depending on which guard conditions are satisfied.

- *Playing link*: it is employed to represent an actor who is playing a role. These links have a *guard* condition (Figure 10) that represents what conditions must be satisfied so that a role can be played by an actor.
- *Responsibility link*: this link is used to assign roles (played by actors) to goals, softgoals or tasks (See Sections 3.4 and 3.5). This link represents which stakeholder is responsible for a goal/task accomplishment.



Figure 10. Graphical representation of RD's elements

Responsibility link

Softgoal

3.4 Task Refinement Diagram (TRD)

Role

In a Task Refinement Diagram (TRD) (Figure 11), the tasks previously identified in RDs are decomposed into individual and collaborative tasks that support awareness features. In order to achieve the different information goals, decision makers need to perform certain analysis tasks using information provided by the information system. In our approach, we specify this type of task by means of the refined

concept *Information Requirement*, used to describe both analysis tasks to be supported and the information to be provided by the system being designed. CSRML4BI identifies the following elements in a TRD:

• *Task:* it represents actions that an actor wants to execute, usually with the purpose of achieving some goal. These tasks are hierarchically refined into subtasks up to leaf-level tasks, which define system requirements (Dalpiaz et al. 2016). As shown in the metamodel (see Figure 1) this element has an importance level according to the task's development priority. This importance is defined by a graphical notation (Moody 2009) based on a color code (green, yellow, orange, red), green being the least and red the most important. The concept of task is refined for the BI domain into:

• Abstract task: an abstraction of a set of concrete tasks and other elements.

- *Concrete task*: a refinement of an abstract task and is related to roles responsible for its accomplishment. There are four types of concrete task: An *Individual task* that an actor can perform without any kind of interaction with other actors; *Collaboration / Communication / Coordination tasks* that require two or more actors to be involved in order to perform any kind of collaboration / communication / communication.
- Information Requirement: it represents the analysis of information that a decision maker will perform in order to satisfy the corresponding Information Goals. They can be decomposed into several Information Requirements, Business Process Contexts and Measures, necessary to support the decision-making process. It can be specialized into Collaborative Information Requirements.
- Collaborative Information Requirement: Information Requirements used by the decision makers to satisfy the Information Goals along with the collaboration of other actors. They can (i) simply represent the involvement of another actor without further consequences for the analysis, as when information is shared; (ii) specify Communication, where one or more roles have to communicate, such as when decision makers request information from analysts or managers; (iii) specify Collaboration, where all the roles involved interact with each other during the analysis; or (iv) specify a Coordination, where every role involved has to coordinate its analysis task.

• *Resource*: a resource is considered as an entity (either physical or informational) required by actors for achieving goals or performing tasks. The main interest when specifying resources is whether they are available and from whom. To properly specify collaborative BI systems, it has been specialized (see Figure 12) into information resources as follows:

- Business Process Context (BPC): it represents information about a certain entity which needs to be specified and provided by the system to ease the analysis of business processes from a certain viewpoint. It can be aggregated into other Contexts, thus forming an analysis hierarchy which will be implemented in the information system being specified. A Context can be also specialized into a Business Process Shared Context.
- *Measure (M)*: it is used to specify numerical information that somehow estimates the throughput of the business activity under study, as well as the needs to be recorded in order to empower the analysis. Can be specialized into *Shared Measure*.
- Business Process Shared Context (BPSC): it represents entity information that is provided by an Actor to the system, instead of being gathered by the system itself. Consequently, the supplier is responsible for providing this information.
- *Shared Measure (SM)*: it represents numerical information related to a certain activity that an *Actor* provides to the system. Its supplier is responsible for the existence of this information.
- Awareness Resource: it represents a perception requirement that helps a role to perform a task by providing the needed awareness. It includes a set of attributes attached to a *participation link* between a task and the role performing it. Note that this kind of element is depicted in the diagrams in two different ways: the expanded and reduced form. In its expanded form, the Awareness Resource shows all the Workspace Awareness features identified in (*Reference removed for blind review*) that can be set (if needed) with their importance according to the contribution to the accomplishment of a task (see Figure 12), these awareness features are categorized into four sections, related to present, past, future and social awareness needs, like in (*Reference removed for blind review*). The importance of each awareness element can be *nice to have* (N), *desirable* (D), *highly desirable* (HD) or *mandatory* (M).

 • *Dependency*: dependencies are relationships between a depender and a dependee for a dependum. Both depender and the dependee are roles played by actors. Dependums can be either goals, tasks, resources or softgoals. Hence, dependers depends on dependees to achieve goals, to perform tasks or to use resources. If the dependee does not provide the depender with the dependum, it may be difficult or even impossible for the former to achieve a goal, perform a task or use a resource. According to the kind of the dependum, 4 types of dependencies can be found: goal, task, resource or softgoal dependencies.

- *Means-end link*: means-end links define whether a softgoal, task, and/or resource contributes to achieving a goal. These links also ease documentation and evaluation of alternative ways to satisfy a goal (Horkoff and Yu 2012), that is, they are used to specify different ways of decomposing goals into several subgoals, tasks or resources.
- *Task decomposition link*: it depicts the fundamental elements of a task. Task decomposition links relate a task to its components. Such components can be any combination of goals, tasks, resources, and softgoals. The decomposition of a task can comprise sub-tasks that have to be performed, sub-goals that have to be achieved, resources that can be needed, and softgoals that usually define quality goals for such a task.
- *Participation-link*: it defines which role is involved in the performance of a task. These links have an attribute to specify cardinality, i.e. the number of users that can be involved in a task. It may be optionally related to an awareness resource for specifying that the role involved has a special perception requirement (specified though an awareness resource) to participate in a task. Without this awareness information, the performance of the task could be negatively affected or the role will not be able to participate in its accomplishment.
- *Restriction*: it represents a temporal restriction between two tasks following the UsiXML style (Limbourg et al. 2004). These temporal restrictions (and symbolic representation) can be:
  - Enabling (>>): it specifies that the second task will not begin unless the first task is performed.
  - *Choice* ([]):it defines that once a task starts, the other task is not enabled anymore.

- Enabling with information passing ([]>): it defines that the second task cannot be performed until the first one is carried out, using the output information of the first task as input for the second.
- *Concurrent tasks* (|||): they define the likelihood of performing tasks in any order, even at same time. It is also possible for a task to start before the other task is finished.
- *Concurrent communicating tasks* ([[]]): they specify that the related tasks can share information while they are performed concurrently.
- *Task independence* (|=|): it defines that the related tasks can be performed in any order. However, when one task starts it has to finish before the other can begin.
- Disabling ([>): it specifies that the first task (commonly an iterative task) is entirely interrupted by the second.
- Suspend-Resume (|>): it defines that the first task can be interrupted by the second. Hence, once the second terminates, the first one can be reactivated from the previously reached state.

Examples of several of these restrictions can be found in Figure 28 (Appendix A).



Figure 12. Graphical representation of TRD elements

There are several ways to achieve system goals. For example, individual users may use the system in isolation, or decisions may involve multiple decision makers, requiring not only business process related information but also requiring the system to be context-aware (Martínez-Carreras et al. 2013).

#### 3.5 Quality Factors Diagram (QFD)

A Quality Factors Diagram (QFD) (Figure 13) specifies the quality factors that contribute to achieving the softgoals (quality factors) and tasks identified in RDs and TRDs. Therefore, these diagrams are used to specify the non-functional part of the system (Zhu et al. 2012) by using softgoals. As a novelty, the following elements may be specified in these diagrams (Figure 14):

- *Softgoal* is a state that an actor wants to achieve. Nevertheless, unlike (hard) goals, the condition for the achievement is not well-defined. Hence, a softgoal is typically an attribute related to the system's quality that constrains other elements, such as goals, tasks or resources.
- *Contribution link* depicts an influence from a task or softgoal to a different softgoal. It is defined by means of some of the following types of attribute:
  - *Make*: A positive contribution strong enough to fulfill a softgoal.
  - *Some*+: A positive contribution with unknown strength.
  - Help: A partial positive contribution, yet not enough by itself to fulfill the softgoal.
  - Unknown: A contribution to a softgoal whose influence is unknown.
  - Some-: A negative contribution with unknown strength.
  - *Hurt*: A partial negative contribution, yet not enough by itself to deny the satisfaction of a softgoal.
  - Or: The parent is fulfilled if any of its children is fulfilled.
  - And: The parent is fulfilled if all its children are fulfilled.





Figure 14. Quality factors

## 3.6 Basic CSRML4BI Model Example

In this section, a basis CSRML4BI model is described as an example of a straightforward BI system (Figure 15). In this system, two kinds of stakeholders will participate, i.e. *Actor 1* and *Actor 2* (Figure 15a). In this system, one *Actor 1* will participate (1), although one or more *Actor 2* will do so (1..\*). These actors will constitute the *GroupActor*, whose leader will be *Actor 1* (hand icon).

As expected, this model will represent a *Business Process*, consisting of one *Strategic Goal* (Figure 15b). The *Strategic Goal* is decomposed into a *Collaborative Decision Goal* which, in turn is decomposed into one *Information Goal*. The latter will be achieved thanks to the *System Main Task*. *Actor 1* will be involved in the accomplishment of the *Strategic Goal*, while the whole GroupActor will be involved in accomplishing the remaining goals.

It can be seen in Figure 15c that there will be two roles, *Role 1* and *Role 2*, which will be played by *Ac*tor 1 and *Actor 2* respectively, when certain guard conditions are accomplished. The previous *System Main Task* will consist of just one sub-task, namely *Task*, for which *Role 1* will be responsible.

The mentioned *Task* is to be specified in Figure 15d and will be decomposed into one *Goal*, which will be fulfilled by means of two *Information Requirements*, an individual and a collaborative. Both Information Requirements will be decomposed into two resources, namely a *Measure* and a *Business Process* 

*Context* (see Section 3.4). In the case of the *Collaborative Information Requirement*, such resources will be shared. The *Information Requirement* will be performed by one (1) *Role 1*. However, in order to perform the *Collaborative Information Requirement*, one *Role 1* and one or more (1..\*) *Role 2* must participate. The *Awareness Resource* indicate that it is mandatory ([M]) for Role 1 to be aware of the *Presence* of others in order to participate in *Collaborative Information Requirement*. There is a restriction between the two *Information Requirements*: *Information Requirement enables Collaborative Information Requirement*. Therefore, the former must be performed prior to the second.

Figure 15e specifies the sole quality factor present in this example, namely *Softgoal*. In this case, the previous *Information Requirement* will contribute positively to the fulfillment of the *Softgoal*.



Figure 15. Basic CSRML4BI model

#### 3.7 CASE Support: CSRML4BI Tool

Since CSRML4BI is an evolution of CSRML for the BI domain, its CASE support has been developed by adapting and extending the original CSRML CASE tool (*Reference removed for blind review*) with the new BI features. Indeed, as the original tool, this new version, shown in Figure 16 has also been integrated with Visual Studio to provide BI practitioners with facilities to specify and verify BI requirements models. This tool is available for the BI community through the Visual studio Marketplace (*Reference removed for blind review*).





#### 4 Evaluation

To evaluate our proposal, a controlled experiment was carried out designed to compare CSRML4BI with  $i^*$ , the language it is based on. We compared both languages by using them to model two different systems and then evaluating these models for their understandability, scalability, efficiency and user satisfaction.

#### 4.1 Experimental context

The main goal of this experiment, defined by using Goal Question Metric (Basili et al. 1994) as: **analyze**  $i^*$  and CSRML4BI **for the purpose of** evaluating the understandability, scalability, efficiency and user satisfaction for both languages, **for** researchers **in the context of** BI practitioner and undergraduate students. With this aim, Table 1 presents the hypothesis that this experiment tried to demonstrate.

ruble 1. multi features of the experiment	Table 1.	. Main	features	of the	experimen
---	----------	--------	----------	--------	-----------

Null hypothesis	H <sub>0A</sub> : CSRML4BI and $i^*$ have the same	e score for understandability of BI models.										
	$H_{1A}$ : $\neg H_{0A}$											
	H <sub>0B</sub> : CSRML4BI and $i^*$ have the same score for scalability of BI models. H <sub>1B</sub> :											
	¬H <sub>0B</sub>											
	H <sub>0C</sub> : CSRML4BI and <i>i</i> * have the same score for efficiency when analyzing BI											
	models. H <sub>1</sub> C: ¬H <sub>0</sub> C											
	H <sub>0D</sub> : CSRML4BI and $i^*$ have the same score for user satisfaction when analyzing											
	BI models. H <sub>1D</sub> : ¬H <sub>0D</sub>											
Dependent varia-	Understandability score (UND), scalability score (SCA), efficiency score (EFF)											
bles	and user satisfaction score (SAT)											
Independent vari-	The language (CSRML4BI or $i^*$ ) used	The language (CSRML4BI or <i>i</i> *) used to specify the experimental models										
ables												
Location	Lucentia Lab (Sant Vicent del	University of Alicante (Sant Vicent del										
	Raspeig, Spain)	Raspeig, Spain)										
Date	January 2018	February 2018										
Subjects	9 Business Intelligence practitioners	62 Computer Science students										

It was decided to perform this experiment on experimental subjects from two different backgrounds. We first ran the experiment in a company specializing in creating BI solutions on participants with experience in understanding BI requirements models. The second experiment involved a large number of Computer Science students. The participants were required to have experience in requirements engineering as well as in elementary BI concepts, but not to have any previous experience of either CSRML4BI or  $i^*$  to avoid any bias.

## 4.2 Experimental design

The experiment consisted of reading and understanding two different BI requirements models created with the languages being compared. To avoid the *learning effect*, a  $2 \times 2$  factorial design with confounded interaction (Winer et al. 1991) was used, as shown in Table 2.

		Do	omain
		Supply Chain (1)	Public Transport (2)
Language	CSRML4BI	Group 1	Group 2
	<i>i</i> *	Group 2	Group 1

It was decided to use a supply chain (domain 1) and a public transport system (domain 2) for the experimental model domains. The former specifies a supply chain for several supermarkets in which suppliers and managers must take collaborative decisions. The latter (partially) collects the requirements of a public transport system of the smart city depicted in Appendix A. The experimental materials were thus requirements models of these two BI domains created with CSRML4BI and *i*\*. It should be noted that both the CSRML4BI and *i*\* models specify the same requirements, but modeled with the languages being analyzed. However, the specification for CSRML4BI was created by using the different diagrams that this language supports.

To analyze the understandability of both languages (UND) (ISO/IEC 9126 1991), we asked the participants 10 multiple-choice questions per model with 4 possibilities. UND was scored as the number of correct answers. Regarding scalability (SCA), half of the questions (even numbers) required reading two or more diagrams (for CSRML4BI) in order to answer them correctly, and SCA was thus scored as the number of correct answers for the even questions.

For the measurement and evaluation of efficiency (EFF), the participants were asked to write down the current time before and after answering the 10 questions of each domain. This enabled us to compute the elapsed time for understanding each model. EFF was calculated as number of correct answers per hour.

User satisfaction (SAT) was evaluated by asking subjects to answer different questions regarding several characteristics of the languages on a scale from 1 (nothing) to 5 (very) and were related to difficulty, understandability, readability, scalability, modifiability, traceability and expressivity. SAT was thus defined as the average score for each question. In the case of difficulty, the scale was defined from 1 (very hard) to 5 (very easy). For the sake of replicability, the models and questionnaires were published under a Creative Commons license (*Reference removed for blind review*). Finally, the experimental material given to the participants consisted of the following items:

- Double-sided A3 sheet of paper with the experimental models. Depending on the group (G1 or G2) the subjects belonged to, they were given a specific document (see (*Reference removed for blind re-view*)).
- A questionnaire to fill in statistical data, answer the questions and express their personal opinion (see *(Reference removed for blind review)*).
- Comprehensive documentation regarding both languages and their graphical notation (double-sided A4 sheet of paper per language) in case a subject had forgotten any concept.

The experimental task consisted of analyzing the provided paper models and then trying to answer the questionnaires.

#### 4.3 Running the experiment

The experiment was carried out in two different locations, first in the meeting room of a BI company with practitioners as the experimental subjects, then in a university classroom with undergraduates, whose experimental results are comparable to those obtained by professionals, according to Höst et al. (Höst et al. 2000). In both places, the experiment started with an introductory session, presenting both languages and the goal and procedure of the experiment. These sessions took around 30 minutes each. After this short introduction, the participants were given the experimental material.

To facilitate the participation in this experiment, both the introduction and the experimental material were translated into Spanish. Since there was no CASE tool able to model both CSRML4BI and *i*\* models, all the material used was provided on paper. There were no dropouts during the experiment.

	Practitioners	Students
Number of participants	9	57
Average age	27.75	21.54
Percentage of female participants	22.22%	14.04%
Maximum elapsed time	0:35:00	0:43:00
Minimum elapsed time	0:21:00	0:17:00
Average elapsed time	0:26:20	0:27:33

Table 3. Statistics about the experiment

 Table 3 summarizes the statistical data gathered, as well as the time participants took to complete the experiment.

## 4.4 Results

After recording the participants' paper questionnaires, we obtained the results shown in Table 4 and Table 5. CSRML4BI surpassed  $i^*$  regarding the four different dependent variables evaluated, regardless of type of subjects. In the following subsection, these dependent variables are analyzed in detail.

					Questions									
Participants	Group	Size	Language	1	2	3	4	5	6	7	8	9	10	
	1	-	CSRML4BI	80%	60%	40%	80%	100%	60%	100%	40%	80%	40%	
Destition	1	5	<i>i*</i>	0%	20%	100%	20%	60%	80%	20%	0%	20%	60%	
Practitioners 2			4	i*	0%	0%	100%	0%	75%	75%	100%	50%	0%	50%
	2	4	CSRML4BI	75%	0%	75%	100%	75%	75%	100%	50%	75%	25%	
Students 2	1	1 31	CSRML4BI	58%	97%	74%	71%	90%	71%	84%	84%	71%	52%	
	1		<i>i*</i>	3%	3%	45%	19%	42%	81%	16%	13%	0%	84%	
	2	2 26	<i>i</i> *	3%	3%	68%	21%	56%	94%	56%	85%	26%	38%	
	20	CSRML4BI	94%	12%	53%	41%	85%	59%	53%	59%	59%	29%		
D.1	1	20	CSRML4BI	61%	92%	69%	72%	92%	69%	86%	78%	72%	50%	
	1	36	<i>i*</i>	3%	6%	53%	19%	44%	81%	17%	11%	3%	81%	
Both	2	20	<i>i</i> *	3%	3%	71%	18%	58%	92%	61%	82%	24%	39%	
	2	2	30	CSRML4BI	92%	11%	55%	47%	84%	61%	58%	58%	61%	29%

Table 4. Questionnaire results per language and question

			Results					Opinion						
Subjects	Language	Q. Odd	Q. Even	Total	Time	Correct an- swers / hour	Difficulty	Readability	Scalability	Under- standability	Modifiabil- ity	Traceability	Expressivity	Total
Drastitionara	CSRML4BI	4.00	2.65	6.65	0:14:35	27.36	3.18	3.63	3.70	3.27	3.23	3.53	3.90	3.49
Practitioners	<i>i*</i>	2.38	1.78	4.15	0:11:45	21.19	2.25	2.27	2.77	2.17	2.60	2.63	2.83	2.50
Studente	CSRML4BI	3.61	2.87	6.48	0:14:00	27.77	2.77	3.46	3.25	2.85	3.25	3.27	3.60	3.21
Students	i*	1.58	2.21	3.78	0:13:31	16.79	2.41	2.71	2.80	2.24	2.67	2.92	2.82	2.65
A	CSRML4BI	3.65	2.83	6.48	0:14:04	27.66	2.82	3.48	3.30	2.89	3.23	3.30	3.63	3.24
Aily	i*	1.68	2.16	3.83	0:13:19	17.27	2.39	2.65	2.79	2.22	2.67	2.88	2.83	2.63

Table 5. Experiment results per language

#### Understandability

As can be seen in Figure 17, CSRML4BI obtained better results than  $i^*$  for understandability (UND), regardless of type of subject. These results were computed as the average score for all the questionnaire questions. Both practitioners and students achieved proportionally similar results. However, practitioners surpassed students in understanding BI requirements models (6.65 and 4.15 vs. 6.48 and 3.78).



Figure 17. Understandability results

In order to accept or reject the null hypothesis  $H_{0A}$ , a 2-Sample t Test was performed (Figure 18) with an alpha of 0.05. Thanks to this test, we could conclude that the means for UND differ at the 0.05 level of significance, with a p-value less than 0.001. With a 95% confidence level, the true difference was between 1.9634 and 3.2528, so that we rejected the null hypothesis  $H_{0A}$ , meaning that CSRML4BI does not have the same score for understandability of the BI models.



Figure 18. Distribution of data for understandability (correct answers)

In the results of question n° 5, directly related to collaboration, CSRML4BI obtained 92% for model 1 and 84% for model 2, while  $i^*$  got 44% and 58% respectively. These results indicates that CSRML4BI is much more understandable than  $i^*$  when dealing with collaboration among users, this being one of the cornerstones of our proposal.

## Scalability

The scalability score was computed as the average number of correct answers for the questions which required consulting several diagrams in the case of CSRML4BI. As these were the even questions, (see *(Reference removed for blind review)*), each subject had to score them between 0 and 5. Similarly to UND, the results were better for CSRML4BI, regardless of the subject type (Figure 19), although in this case the difference was not so high.



Figure 19. Scalability results

To assess that difference, a t-Test was performed again (Figure 20). Here again we concluded that the means for SCA differ at the 0.05 level of significance (p-value=0.001). However, the obtained confidence interval was lower in this case, being (0.25843, 1.0389) at a 95% confidence level. We thus rejected the null hypothesis  $H_{0B}$ , so CSRML4BI and *i*\* do not have the same score for scalability of BI models.



Figure 20. Distribution of data for scalability (correct even answers)

## Efficiency

Efficiency (EFF) was measured as the number of correct answers per hour. For this variable, the 10 questions were taken into account. Once again, CSRML4BI obtained a better score in both types of subject (Figure 21), although there was a difference in the means of 6.17 correct answers per hour for the practitioners and 10.98 for the students. Therefore, the use of CSRML4BI instead of  $i^*$  affected the efficiency of students more than practitioners.



Figure 21. Efficiency results

Once again, a t-Test was performed regarding the null hypothesis  $H_{0C}$  (Figure 22). This null hypothesis was also rejected, since the means differed at the 0.05 level of significance, with a p-value <0.001 and a 95% confidence interval of (8.0707, 17.333). Hence, CSRML4BI and *i*\* do not have the same score for efficiency when analyzing BI models.



Figure 22. Distribution of data for efficiency (correct answers per hour)

## **User Satisfaction**

User satisfaction (SAT) was measured as the average score of the personal opinion questions (See (*Reference removed for blind review*)). For this variable, CSRML4BI obtained a better score than *i*\* for individual questions (difficulty, understandability, readability, scalability, modifiability, traceability and ex-



pressivity), as shown in Figure 23. The total (average of these 7 metrics) was also better, as would later be confirmed by a t Test.

Figure 23. User satisfaction results

The t-Test also rejected the null hypothesis  $H_{0D}$ , since the means differ at the 0.05 level of significance, with a p-value <0.001, and a 95% confidence interval of (0.26242, 0.74085) (Figure 24). Because of this result,  $H_{0D}$  is rejected, so CSRML4BI and *i*\* do not have the same score for user satisfaction when analyzing BI models.



Figure 24. Distribution of data for user's satisfaction (score)

Two of the questions about personal opinion can also be used as a subjective measurement for UND and SCA variables (see (*Reference removed for blind review*), *It is understandable by a non-expert* and *It is scalable*). These results on personal opinion coincide with the objective ones (UND and SCA), as can

be seen in Table 6. The ratio between both subjective and objective results is closer to 1 for understandability than scalability.

			UND		SCA					
	Language	Objective $\Psi$	Subjective	Ratio	Objective	Subjective	Ratio			
Practitioners	CSRML4BI	6.65	3.27	1.02	2.65	3.70	0.72			
	<i>i*</i>	4.15	2.17	0.96	1.78	2.77	0.64			
Students	CSRML4BI	6.48	2.85	1.14	2.87	3.25	0.88			
	<i>i</i> *	3.78	2.24	0.84	2.21	2.80	0.79			
Both	CSRML4BI	6.48	2.89	1.12	2.83	3.30	0.86			
	<i>i</i> *	3.83	2.22	0.86	2.16	2.79	0.77			

Table 6. Comparison between objective and subjective results for understandability and scalability

<sup> $\Psi$ </sup>Scale from 0 to 10

To sum up, in view of the results for the four dependent variables, it can be said that CSRML4BI is more suitable for modeling BI requirements than  $i^*$ , regardless of the user's background (practitioners or students).

## 4.5 Threats to the Validity

As suggested by Wohlin et al. (Wohlin et al. 2012), in the following the most relevant threats to the validity of the controlled experiment described here are analyzed.

Internal validity is related to the influences on the independent variable (Wohlin et al. 2012). The different subjects that participated in the experiment were not informed previously, avoiding social threats. A 2x2 factorial design was applied, so that the language and the system were changed after a break between the two sessions in each group. The subjects were randomly assigned within the groups to cancel out both learning and fatigue effects.

According to Wohlin et al (Wohlin et al. 2012), external validity threats are related to the generalization of the experiment. The experimental subjects in the experiment had enough maturity level because the tasks to be carried out were not highly demanding in terms of industrial experience (Höst et al. 2000). However, it should also be noted that the practitioners who participated in the experiment had similar results for the experimental tasks. The models used for the evaluation were a partial description that could have been used to describe a real system in an industrial setting.

The method used to evaluate the outcome of the experimental task may threaten the Construct validity (Wohlin et al. 2012) of the experiment. In order to avoid this threat, a questionnaire was used to evaluate its understandability. Prior to conducting the experiment, external experts tried to complete the question-naires after analyzing the models, and we then refined the models and questions until they reached an understandability score of 100% for each model, thus reducing the chance of bias towards one of the languages.

Conclusion validity threats are related to the statistical relationship between the independent and dependent variables (Wohlin et al. 2012). The statistical power can be considered high, since 66 subjects participated in the experiment, being enough according to the central limit theory. We also avoided the "fishing for the result" effect as we focused the analysis on which language, *i*\* or CSRML4BI provided the best support for the specification of BI requirements. Finally, the experiment was not balanced, that is, the number of participants per group were different as shown in Table 4, the first group being bigger than the second. The authors of this paper did not monitor the running of the experiment to avoid introducing any bias.

### 5 Conclusions and Future Work

Collaborative BI is commonly practiced in companies as it helps decisions makers to make the most of the available information as well as to analyze the problem from other points of view. However, this approach is barely supported by currently available tools. Recent proposals focus on the improvement of the technical side of the development of these systems. However, they may significantly benefit from a requirements modeling technique that enables designers to perform the specification of collaborative system requirements. These systems could really provide the expected and needed functionality as well as consider important aspects from the CSCW community, such as awareness.

In this work, we present CSRML4BI, a goal-oriented and i\*-based framework for collaborative BI that offers expressive facilities to identify and model (i) the decision-making tasks that require collaboration

among the participants, (ii) the participants involved in collaborative decision making, and (iii) the information required and shared among them. Of the new elements in CSRML4BI it is important to highlight the facilities for specifying the business goals to be defined, the collaborative information requirement that helps stakeholders to analyze the goals of the system from different perspectives, as well as the resources, especially those related to measures and shared contexts, which may help decision makers to make decisions based on quantitative data.

In order to guide the process of applying CSRML4BI, its core elements were applied to specify the Smart City Dashboard (see Appendix A). Its specification was carried out using the different diagrams that CSRML4BI recommends, making an iterative refinement process from the actors and goals of the system to the tasks to be supported. Special attention was paid to the collaborative side of the specification, one of the main strengths of CSRML4BI.

To facilitate the specification of a collaborative BI system, a CASE tool was developed and made available for BI practitioners. This tool consists of an extension of an already existing tool (*Reference removed for blind review*), which was extended to support the new CSRML4BI. The new tool, which is currently available for free, was fully integrated with Visual Studio to obtain a complete IDE for the development of BI collaborative applications. With this tool, BI designers could benefit from the advantages it provides, such as improved correctness of the models, thanks to its automatic validation. Traceability of model diagrams is also improved, since it will manage the links between diagrams and elements.

A controlled experiment was performed to compare CSRML4BI against  $i^*$  for modeling BI requirements by taking into account four different characteristics (understandability, scalability, efficiency and user satisfaction) and using two different types of experimental subjects, namely BI practitioners and Computer Science undergraduate students. CSRML4BI obtained a better score than  $i^*$  for the four variables in both types of participants. Understandability and scalability were assessed in an objective (number of correct answers and correct answers per hour) and subjective manner (score for personal opinion). Once again, CSRML4BI obtained a better score for both the objective and the subjective measurement of understandability and scalability, regardless of the subjects' background.

Our future work will consist of several lines of research. It is planned to evaluate which goal reasoning techniques, such as that presented by Giorgini et al. (Giorgini et al. 2008), may be used to exploit the different alternatives specified by CSRML4BI. As part of our on-going work, we are analyzing how to exploit the collaborative requirements modeled in CSRML4BI to provide collaborative support automatically within BI platforms. Finally, we will develop a series of guidelines and video tutorials to promote its use among the BI community. By following these guidelines, a BI practitioner will be able to use CSRML4BI for modeling a BI system following a step-by-step procedure that will lead to the identification of BI goals, actors, collaborative tasks and so on. We are also considering developing a model-driven tool that generates scaffolding code to build the final application, taking a complete specification of a collaborative BI system as its input.

In a different vein, an additional experiment will be performed to evaluate the scalability of CSRML4BI from a different point of view. Different-size case studies will be considered, thus enabling us to assess whether our proposal's scalability depends on the size of the BI system.

#### References

- Asur S, Huberman BA (2010) Predicting the Future with Social Media. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. IEEE, Toronto, Canada, pp 492–499
- Basili VR, Caldiera G, Rombach HD (1994) The Goal Question Metric Approach. In: Encyclopedia of Software Engineering. Wiley, pp 528–532
- Berthold H, Rösch P, Zöller S, Wortmann F, Carenini A, Campbell S, Bisson P, Strohmaier F (2010) An architecture for ad-hoc and collaborative business intelligence. In: 1st International Workshop on Data Semantics (DataSem'10). ACM Press, Lausanne, Switzerland, p 13:1-13:6

Dalpiaz F, Franch X, Horkoff J (2016) iStar 2.0 Language Guide.

Dayal U, Vennelakanti R, Sharma R, Castellanos M, Hao M, Patel C (2008) Collaborative Business

Intelligence: Enabling Collaborative Decision Making in Enterprises. In: Meersman R, Tari Z (eds) On the Move to Meaningful Internet Systems: OTM 2008. Springer, Berlin Heidelberg, pp 8–25

- de Moor A (1999) Empowering communities: A method for the legitimate user-driven specification of network information systems. Tilburg University
- Devlin B (2012) Collaborative Business Intelligence: Socializing Team-Based Decision Making. Bus Intell J 17:9–17.
- Embley DW, Liddle SW (2013) Big Data—Conceptual Modeling to the Rescue. In: Ng W, Storey VC, Trujillo JC (eds) ER 2013: Conceptual Modeling. Springer, Berlin Heidelberg, pp 1–8

Figueroa-Martinez J, López-Jaquero V, Gutiérrez Vela FL, González P (2013) Enriching UsiXML language

- to support awareness requirements. Sci Comput Program 78:2259–2267. doi: 10.1016/j.scico.2012.09.020
- Giorgini P, Mylopoulos J, Nicchiarelli E, Sebastiani R (2003) Formal reasoning techniques for goal models. In: Journal on Data Semantics I. Springer, pp 1–20
- Giorgini P, Mylopoulos J, Sebastiani R (2005) Goal-oriented requirements analysis and reasoning in the tropos methodology. Eng Appl Artif Intell 18:159–171.
- Giorgini P, Rizzi S, Garzetti M (2008) GRAnD: A goal-oriented approach to requirement analysis in data warehouses. Decis Support Syst 45:4–21. doi: 10.1016/j.dss.2006.12.001
- Golfarelli M, Mandreoli F, Penzo W, Rizzi S, Turricchia E (2012) BIN: Business Intelligence Networks. In: Zorrilla ME, Mazón J-N, Ferrández Ó, Garrigós I, Daniel F, Trujillo J (eds) Business Intelligence Applications and the Web: Models, Systems and Technologies. IGI Global, Hershey, pp 244–265
- Gross T (2013) Supporting Effortless Coordination: 25 Years of Awareness Research. Comput Support Coop Work 22:425–474. doi: 10.1007/s10606-013-9190-x
- Gutwin C, Greenberg S (2002) A Descriptive Framework of Workspace Awareness for Real-Time Groupware. Comput Support Coop Work 11:411–446. doi: 10.1023/A:1021271517844
- Gutwin C, Greenberg S, Roseman M (1996) Workspace Awareness Support with Radar Views. In: Conference companion on Human factors in computing systems common ground (CHI'96). ACM Press, Vancouver, Canada, pp 210–211

Horkoff J, Yu E (2010) Finding solutions in goal models: an interactive backward reasoning approach. In:

International Conference on Conceptual Modeling. Springer, Aachen, Germany, pp 59–75

- Horkoff J, Yu ES-K (2012) Comparison and evaluation of goal-oriented satisfaction analysis techniques. Requir Eng 18:199–222. doi: 10.1007/s00766-011-0143-y
- Höst M, Regnell B, Wohlin C (2000) Using Students as Subjects—A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. Empir Softw Eng 5:201–214. doi: 10.1023/A:1026586415054
- ISO/IEC 9126 (1991) Software engineering Product quality.
- Kaufmann J, Chamoni P (2014) Structuring Collaborative Business Intelligence: A Literature Review. In: 2014 47th Hawaii International Conference on System Sciences. IEEE, Waikoloa, USA, pp 3738–3747
- Kelly J, Jones P, Barta F, Hossinger R, Witte A, Wolf A (2004) Successful Transport Decision-making: a project management and stakeholder engagement handbook. European Commission
- Lakomaa E, Kallberg J (2013) Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs. IEEE Access 1:558–563. doi: 10.1109/ACCESS.2013.2279164
- Limbourg Q, Vanderdonckt J, Michotte B, Bouillon L, López-Jaquero V (2004) USIXML: A Language
  Supporting Multi-path Development of User Interfaces. In: Bastide R, Palanque P, Roth J (eds)
  Engineering Human Computer Interaction and Interactive Systems. Springer, Berlin Heidelberg, pp 200–220
- Lindman J, Rossi M, Tuunainen VK (2013) Open Data Services: Research Agenda. In: 2013 46th Hawaii International Conference on System Sciences. IEEE, Wailea, USA, pp 1239–1246
- Liu L, Daniels H (2012) Towards a value model for collaborative, business intelligence-supported risk assessment. In: 6th International Workshop on Value Modeling and Business Ontology (VMBO'12). Vienna, Austria, pp 1–5
- Martínez-Carreras MA, Muñoz A, Botía J (2013) Building and evaluating context-aware collaborative working environments. Inf Sci (Ny) 235:224–241. doi: 10.1016/j.ins.2013.02.009
- Mettler T, Raber D (2011) Developing a collaborative business intelligence system for improving delivery reliability in business networks. In: 2011 17th International Conference on Concurrent Enterprising. Aachen, Germany, pp 1–7

Moody DL (2009) The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in

Software Engineering. IEEE Trans Softw Eng 35:756–779. doi: 10.1109/TSE.2009.67

Oh O, Agrawal M, Rao HR (2013) Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. MIS Q 37:407–426.

Pohl K (2010) Requirements Engineering: Fundamentals, Principles, and Techniques, 1st edn. Springer

- Pressman RS (2009) Software engineering: a practitioners approach, 7th editio. McGraw-Hill Science/Engineering/Math
- Rizzi S (2012) Collaborative Business Intelligence. In: Aufaure M-A, Zimányi E (eds) Business Intelligence. Springer, Berlin Heidelberg, pp 186–205
- Schmidt K, Bannon L (1992) Taking CSCW seriously. Comput Support Coop Work 1:7–40. doi: 10.1007/BF00752449
- Vitali M, Pernici B, O'Reilly U-M (2015) Learning a goal-oriented model for energy efficient adaptive applications in data centers. Inf Sci (Ny) 319:152–170. doi: 10.1016/j.ins.2015.01.023
- Winer BJ, Brown DR, Michels KM (1991) Statistical Principles in Experimental Design, 3rd edn. McGraw-Hill Humanities/Social Sciences/Languages
- Wohlin C, Runeson P, Höst M, Ohlsson MC, Regnell B, Wesslén A (2012) Experimentation in Software Engineering, 1st edn. Springer Berlin Heidelberg, Berlin, Heidelberg
- Yu ESK (1997) Towards modelling and reasoning support for early-phase requirements engineering. In: 3rd IEEE International Symposium on Requirements Engineering (ISRE'97). IEEE Comput. Soc. Press, Annapolis, USA, pp 226–235
- Zhu M-X, Luo X-X, Chen X-H, Wu DD (2012) A non-functional requirements tradeoff model in Trustworthy Software. Inf Sci (Ny) 191:61–75. doi: 10.1016/j.ins.2011.07.046
- Zikopoulos PC, Eaton C, de Roos D, Deutsch T, Lapis G (2011) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne, New York

#### Appendix A - Running example: Smart City Dashboard

In this annex, this new version of CSRML4BI is put into practice by specifying a Smart City Dashboard considering the ideas from the literature (*Reference removed for blind review*). In this system, decisions about public transport are amongst the most important for the development of a city. They determine where and how fast citizens can move around the city. For example, setting the locations of bus stops in order to optimize travel times or monitoring and establishing the capacity of buses to ensure safe journeys are crucial decisions for commercial activity. For these cases, multiple actors are involved in the decision process. These actors include the Transport City Council members, transport company representatives, and even citizens whose needs should be heard to provide a satisfying solution to the transport take into account the necessary collaboration features among these different actors. In the following subsections, it will be described how, thanks to the new version of CSRML4BI, we can model the requirements of this system for facilitating its later implementation.

#### A.1 Group Hierarchy Diagram (GHD)

First, we will focus on identifying the actors and groups involved in a *Smart City Dashboard*. These elements will be specified in the GHD, as shown in Figure 25. Revising relevant instances of transport decision-making processes, we can see that a "Regional Transport Consortium" is one of the most usual ways to group these actors. Accordingly, we can see in Figure 25 how "Council Member" and "Transport Representative" are grouped into the "Regional Transport Consortium" in our case study. Consortium members collaborate in equal conditions, so that they have access to the same information and consequently make better decisions.



Figure 25. Actor and groups involved in the Dashboard System

Furthermore, in a *Smart City*, the "Citizens" are also involved. They should have access to nearly the same data and tools as the previous actors to improve their use of transport to make it more efficient. They can analyze data by means of the dashboard application, individually or in groups. However, they do not collaborate directly with the "Regional Transport Consortium" and their decisions are just for their personal use, to create transport guidelines for a community of people with similar interests or to gather feedback for carrying out a petition to the Regional Transport Consortium.

Once we have identified the actors involved in our system, we will assign responsibilities to them by identifying their possible roles. This will be done by means of the Responsibility Diagram (RD). Never-theless, the main goals of the BI system should be identified first in the System Goals Diagram (SGD).

## A.2 System Goals Diagram (SGD)

System goal identification can be easier if analyzed from the perspective of the relevant business processes. The *Business Process* in our *Smart City* scenario is the "Public Transport", whose decision makers aim to improve the transport system by fulfilling certain strategic goals. The "Regional Transport Consortium" group has two strategic goals related to this process. Firstly, they reckon that "Minimize delays" at bus stops is a goal that causes an immediate benefit for the improvement of public transport. In order to achieve this, they have to make decisions regarding which actions should be taken to reduce delays in bus routes.



59

64 65

1



Figure 26. Goals modeled in the public transport scenario

For example, to determine the number and location of the bus stops in a route is a decision goal that has a direct impact on the possible delays. We called this goal "Set route stops" as shown in Figure 26, which is related to the information goals: "Unpunctual stops identified" and "Route duration analyzed". It is labelled as CD in Figure 26 because it is considered a Collaborative Decision.

Along with the previous strategic goal, the "Increase Road Safety" goal has also been identified. In order to achieve it, the Regional Transport Consortium considers that it is necessary to determinate how many passengers can travel in each type of vehicle due to the fact that many passengers do not get a seat in urban buses. This can affect the passengers' safety negatively in case of an accident. The decision goal is to "Establish the maximum legal capacity" in order to increase road safety. As before, this decision has to be made collaboratively by all the members of the consortium. We can also see that this goal requires to meet the information goal "At stop occupancy analyzed", meaning that both legal maximum as well as historical real occupation at each stop have to be analyzed. The analysis of historical accidents in each route can also be useful to make this decision, as some routes may be more dangerous than others. This is represented by the information goal "Accidents in route analyzed".

On the other hand, citizens pursue the strategic goal "Minimize travel time" using bus services in order to make more efficient use of public transport. To achieve this, they have to decide the optimal routes to reach the desired locations. To make this decision, the citizens need to achieve the information goal "Compare occupancy and delays per stop" by comparing the information about each route they can follow to reach their destination.

#### A.3 Responsibility Diagram (RD)

In our case study, the "Citizen" and "Council Member" actors have just one role, although the "Transport Representative" actor is able to play two different roles regarding the information system, as shown in Figure 27.



Figure 27. Actor and roles involved in the Smart City Dashboard

The Responsibility Diagram depicts the different ways that actors may get involved in a system. The "Council Members" and the "Transport Representative" become "Consortium Decisions Makers" when they meet at the "Regional Transport Consortium". These "Consortium Decisions Makers" are responsible for the accomplishment of the task "Analyze Smart City Information". While playing this role, they are supposed to use the Smart City Dashboard in order to make decisions collaboratively and analyze information.

However, some actors can play more than one role with regard to the system. In our case, "Transport Representative" actors can assume the role of "Transport Information Supplier". Since private information owned by transport companies is not initially available in the system, "Transport Representatives"

can be of help to the decision-making process by sharing their private information if deemed necessary by the Regional Transport Consortium. Therefore, they are responsible for facilitating these data by using their company system to connect with the *Smart City Dashboard* system and share the data between both systems. This is represented in the Responsibility Diagram by the guard condition "Logged through Company System" as seen in Figure 27, along with the other roles mentioned.

Regarding the citizens, as explained before, they can use the system to analyze the data on their own, acting as "Citizen Analyst" as shown in Figure 27.

#### A.4 Task Refinement Diagram (TRD)

In order to complete our specification of the Smart City Dashboard system, we have identified the following *Information Requirements* (IR) listed below (shown in Figure 28) aimed at fulfilling the *Information Goals* (IG) described in the previous section. In the following these tasks are explained:

• "Analyze Accident Information": This IR supports the need for analyzing the historical accidents records in each bus route. For this aim, it is decomposed into several information resources. Accidents are represented by the "Accident" business process context and can be grouped according their severity as the "Severity" business intelligence context specifies in the hierarchy illustrated in Figure 28. Next, the "Route" business intelligence context collects information about the "Stops" that constitute a route. It is an information resource used to represent information about routes of the public transport network. Using "Accident" and "Route" business intelligence contexts, we have the location and severity of accidents that occurred in the transport network. However, since we must also analyze if the number of accidents is increasing or decreasing over time, we also need to include time information. The "Day" business intelligence contexts can be used for facilitating different levels of historical analysis. This IR is also decomposed into the "Total Route Accidents" measure, which represents the total value of accidents in a route. However, as this is a collaborative analysis task, we also have to specify the *workspace awareness* that needs to be provided by means of the awareness resource "Acci-

dents Shared Radar" (Gutwin et al. 1996) depicted in Figure 28. This resource enables every decision maker to know *who* are the other analysts involved, *where* they are working, and *what* actions they have performed, in order to avoid possible sources of conflict.

• "Analyze Routes Information": This IR specifies that decision makers analyze the information about the several routes that exist in the public transport network in order to determine whether new routes are necessary. To perform this analysis, decision makers need to examine each stop involved in each bus route. This IR is decomposed into "Stop"→"Route" contexts, as well as into "Estimated Travel Time" measure, as shown in Figure 28. Additionally, as in the previous IR, we need to provide another radar to cover the awareness required by decision makers, to know information about *who* is working with *what* routes as well as *what* actions were performed in the past. To make proactive decisions, decision makers should be aware of future citizens' needs, knowing *what* will be the next status of the different zones (what is the likelihood that they require transportation) and *where* they are or will be located (for developing areas). In order to provide this awareness information, the dashboard will provide feedback about relevant news related to the development of different areas of the city.

• "Analyze Citizen Stop Rating": Within a *Smart City*, citizens are not just information consumers, but also information providers. In this case study, the occupation and punctuality of each bus stop are evaluated by the citizens by using an application. Given that citizens' opinion is critical, the Regional Transport Consortium wishes to analyze the information provided by the citizens. Such information is represented by the measures "Average of Occupation Rating" and "Average of Punctuality Rating" included in this IR. In addition, the analysis needs to include information about the "Stop" and the "Day"→"Month"→"Year" contexts. As in other IRs involving the "Consortium Decision Makers", they require being aware of *who* is analyzing *what* stop to avoid conflicts. Such awareness is specified by using an "Alert Panel" where all the information added is shown.



Figure 28. Information requirements and their respective contexts in Smart City scenario 46

• "Analyze Service Information": Analyze the information about the historical occupancy and delays for each bus service is an IR that the decisions makers of the Regional Transport Consortium need to perform in order to achieve the information goals "Occupancy per stop analyzed" and "Unpunctual stops identified". However, the information resources about services are owned by each transport company, thus they are not initially stored in the BI system of our *Smart City Dashboard* so that they have to be shared by the respective companies. Therefore, these resources are provided by another IR "Transport Information".

- "Analyze Occupancy and delay per route stop": This IR specifies the analysis of the historical service information. It requires several analysts, who can be citizens or consortium members, to collaborate in this task.
- "Provide Transport information": Since Transport Information must be provided previously to carrying out the analysis described in the previous tasks, a temporal restriction has to be defined among them. This explains why "*Provide Transport information*" is related to both "*Analyze service information*" and "*Analyze occupancy and delay per route stop*" in Figure 28. The role "*Transport Information supplier*" is responsible for providing this shared information that is made up of the Business Process *Shared Contexts* (BPSC) "Service" and "Average Stop Delay" as *Shared Measures* (SM).

The aforementioned IRs are carried out collaboratively, so are modeled as Collaborative Information Requirements (CIR). The roles who collaborate in these IRs are the Consortium Decisions Makers, i.e. the Regional Transport Consortium members who use the dashboard system to make decisions regarding public transport. These CIRs capture the needs of collaboration to execute tasks related to each *Decision Goal* and thus establish the collaborative aspects that have to be supported by the system.

We have also highlighted how part of the information cannot be gathered or stored by the system at hand and must be provided by an external entity, such as the transport company. We therefore model here an inter-enterprise collaborative system (Embley and Liddle 2013), in which the actors involved share the

information that is required. As shown, we are able to identify who is sharing the information, i.e. "Transport Information Supplier", as well as what information is being shared.

Regarding the use of the dashboard system by regular citizens, we define the IR "Analyze occupancy and delay per route stop" necessary to fulfill the information goal "Compare occupancy and delays per stop". Because the decision goal pursued is individual, and no other analysts are involved in the analysis task, this IR is not collaborative. However, this situation still depicts a collaborative BI scenario, where actors share certain information across the system although they do not interact with each other to perform the analysis task or achieve common goals.

#### A.5 Quality Factors Diagram (QFD)

In our example, three quality factors (softgoals) were considered, namely "*Citizens' Satisfaction*", "*Service Efficiency for Consortium*" and "*Quality of service*". For instance, two tasks in the TRD contribute positively (Some +) to achieve the quality factor "*Service Efficiency for Consortium*". Such tasks are "Analyze Service Information" and "Analyze citizen Stop Rating" (Figure 29).



Figure 29. Quality factors in the public transport scenario