Modelling Knowledge about Data Analysis Processes in Manufacturing

Thomas Neuböck* Michael Schrefl**

* solvistas GmbH, Graben 18, 4020 Linz, Austria (e-mail:
thomas.neuboeck@solvistas.com).
** Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz,
Austria (e-mail: schrefl@dke.uni-linz.ac.at)

Abstract: In industry 4.0, analytics and business intelligence (BI) are of particular importance to increase productivity, quality, and flexibility. It is necessary to make right and quick decisions for effective and efficient problem solving and process improvements. Modern technologies allow to collect a large amount of data that can be analysed. Heterogeneity and complexity of industrial environments require considerable expert knowledge to perform meaningful and useful data analysis. BI analysis graphs represent expert knowledge about analysis processes. This knowledge can be modelled pro-actively at schema level and used at instance level. Analysis situations can be considered as multi-dimensional queries and represent nodes of a BI analysis graph. An arc between two nodes is a relationship between two analysis situations describing the difference of both. It represents a navigation step, e.g., an online analytical processing (OLAP) operation, of the analysis process. We demonstrate BI analysis graphs by a use case originated from manufacturing of brushes. Complex analysis paths, e.g., to analyse substitute material in the case of delayed delivery, are modelled by BI analysis graphs and can be used multiple times (also by non-experts). Reinvention of analysis knowledge is prevented – right and quick decisions for finding effective and efficient problem solutions can be made.

Keywords: decision making, knowledge representation, data models, process models, manufacturing, industry 4.0, data warehouses, business intelligence

1. INTRODUCTION

In many areas, business analysts need to explore a large amount of data to answer management questions or questions of other stakeholders. In most cases the data is heterogeneous and considerable expert knowledge is required to perform meaningful and useful data analysis. Based on the result of data analysis, strategic decisions (e.g., decision whether or not to outsource the production of a whole product line) or operative actions (e.g., ordering appropriate substitute material for production, if there is a serious delayed delivery of planned material) are made depending on the given analysis goal. General contributions to the integration of various decision levels in manufacturing companies can be found in Gerber et al. (2012).

Business intelligence (BI) and analytics give a wide range of opportunities for comprehensive data analyses. Data is collected in data warehouses, organized in multi-dimensional cubes, and queried by online analytical processing (OLAP) operations. Whereas it is common to model "static" knowledge about the underlying data, e.g., as a dimension fact model (DFM), see Golfarelli et al. (1998), there are no appropriate means for modelling "dynamic" knowledge about analysis processes – such as, e.g., it can be found analogously in business process modelling notation (BPMN), see Silver (2011).

Industry 4.0 — an initiative of the German government — has become a new catchword emphasizing a "new industrial revolution" that automates customization of products on demand. Whereas in conventional production systems large quantities of a small range of products are manufactured, in industry 4.0 companies have to produce large quantities of a wide range of items with many options of individual customer configurations (mass customization). Production systems are coped with massive order-related manufacturing. With respect to these trends, smart factories and internet of things are visions that become reality, see Zühlke (2009). Dealing with big data issues for rapid decision making to improve productivity rises new challenges for companies, see Lee et al. (2014). In this context, analytics and business intelligence are of particular importance to increase productivity, quality, and flexibility. Right and quick decision making is necessary to guaranty effective and efficient problem solutions and process improvements. Data exploration is one of the key factors necessary for this endeavour. Heterogeneity and complexity of industrial environments issues a challenge to business analysts – considerable expert knowledge about data analysis processes in industrial environments is required.

To overcome these analysis requirements for industry 4.0, data must be integrated quickly into data warehouses and actions, as a consequence of the analysis result, should be executed automatically. Near real-time data warehouses

set the focus to fast data integration, see Bruckner et al. (2002). Active data warehouses offer support to automate the routine elements of decision tasks by extending conventional data warehouse architecture with analysis rules, see Thalhammer et al. (2001). Analysis and judgement rules can be extended for ontology-driven comparative data analysis, see Steiner et al. (to appear in 2015).

Another crucial point for a successful analysis environment in manufacturing is the provision of expert knowledge and its flexible application. E.g., in the case of analysing substitute material, an analyst must have knowledge about material properties, about its usage within the production process, and about customers' requirements. Knowledge about data is made visible by elaborating conceptual models, e.g., dimension fact models. Business and technical terms are documented in business glossaries. The meaning of measures is described by mappings that relate original attributes of data sources to measures used for analysis. In contrast to these modelling and documentation support there exists no adequate means to model analysis processes itself, although these contain tacit valuable expert knowledge. A business analyst performs an analysis and evaluates the results that again induces a subsequent analysis, and so on. The difference between two analysis situations can be considered as an application of the analyst's expertise, or, in other words, navigation from one analysis situation to another one represents knowledge.

To achieve this type of knowledge integration, analysis processes should be modelled *pro-actively* to provide analysis guidance. Operational business processes are characterized by online transaction processing (OLTP) allowing a high degree of automation. In contrast, analysis processes can be considered as semi-routine processes comprising routine and non-routine elements. In analysis processes query execution and user interaction about how to proceed in the process (depending on the analysis situation) alternate frequently. Whereas other approaches about analysis processes focuses on evaluation of analysis situations' instances — see Romero et al. (2011), Giacometti et al. (2009), and Jerbi et al. (2009) — our approach sets priorities to pro-active modelling at schema level.

The focus of our approach lies on preparation for decision-making. The aim is to prepare decision-relevant data for management, e.g., required in emergency meetings. A business analyst searches for opportunities and impacts to support decision-making. To support the business analyst we provide more than fixed guided reports. By *pro-active modelling*, analysis processes can be flexibly defined, adapted, and subsequently executed. The business analyst gets suitable analysis guidance.

In this paper we show how knowledge about data analysis processes can be modelled *pro-actively* and used in the field of manufacturing. The approach is based on BI analysis graphs, see Neuböck et al. (2012) – our previous work focused on the instance level of analysis situations and not on modelling analysis situation schemas pro-actively. A BI analysis graph consists of analysis situations as nodes representing multi-dimensional queries and navigation operations (e.g., OLAP operations) as arcs representing relationships between analysis situations (the difference of two analysis situations). In the paper in hand we show
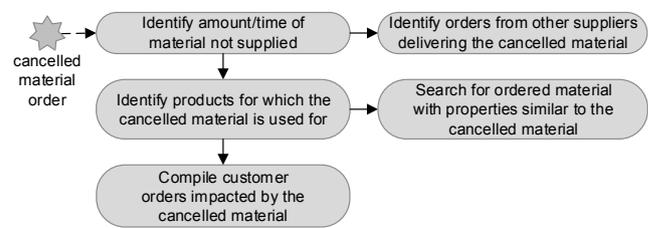

Fig. 1. Analysis graph (undetailed view)

how knowledge about analysis processes can be modelled at schema level (modelling) and used at instance level (analysing). The approach is demonstrated by a use case from the manufacturing area (production of various types of brushes with a high level of customization). It gives an understanding of how complex analysis paths can be modelled and used multiple times (also by non-experts; preventing reinvention). Right and quick decisions can be made to solve acute issues or to enforce effective and efficient process improvements.

Section 2 introduces the manufacturing use case. Section 3 presents the schema of a single analysis situation and shows how an analysis situation is instantiated. Analysis graph schemas are described in section 4 starting with the schema of a single navigation operation. Linking a set of analysis situations schemas by navigation operations returns the schema of an analysis graph. Section 5 shows the use of analysis graphs. A business analyst traverses analysis situations including backtracking. Finally, section 6 concludes our presentation with an outlook.

2. USE CASE

Our use case falls into the field of manufacturing brushes of various types. It comes from our Austrian business partner KOTI Kobra — a member of the European company group KOTI. To satisfy customers' requirements, the company has to offer both large scale production and strong customization leading up in an extreme case to a batch size of one. The focus lies on an order-related production process. Manufacturing of brushes covers a wide application range, i.e., everything that consists of a body material and bristles can be considered as a brush. Industrial and technical brushes, strip and sealing brushes, work tool brushes, sweeping and cleaning brushes, runway brushes, hygiene brushes, or entrance brush mats are important brush applications. Brushes are produced for various markets, e.g., automotive, airport and winter equipment, chemical industry, electronics, food and beverage industry, etc. Various production parameters are important for customization: base types of brushes (strip brush, roller brush, brush discs), body and bristle material (e.g., with respect to temperature resistance, lifetime, mechanical load, etc.), number and ordering of drill holes for bristles, colour, etc.

KOTI Kobra is faced with analysis tasks for solving strategic and operational issues. Here we demonstrate an analysis process that is triggered by a cancelled material order. In this case the situation and the courses of actions must be analysed and relevant information has to be gathered (preparation stage for decision making). The result of this analysis process can be used in an emergency meeting of the management as a well-founded basis for
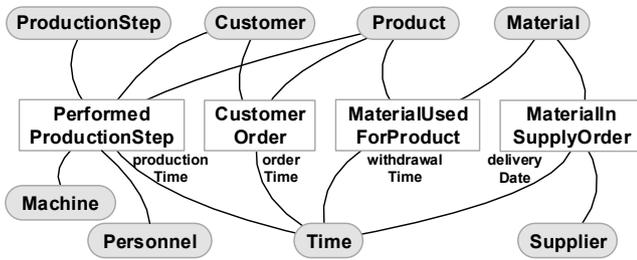
Fig. 2. Overview facts and dimensions

decision making. Fig. 1 depicts an undetailed view (bird's eye view) of a so called analysis graph that sketches an analysis process for our use case. [1] If a material order is cancelled or considerably delayed, the business analyst has to identify the amount and time of the missing material of the cancelled order. Maybe the cancelled material is contained in orders from other suppliers that can serve as a short-term compensation for the cancelled order. Thus the analyst searches for orders from other suppliers containing the cancelled material. One retrieves a list of such suppliers that comprises delivery date, ordered quantity, and costs. Additionally, the business analyst identifies products the cancelled material is used for. This report shows the urgency of the missing material. Next it is important to compile customer orders impacted by the cancelled material. With this information the management elaborates a priority list of critical and less critical customer orders and searches for scope of action. Another possibility for mitigating the issue lies in searching alternative material. The aim is to find material with properties similar to the cancelled material or with sufficient properties that also satisfies the customers' requirements. To get essential information, the analyst must check product requirements and evaluate material with regard to appropriate properties.

Pro-active modelling allows flexible definition and rapid usage of analysis processes. As a company that prefers just in time production to reduce storage and capital commitment costs, and, nevertheless, to satisfy customers' requirements for quality and on-time delivery, it is important to rapidly react to delayed or cancelled supply of production material. Especially, the situation is exacerbated in that considerable expert knowledge is required to analyse effects of and solutions for missing supplies — e.g., about five hundred different items of trimming material (bristles) and more than four hundred different items of body material are used satisfying various costumer requirements (increasing complexity of the analysis process).

Fig. 2 presents an overview of the data of our simplified use case description. Business events (e.g., MaterialInSupplyOrder) are stored as facts (cubes) in a data warehouse. Facts can be analysed with respect to dimensions (e.g., Supplier). Both dimensions and facts are tables organized in a star schema that can be conceptually visualized by a dimension fact model (DFM). In the subsequent presentation, we only use facts CustomerOrder, MaterialUsedForProduct, and MaterialInSupplyOrder, and their related dimensions. Actually, there are more than hundred fact and
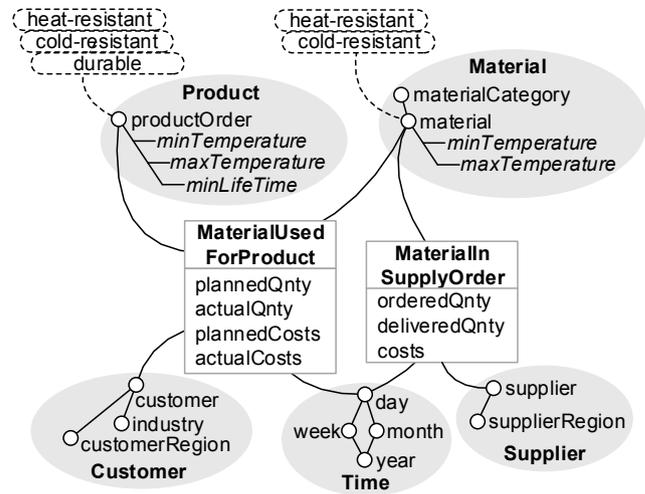
dimension tables containing analysis data, e.g., PerformedProductionStep, Machine, Personnel, or ProductionStep.

Fig. 3 shows two stars containing material used for products and material in supply orders as fact tables. Ordered quantity, delivered quantity, and costs are measures of fact table MaterialInSupplyOrder. Dimensions consist of dimension hierarchies that comprises dimension levels, e.g., dimension Time contains a hierarchy comprising dimension levels day (date), month, and year. The levels day, week, and year form another hierarchy of dimension Time. A measure is related to the finest granularity level, e.g., the measures of MaterialInSupplyOrder are stored per material, supplier, and day. They can be aggregated to higher dimensional levels, e.g., sum of ordered quantities over all suppliers.

Concrete elements of dimensions are called dimension nodes (e.g., *21 October 2014* is a node of dimension Time at level day; *2014* is another node of dimension Time at level year). A dimension node can be described by additional attributes (describing attributes) related to a level, e.g., minTemperature and maxTemperature are describing attributes of level material in dimension Material. These attributes describe the temperature range of the material's application. Describing attributes can be used to define predicates that can be applied in queries, e.g., predicate heat-resistant is defined at level material as $maxTemperature \geq 100$. Predicates can be extended to an ontological approach we will not deal with in the present paper. For details see Neuböck et al. (2014).

Based on cubes, a business analyst can execute queries and apply OLAP operations. E.g., she or he takes cube MaterialInSupplyOrder, selects node *2014* in dimension Time at level year and node *Bristle Material* in dimension Material at level materialCategory (selected nodes and levels are called dice node and dice level respectively), restricts the query to heat-resistant material (predicate heat-resistant), and lists the sum of ordered quantity (measure orderedQnty) per material (level material). Next the user can, e.g., move down from node *2014* to a subordinate node *May 2014* or drill down additionally to level day, such that the sum of measure orderedQnty is additionally listed per day.

Fig. 3. Dimension Fact Model (DFM)

---

[1] For demonstration the illustrated analysis process is a simplification of the real one.
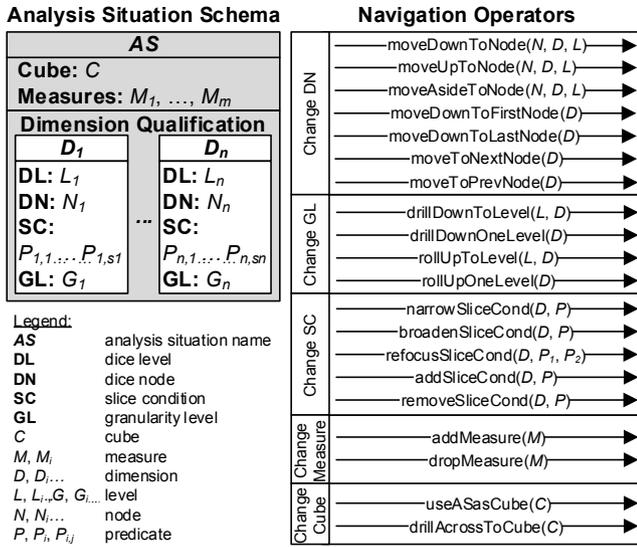
Analysis Situation Schema

Fig. 4. Analysis situation schema and navigation operators

3. ANALYSIS SITUATION

An analysis situation represents a multi-dimensional query. It is modelled at schema level by an analysis situation schema (see Fig. 4, left side) which comprises an identifier AS, a cube C, m measures $M_1 \cdots M_m$, and a dimension qualification. The measures must be defined over cube C. The dimension qualification specifies the selected nodes in each dimension and defines the granularity in each dimension, i.e., the level at which measures are aggregated. Each dimension D_i ($1 \leq i \leq n$) must be a dimension of cube C. For dimension D_i the modeller of an analysis situation can state a dice node N_i at dice level L_i meaning that the multi-dimensional query includes node N_i and all its subordinate nodes. Per dimension there is a special level top representing the highest level to which all dimension nodes can be rolled up. all is the single node of level top. Moreover, the dice node and all of its subordinate nodes can be restricted by an additional slice condition that consists of s_i predicates $P_{i,1} \cdots P_{i,s_i}$. The conjunction $P_{i,1} \wedge \cdots \wedge P_{i,s_i}$ of all predicates forms the slice condition which only selects those nodes of the dice node that satisfies all predicates. Finally, granularity G_i defines the level of dimension D_i the measures are aggregated. The pair dice level and dice node can be missing, meaning that analysis situation AS defines dice level top and dice node all for the corresponding dimension. A missing slice condition expresses that there are no further restrictions, i.e., it is equivalent to slice condition with constant truth value true. If there is no granularity definition, the granularity level of the corresponding dimension is top. Instead of concrete values for measures, dice levels, dice nodes, predicates, and granularities, one can use variables. Prefix ? is used to denote variable names syntactically (e.g., ?mat, ?tm, etc.).

Fig. 5 shows four examples of analysis situation schemas.[2] Using AS2, a business analyst can retrieve a list of suppliers comprising delivery date, ordered quantity, already delivered quantity, and costs of ordered material ?mat within a certain time period ?tm. Maybe ordered items can

---

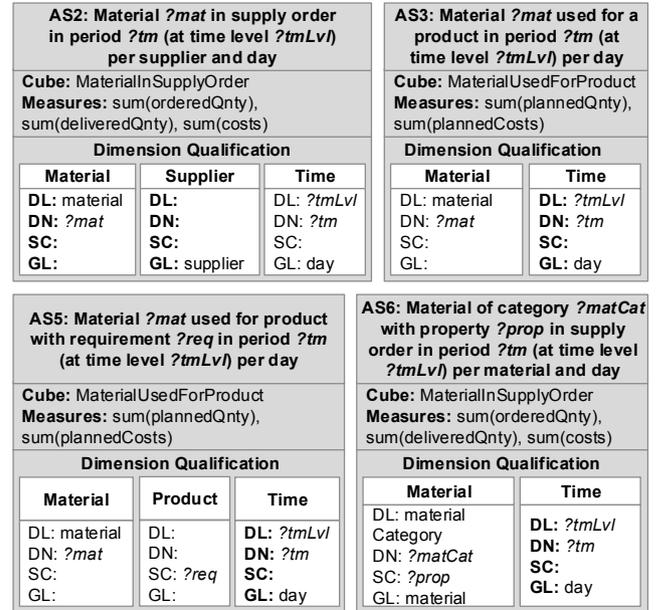[2] Analysis situation schemas AS1 and AS4 are explained later.

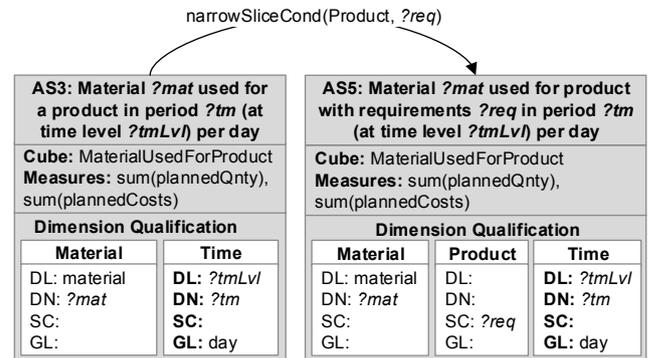Fig. 5. Four examples of analysis situation schemas

Fig. 6. Example of a navigation step schema

serve as a short-term compensation for the cancelled order. Using AS3, the business analyst can identify scheduled production that would need the cancelled material. From the resulting report, one can derive the urgency of the missing material. To search for alternative material, a business analyst can use AS5 for checking product requirements and AS6 for evaluating ordered material with regard to appropriate properties.

To use an analysis situation, one binds its variables, e.g., for AS2 variable ?mat is bound to *Polymex 0.18 black*, ?tmLvl to month, and ?tm to *Oct. 14*. An instantiated analysis situation represents an executable OLAP query. Tab. 1 shows a possible query result.

Table 1. Query result

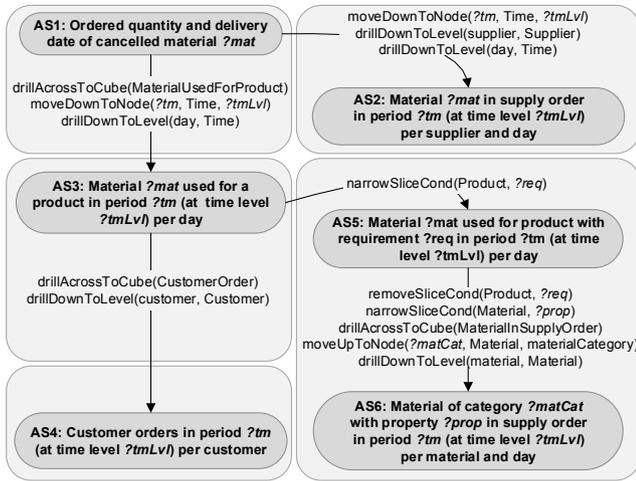| supplier	day	sum(orderedQnty)	sum(deliveredQnty)	sum(costs)
4711	2014-10-01	100	100	2,200.00
4711	2014-10-31	300	0	6,000.00
9966	2014-10-22	200	0	4,100.00

Fig. 7. Analysis graph (detailed view)

Fig. 8. Example of a navigation step instance

4. ANALYSIS GRAPH SCHEMA

Two analysis situation schemas [3] can be linked. A link represents a navigation step that also can contain variables. A navigation step corresponds to one or more OLAP operations (e.g., drill down, roll up, etc.) that indicate how an instance of the target situation schema is derived from an instance of the source situation schema. The right side of Fig. 4 shows available navigation operators. They are grouped into operators each changing one component of the source analysis situation — changing dice node (and possibly dice level), granularity, slice condition, measure, or cube. E.g., moveDownToNode(N, D, L) moves down to node N which is situated at finer granularity level L of dimension D, drillDownToLevel(L, D) changes granularity of dimension D to a finer granularity level L, narrowSliceCond(D, P) restricts slice condition of dimension D to predicate P, removeMeasure(M) removes measure M from the query result, and drillAcrossToCube(C) leaves the cube of the source situation and moves to cube C respecting common dimensions. The analysis graph schema is a directed graph of nodes representing analysis situation schemas with arcs as navigation steps.

Fig. 6 shows a navigation step from AS3 to AS5, where slice condition in AS3 is narrowed in dimension Product to predicate ?req resulting in AS5. A detailed view of the analysis graph of Fig. 1 can be found in Fig. 7. By AS1, ordered quantity and delivery date of the cancelled material can be obtained. To look for opportunities of short-term compensation, a list of suppliers and delivery dates can be retrieved by AS2. A report can be generated by AS3 listing the days where the cancelled material was scheduled for production (detecting the urgency of the missing material). It is important to rank customer orders with respect to critical and less critical ones. Analysis situation AS4 can be used to retrieve customer orders impacted by the missing material. Searching for substitute material represents another alternative to cope with the issue of cancelled material. Product requirements can be evaluated by AS5. One can navigate to AS6 to identify material of corresponding categories and with appropriate

properties, such that the product requirements can still be satisfied.

5. ANALYSIS

Once an analysis graph schema has been modelled it can be used multiple times. An analysis corresponds to a sequence of instantiated analysis situations which are retrieved from an analysis graph following the modelled navigation steps (arcs). Instantiating a navigation step, the user binds all its variables. Thereby variable bindings of the source analysis situation are transferred to the target analysis situation according to the navigation operator associated with the navigation step.

Fig. 8 shows an instantiated navigation step with instantiated source and target analysis situations. In AS3 the analyst identifies all days of October 2014 the missing material *Polymex 0.18 black* is planned for production. The same query is executed in AS5 except product requirements are restricted to *heat-resistant*. The analyst intends to evaluate cancelled material with respect to special product requirements.

An execution of an analysis graph, i.e., an analysis, can be represented as an analysis trace. Such a trace may also contain backtracking steps to a previous analysis situation (not modelled explicitly).

Tab. 2 shows an excerpt of an analysis trace of the analysis graph in Fig. 7. Analysis situation instances are listed in rows. Column AS/I contains analysis situation identifier (AS) and instance number (I). The value a variable is bound to is underlined. AS3/1 is an instance of analysis situation AS3. The analyst focuses on material *Polymex 0.18 black* planned for production in October 2014. In the next step (AS3/5) she or he is interested in products satisfying heat-resistance where *Polymex 0.18 black* is used. Subsequently, the analyst wants to retrieve a list containing material of category *Bristle Material* having property *heat-resistant* (AS6/1). The aim is to identify possible substitute material satisfying special product requirements, too. In the last visible row (AS3/2) of Tab. 2, a second instance of analysis situation AS3 is listed. The analyst changes the focus to material *PA 66, 0.25 white* for searching again appropriate substitute material.

---

[3] If the context is unambiguous, we also use – for abbreviation – the notion analysis situation instead of analysis situation schema and the notion analysis graph instead of analysis graph schema.
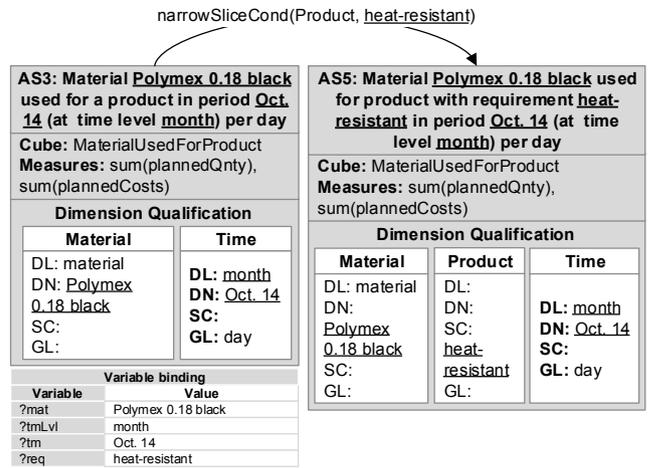
The pro-actively modelled analysis graph of the presented use case provided support to find meaningful analysis situations for solving the issue of cancelled material orders. Various courses of actions could be identified (material of other suppliers, substitute material with sufficient properties, identification of critical customer orders, etc). Although the analysis graph did not return concrete decisions, it provided assistance how to proceed in the analysis process to get a basis for effective and efficient decision-making.

6. CONCLUSION

Pro-active modelling of analysis processes becomes more significant in the context of industry 4.0. The requirements of strong customization increases the analysis complexity that impacts pace of decision-making and increases the risk of wrong decisions. Pro-active modelling of analysis processes by BI analysis graphs allows to cope with increasing complexity. The analysis process is immediately available when needed and provides analysis guidance.

In the past KOTI Kobra used means of OLTP systems for data analysis. Analysis processes were implemented only partially and hard-coded as parts of business processes without flexibility for adaptation. Nowadays the company has established a data warehouse and uses OLAP cubes. Pure OLAP allows new flexibility in the sense of a high degree of analysis capabilities. Nevertheless, there is a lack of analysis guidance that can be adapted rapidly depending on fast changing business situations. Canned and hard-linked reports provide less flexibility. Our pro-active modelling approach allows that analysis processes based on OLAP cubes can be implemented, adapted, and executed immediately. The modelled analysis guidance provides expert knowledge and variability that respects users' creativity necessary for problem solving. The user is supported to navigate through comprehensive and highly interrelated data structures in an effective and efficient way.

The presented approach was evaluated by establishing and using a data analysis manual of analysis graphs for several use cases but without specific tool support. Currently we develop a prototype tool for modelling and using analysis graphs that includes also further modelling primitives such as generalization and composition of analysis graphs.

REFERENCES

Bruckner, R.M., List, B., and Schiefer, J. (2002). Striving towards near real-time data integration for data warehouses. In *Data Warehousing and Knowledge Discovery, 4th International Conference, DaWaK 2002, Aix-en-Provence, France, September 4-6, 2002, Proceedings*, 317–326.

Gerber, T., Bosch, H.C., and Johnsson, C. (2012). Vertical integration of decision relevant production information into it-systems of manufacturing companies. In *Proceedings of the 14th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 12), May 23-25, Bucharest, Romania*. IFAC.

Giacometti, A., Marcel, P., Negre, E., and Soulet, A. (2009). Query recommendations for OLAP discovery driven analysis. In I.Y. Song and E. Zimányi (eds.), *DOLAP*, 81–88. ACM.

Golfarelli, M., Maio, D., and Rizzi, S. (1998). The dimensional fact model: A conceptual model for data warehouses. *Int. J. Cooperative Inf. Syst.*, 7(2-3), 215–247.

Jerbi, H., Ravat, F., Teste, O., and Zurfluh, G. (2009). Preference-based recommendations for OLAP analysis. In T.B. Pedersen, M.K. Mohania, and A.M. Tjoa (eds.), *DaWaK*, volume 5691 of *Lecture Notes in Computer Science*, 467–478. Springer.

Lee, J., Bagheri, B., and Kao, H.A. (2014). Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. In *IEEE Int. Conference on Industrial Informatics (INDIN) 2014*. IEEE.

Neuböck, T., Neumayr, B., Rossgatterer, T., Anderlik, S., and Schrefl, M. (2012). Multi-dimensional navigation modeling using BI Analysis Graphs. In S. Castano, P. Vassiliadis, L.V. Lakshmanan, and M.L. Lee (eds.), *ER Workshops*, volume 7518 of *Lecture Notes in Computer Science*, 162–171. Springer.

Neuböck, T., Neumayr, B., Schrefl, M., and Schütz, C. (2014). Ontology-driven business intelligence for comparative data analysis. In *Business Intelligence - Third European Summer School, eBISS 2013, Dagstuhl Castle, Germany, July 7-12, 2013, Tutorial Lectures*, volume 172 of *Lecture Notes in Business Information Processing*, 77–120. Springer.

Romero, O., Marcel, P., Abelló, A., Peralta, V., and Bellatreche, L. (2011). Describing analytical sessions using a multidimensional algebra. In A. Cuzzocrea and U. Dayal (eds.), *DaWaK*, volume 6862 of *Lecture Notes in Computer Science*, 224–239. Springer.

Silver, B. (2011). *BPMN Method and Style, 2nd Edition, with BPMN Implementer's Guide: A Structured Approach for Business Process Modeling and Implementation Using BPMN 2.0*. Cody-Cassidy Press.

Steiner, D., Neumayr, B., and Schrefl, M. (to appear in 2015). Judgement and analysis rules for ontology-driven comparative data analysis in data warehouses. In *Proceedings of the 11th Asia-Pacific Conference on Conceptual Modelling (APCCM 2015)*.

Thalhammer, T., Schrefl, M., and Mohania, M.K. (2001). Active data warehouses: complementing OLAP with analysis rules. *Data Knowl. Eng.*, 39(3), 241–269.

Zühlke, D. (2009). SmartFactory - a vision becomes reality. In *Keynote Papers of the 13th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 09), June 3-5, Moscow, Russian Federation*. ICS / RAS.

Table 2. Analysis trace

AS/I	Description
...	...
AS3/1	Material Polymex 0.18 black used for a product in period Oct. 14 (at time level month) per day
AS5/1	Material Polymex 0.18 black used for product with requirement heat-resistant in period Oct. 14 (at time level month) per day
AS6/1	Material of category Bristle Material with property heat-resistant in period Oct. 14 (at time level month) per material and day
AS3/2	Material PA 66, 0.25 white used for a product in period Oct. 14 (at time level month) per day
...	...