



University of Augsburg
Prof. Dr. Hans Ulrich Buhl
Research Center
Finance & Information Management
Department of Information Systems
Engineering & Financial Management



Discussion Paper WI-325

A Metadata-based Approach to Leveraging the Information Supply of Business Intelligence Systems

by

Benjamin Mosig, Maximilian Röglinger

December 2011

in: Proceedings of the 31st International Conference on Conceptual Modeling,
Florence, Italy, October, 2012, p. 537-542

A Metadata-based Approach to Leveraging the Information Supply of Business Intelligence Systems

Benjamin Mosig and Maximilian Röglinger

FIM Research Center, University of Augsburg,

Universitätsstraße 12, 86159 Augsburg, Germany

{benjamin.mosig,maximilian.roeglinger}@wiwi.uni-augsburg.de

Abstract. Ensuring adequate information provision continues to be a key challenge of corporate decision making and the usage of business intelligence systems. As a matter of fact, the situation becomes increasingly paradox: Whereas decision makers struggle to specify their information requirements and spend much time on obtaining the information they believe to require, the amount of information supplied by business intelligence systems grows at a speed that makes it hard to keep track. Thus, it is very likely that the required information or suitable alternatives are available, but neither found nor used. Instead, manual searching causes considerable opportunity cost. Existing approaches to information requirements analysis pay attention to incorporate information supply, but do not provide means for leveraging it in a systematic and IT supported manner. As a first step to close this research gap, we propose a metadata-based approach consisting of a procedure model and formalism that help identify a suitable subset of the information supplied by an existing business intelligence system. The formalism is specified using set theory and first-order logic to provide a general foundation that may be integrated into different conceptual modelling approaches.

Keywords: Business intelligence, Data warehouse, Metadata, Information provision, Information supply.

1 Motivation

Ensuring adequate information provision has been a recurring topic over the last decades. Numerous approaches to information requirements analysis and various types of management support systems – most recently business intelligence (BI) systems – have been proposed to increase the clarity in corporate decision making [5]. Despite these efforts, decision makers still face information overload with negative consequences such as mental stress, loss of clarity, and reduced decision quality [1]. A particular problem in recent times is that the convenient access to BI systems and the high storage capacity of underlying data warehouses entice companies into accumulating large amounts of information [4]. As BI systems are historically grown and have been subject to uncontrolled growth in many organizations, it is hard to keep track with the information they supply. Academics approvingly report that “not missing information [is] the primary problem” and that “all information is available somewhere” [6] in most companies. Against this backdrop, it is very likely that the required information or suitable alternatives are available within an organization, but neither found nor used. The potential of existing information supply to satisfy information requirements is not sufficiently tapped [6].

As mentioned, literature contains numerous approaches dedicated to the elicitation and specification of information requirements particularly for the development of data warehouses and BI systems [3, 5]. Apart from few exceptions, the proposed approaches pay attention to incorporating existing information supply. The approaches share several characteristics: First, most activities related to leveraging existing information supply require manual effort and are hardly IT supported. Second, the approaches center on informal or semi-formal concepts, which makes it difficult to cover large amounts of existing information supply systematically. Third, some approaches deal with the initial development of a data warehouse or BI system and thus focus on the information supply of operational information systems. Fourth, the approaches provide no explicit means for coping with decision makers’ struggles when specifying information requirements. Despite the value of the presented approaches, there is a need for additional support to leverage the information supply of existing BI systems. This leads to the following research question: *How can the information supply of existing BI systems be leveraged in a systematic and IT supported manner?*

We address the research question by proposing a metadata-based approach consisting of a procedure model and formalism that complement the approaches discussed above and help identify a suitable subset of the information supplied by an existing BI system. We rely on metadata because they play an important role in BI systems and have the potential to structure large amounts of data [2]. Following an axiomatic and deductive research approach, we first sketch the general setting and explicate our assumptions (section 2). We then derive the procedure model on this foundation (section 3). The paper concludes with a discussion (section 4). The formalism is specified using set theory and first-order logic. It has been omitted due to space restrictions, but can be requested from the authors.

2 General Setting

Our unit of analysis is a single historically grown BI system that is based on a data warehouse as informational infrastructure. The data warehouse is based on a multi-dimensional data schema whose core elements on schema level are measures and dimensions. We treat all dimensions as orthogonal and abstract from structural abnormalities such as parallel hierarchies [3]. While an examination on the schema level is reasonable in the context conceptual modeling, information requirements analysis extends to the instance level because information requirements typically relate to the actual values of measures and hierachic levels. We assume:

- (A.1) The multi-dimensional data schema consists of measures and dimensions. Each dimension includes hierachic levels.

To incorporate metadata into the procedure model and the formalism, information requirements need to be split into two parts where the first part includes requirements that directly relate to the core elements of the multi-dimensional data schema and the second part comprises requirements that relate to meta-attributes (see Table 1).

Table 1. Considered components of the information requirements

	<i>Related to the core elements of the multi-dimensional data schema</i>	<i>Related to additional meta-attributes</i>
<i>Schema level</i>	<ul style="list-style-type: none">• Requirements regarding measures• Requirements regarding dimensions• Requirements regarding hierachic levels	
<i>Instance level</i>	<ul style="list-style-type: none">• Requirements regarding the domain of selected measures• Requirements regarding the domain of selected hierachic levels	<ul style="list-style-type: none">• Requirements regarding the value of a meta-attribute <u>for each single</u> selected measure• Requirements regarding the value of a meta-attribute <u>for all</u> selected measures

The first part of the information requirements helps specify requirements where the decision makers know precisely which combinations of measures and dimensions they need. These requirements can be elicited using the existing approaches to information requirements analysis. As known from conceptual modeling, there is a dependency between requirements on schema level and on instance level. That is, requirements regarding the instance level of measures or hierachic levels relate to the domains of the measures or hierachic levels selected on schema level.

Requirements belonging to the second part of the information requirements relate to additional meta-attributes. What is special about using meta-attributes is that usually multiple subsets of the information supply exist that meet the related requirements. Requirements can be defined at two distinct reference levels (see Table 1). Either

each single selected measure has to fulfill a requirement individually (reference level: each single measure) or all selected measures together have to fulfill a requirement (reference level: all measures). Moreover, the collection effort of all selected measures must not exceed a defined limit (meta-attribute: ‘collection effort’). We assume:

- (A.2) Each measure features the same meta-attributes. Meta-attributes are only assigned to measures.
- (A.3) The information requirements $I^{\text{req}} = \{F^{\text{model,schema}}, F^{\text{model,instance}}, F^{\text{meta}}\}$ comprise requirements related to core elements of the multi-dimensional data schema on schema level, $F^{\text{model,schema}} = f_1^{\text{model,schema}} \wedge \dots \wedge f_s^{\text{model,schema}}$, requirements related to the core elements of the multi-dimensional data schema on instance level $F^{\text{model,instance}} = f_1^{\text{model,instance}} \wedge \dots \wedge f_t^{\text{model,instance}}$, and requirements related to meta-attributes $F^{\text{meta}} = f_1^{\text{meta}} \wedge \dots \wedge f_u^{\text{meta}}$. All requirements are specified in first-order logic. $F^{\text{model,schema}}$ only contains requirements that can be covered by the information supply.
- (A.4) The subset of I^{supply} that meets all requirements related to the core elements of the multi-dimensional data schema on schema level ($F^{\text{model,schema}} = \top$) is denoted by $I^{\text{selected,model,schema}}$. The subsets of I^{supply} that meet all requirements related to meta-attributes ($F^{\text{meta}} = \top$) are denoted as set family $(I^{\text{selected,meta}})_v$, where V is an index set, $|V|$ is the number of different sets, and $v \in V$.

Due to the logical AND operator (\wedge) in (A.3), $F^{\text{model,schema}}$, $F^{\text{model,instance}}$, and F^{meta} only evaluate to *true* (\top) if all respective requirements are met. Each of the v set unions $I^{\text{selected,model,schema}} \cup (I^{\text{selected,meta}})_v$ – subsequently referred to as I_v – is a feasible alternative containing the required information on schema level. $F^{\text{model,instance}}$ has not been considered so far as the enclosed requirements can partly be formulated after one of the I_v has been selected. In order to determine which of the I_v should be selected, we assume that the decision makers assess the utility and disutility of each alternative.

- (A.5) Decision makers strive to maximize the net benefit they receive from the selected subset of the information supply. Each measure m_p features a subjective utility value $u(m_p)$ and disutility value $d(m_p)$. The utility and disutility values of a particular subset of the information supply I_v are calculated as follows: $U(I_v) = \sum_{m_p \in \{m_z | \exists (m_z, \dots) \in I_v\}} u(m_p)$ and $D(I_v) = \sum_{m_p \in \{m_z | \exists (m_z, \dots) \in I_v\}} d(m_p)$. The overall net benefit is calculated as $U^{\text{net}}(I_v) = U(I_v) - D(I_v)$.

Finally, we need to know how decision makers specify their information requirements. Based on our experience from related industry projects, we assume:

- (A.6) Decision makers base their information requirements primarily on measures. Moreover, decision makers are able to specify information requirements related to the core elements of the multi-dimensional data schema and requirements related to meta-attributes independent of one another.

3 Procedure Model

Based on the elaborations concerning the general setting, we are able to derive properties of a procedure model for leveraging the information supply of existing BI systems. The overall procedure model is shown in Fig. 1. First, the procedure model can start with the simultaneous specification of requirements from $F^{\text{model,schema}}$ regarding measures as well as F^{meta} . This is because decision makers base their information requirements primarily on measures (see A.6) and meta-attributes are only assigned to measures (see A.2). This results in steps ① and ② of the procedure model. Second, the utility and disutility of the selected measures can be assessed directly afterwards as only measures are assessed (see A.5). This results in steps ③ and ④ of the procedure model. Due to the interdependency of requirements on schema level and on instance level, the requirements regarding dimensions and hierachic levels from $F^{\text{mod-el,schema}}$ have to be specified first. After that, $F^{\text{model,instance}}$ can be formulated when it comes to report parameterization. This results in steps ⑤ and ⑥ of the procedure model. The position of steps ③ and ④ is also reasonable because labour-intense effort is reduced as decision makers would otherwise have to assess the (dis-) utility of measures, dimensions, and dimensional hierarchy levels that are not implemented.

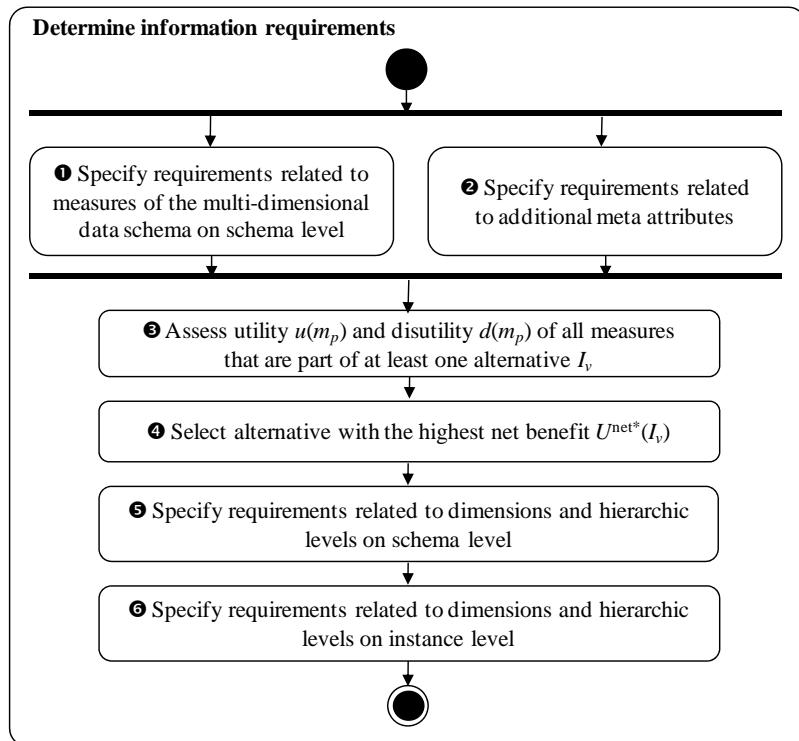


Fig. 1. Procedure model for leveraging the information supply of existing BI systems

4 Discussion

The proposed approach is beset with limitations that need to be taken into account when applying it in industry settings. Other limitations motivate future research endeavors:

1. Prior to application, appropriate meta-attributes have to be identified and – if not already available – filled with values. While this may be quite costly for a single use case, it is worth the effort in case of repeated applications for multiple groups of decision makers. As Stroh et al. [5] point out, information requirements analysis is not a one-time project, but a continual process. The proposed formalization based on metadata is a first step in this direction since it enables the realization of meaningful automation potential.
2. The approach restricts itself to consider existing information supply – which will in general not fully satisfy a decision maker's information requirements. In this case, the remaining parts of the information requirements have to be covered using existing approaches to information requirements analysis.
3. Although IT support plays an important role, an implementation is pending. This is already part of on-going research and will shape up useful for practical application and evaluation issues.
4. The information requirements are currently treated as constant. While this is approximately appropriate for standard reporting and well-structured problems, it is not always the case in a complex and disruptive business environment.

Despite its limitations, the proposed approach is a first step to address the research gap of leveraging the information supply of existing BI systems and towards an enhanced usage of metadata in the context of BI systems.

References

1. Bawden, D; Robinson, L: The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science* 35(2), 180-191 (2009)
2. Foshay, N; Mukherjee, A; Taylor, A: Does Data Warehouse End-User Metadata Add Value? *Communications of the ACM* 50(11), 70-77 (2007)
3. Kimball, R; Ross, M; Thornthwaite, W; Mundy, J and Becker, B: *The Data Warehouse Lifecycle Toolkit*. 2. John Wiley & Sons, Indianapolis (2008)
4. Oppenheim, C: Managers' use and handling of information. *International Journal of Information Management* 17(4), 239-248 (1997)
5. Stroh, F; Winter, R; Wortmann, F: Method Support of Information Requirements Analysis for Analytical Information Systems. State of the Art, Practice Requirements, and Research Agenda. *Business & Information Systems Engineering* 3(1), 33-43 (2011)
6. Winter, R; Strauch, B: A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects. In: Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), 231.1 (2003)