The Manufacturing Knowledge Repository
Consolidating Knowledge to Enable Holistic Process Knowledge Management in Manufacturing

Christoph Gröger, Holger Schwarz, Bernhard Mitschang[1]
[1]*Institute of Parallel and Distributed Systems, University of Stuttgart, Universitätsstr. 38, 70569 Stuttgart, Germany*
{christoph.groeger, holger.schwarz, bernhard.mitschang}@ipvs.uni-stuttgart.de

Keywords: Knowledge Integration, Manufacturing, Process Optimization

Abstract: The manufacturing industry is faced with strong competition making the companies' knowledge resources and their systematic management a critical success factor. Yet, existing concepts for the management of process knowledge in manufacturing are characterized by major shortcomings. Particularly, they are either exclusively based on structured knowledge, e. g., formal rules, or on unstructured knowledge, such as documents, and they focus on isolated aspects of manufacturing processes. To address these issues, we present the Manufacturing Knowledge Repository, a holistic repository that consolidates structured and unstructured process knowledge to facilitate knowledge management and process optimization in manufacturing. First, we define requirements, especially the types of knowledge to be handled, e. g., data mining models and text documents. On this basis, we develop a conceptual repository data model associating knowledge items and process components such as machines and process steps. Furthermore, we discuss implementation issues including storage architecture variants and finally present both an evaluation of the data model and a proof of concept based on a prototypical implementation in a case example.

1 INTRODUCTION

Today, manufacturing companies are exposed to intense competition due to globalization, high market volatility and rapid technological changes (Monauni and Foschiani, 2013). In addition, worldwide homogenization and dissemination of production technologies and materials diminish the competitive potential of tangible assets. Thus, knowledge, that is the intangible intellectual capital of a company, becomes a critical source for competitive advantages emphasizing the need for a systematic knowledge management (Goossenaerts *et al.*, 2005).

Existing knowledge management systems in manufacturing mainly focus on product knowledge and customer knowledge. For example, knowledge-based engineering systems integrate computer aided design (CAD) data and additional product knowledge to enrich product models (Chapman and Pinfold, 2001). Yet, there are only rudimentary concepts for the management of process knowledge in manufacturing.

Existing approaches are characterized by three major shortcomings limiting process knowledge management and continuous process improvement: (i) they are either exclusively based on structured knowledge, e. g., formal rules, or they only deal with unstructured knowledge like documents; (ii) they make use of tailored and application-specific databases to store knowledge items; (iii) they focus on isolated aspects of manufacturing processes, e. g., specific resources, or selected phases of the process lifecycle, e. g., process planning. This leads to an ineffective, costly and time consuming discovery, application and sharing of manufacturing knowledge (Economist Intelligence Unit, 2007). For example, production supervisors typically have to access different isolated IT systems and paper-based documents to find failure reports and improvement suggestions in order to manually correlate them with additional process information like metrics.

To address these issues, we present the Manufacturing Knowledge Repository (MKR), a universal holistic repository that consolidates structured and unstructured process knowledge to facilitate knowledge discovery, knowledge management and

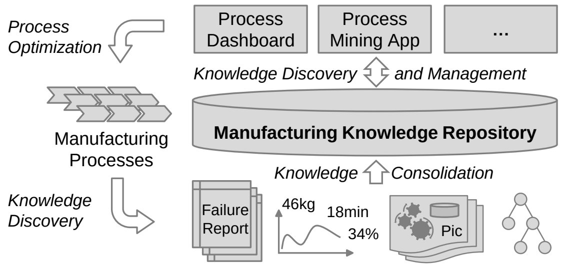

Figure 1: The Manufacturing Knowledge Repository.

knowledge-based process optimization in manufacturing (see Figure 1).

The remainder of this article is organized as follows: First, we structure related work with respect to process knowledge repositories in Section 2. Next, we define contents and requirements for the MKR in Section 3. This provides the basis for the conceptual repository data model presented in Section 4. In Section 5, we focus on implementation issues and present a prototypical implementation. A qualitative evaluation of the MKR and a technical proof of concept based on a case example are described in Section 6. Finally, we conclude in Section 7 and highlight future work.

2 RELATED WORK: PROCESS KNOWLEDGE REPOSITORIES

Process Knowledge repositories are databases for integrating, structuring and storing process knowledge (Davenport and Prusak, 2000). The latter comprises all types of insights related to processes. In this work, we focus on explicit knowledge in the sense of contextualized data connected by patterns and relations (Ackoff, 1989). We further subdivide explicit knowledge in structured knowledge having a predefined technical data structure, e. g., formal rules and metrics, and unstructured knowledge, e. g., photos and documents.

With respect to related work, we distinguish between manufacturing-specific repositories for process knowledge and concepts from the business process and workflow context. Manufacturing-specific approaches can be found as part of various expert systems for process planning (Kiritsis, 1995), (Giovannini *et al.*, 2012). They make use of formal rules and logics to support the generation of work plans. These kinds of repositories are typically based on structured knowledge and focus on process planning aspects. Besides process planning, there are only rudimentary repository approaches focusing on the other lifecycle phases, that is, process execution and process analysis. The tools presented in (Fischer *et al.*, 2000) share a common knowledge repository for process analysis in manufacturing. It integrates structured knowledge for rule-based, case-based and model-based reasoning to identify root causes of production failures. In (Mazumdar *et al.*, 2012), a manufacturing knowledge repository is presented. It integrates and annotates process-related documents, e. g., failure and performance reports, using manufacturing-specific ontologies to support semantic search capabilities for process execution and analysis. All these approaches make use of application-specific databases and are either exclusively based on structured or on unstructured knowledge.

Regarding process knowledge repositories in the business process context, process repositories with semantic search capabilities, e. g., (Ma *et al.*, 2007), can be seen as initial approaches. Most similar to the concept presented in this article is the work in (Niedermann *et al.*, 2011). The authors present a universal process knowledge repository that stores results of workflow analyses, especially metrics and data mining models. Yet, it focuses on structured knowledge and cannot simply be applied to manufacturing as it is based on workflow standards, especially the Business Process Execution Language.

The MKR goes significantly beyond existing approaches by integrating various types of structured and unstructured process knowledge in a universal database to support different analytics- and knowledge-driven applications across the entire process lifecycle in manufacturing.

3 REPOSITORY CONTENTS AND REQUIREMENTS

The MKR integrates different kinds of process knowledge, called insights, by associating them with corresponding process components. Hence, the two core building blocks of the MKR's content are a holistic process meta model as well as a catalogue of different types of insights. The main requirements for these building blocks are described in the following and are used as a basis for the definition of the data model in Section 4.

The holistic process meta model defines essential components of discrete manufacturing processes, e. g., process steps and resources, whereas it is *independent of a concrete industry* in order to be universally applicable. It has to integrate both *design-time and a run-time perspective*, that is, aspects of process planning and execution, to provide a holistic view. The design-time perspective comprises the process model defining, e. g., the types of resources needed, whereas the run-time perspective covers all aspects of the execution of the model, e. g., individual employed resources or occurred failures. Thereby, both a *process view* referring to the flow of process steps including the routing of materials as well a *resource view* referring to the detailed specification and dependencies of resources like machines have to be combined. Moreover, *changes of process models*, that is, their evolution over time, have to be traceable in order to support process optimization purposes (Niedermann *et al.*, 2011). It is important to remark, that there is no need for a highly detailed meta model like in computer aided planning systems. Instead, the meta model has to cover *all major components of manufacturing processes* to associate corresponding insights while remaining *easy to understand* for non-expert users in IT like production supervisors. Finally, it has to be able to be *implemented in a database environment* in order to use it for repository storage.

Regarding insights, a huge variety of knowledge-relevant objects exists in manufacturing ranging from work instructions over failure reports to key performance indicators. Thus, we analyzed insights across the entire process life cycle from a technical point of view differentiating structured and unstructured insights. We observed the following types of structured insights:

- *Metrics*, e. g., lead time, aggregating quantitative process attributes (Brown, 1996)
- *Data mining models*, e. g., decision trees or cluster models, representing patterns and relationships of process attributes (Han *et al.*, 2012)
- *Formal rules* in terms of if-then relations, which can be used for rule-based reasoning (Giarratano and Riley, 2005) or as business rules (Morgan, 2002)
- *Special process constructs*, e. g., rework sequences, which refer to sets of process steps with certain business semantics (Niedermann *et al.*, 2011)
- *Ontology concepts* in terms of semantic annotations using manufacturing-specific ontologies like MASON (Lemaignan *et al.*, 2006) to enable reasoning and semantic search capabilities

In addition, we identified the following types of unstructured insights:

- *Text* referring to any kind of unstructured textual data, e. g., emails or reports
- *Images* like photos, graphics or diagrams
- *Audio* comprising any kind of sound recordings
- *Videos*

4 CONCEPTUAL REPOSITORY DATA MODEL

The conceptual repository data model realizes the contents and requirements discussed in Section 3 and comprises a holistic process meta model as well as an insight model. In the following, we represent both parts as class diagrams in the Unified Modeling Language (UML) and describe their association.

4.1 Holistic Process Meta Model

The basis of the holistic process meta model is the basic meta model described in (Gröger *et al.*, 2012a). The latter comprises a manufacturing process meta model which takes a run-time perspective on manufacturing processes and is designed for the implementation in a data warehouse environment. We refine and extend this meta model with respect to design-time aspects in order to derive the holistic process meta model. To this end, we analyze existing process-oriented manufacturing meta models, especially (Erlach, 2011), (Zor *et al.*, 2011), (International Society of Automation, 2000), (Lemaignan *et al.*, 2006). Figure 2 shows the main components of the resulting process meta model, which we describe in the following. For the sake of simplicity, we omit many additional classes of the model, e. g., for spatial aspects of process steps, and do not detail attributes.

4.1.1 Design-Time Aspects

From a design-time point of view, that is, with respect to process planning and design, a *manufacturing process* in terms of a process model produces one or more types of *products*. A product can be described by *features* referring to informational aggregations of product characteristics, like geometric or functional aspects (Shah and Mäntylä, 1995). Features relevant for a certain process step are associated with the latter to enable both feature-oriented analysis across different manufacturing processes as well as the association of feature-oriented insights,

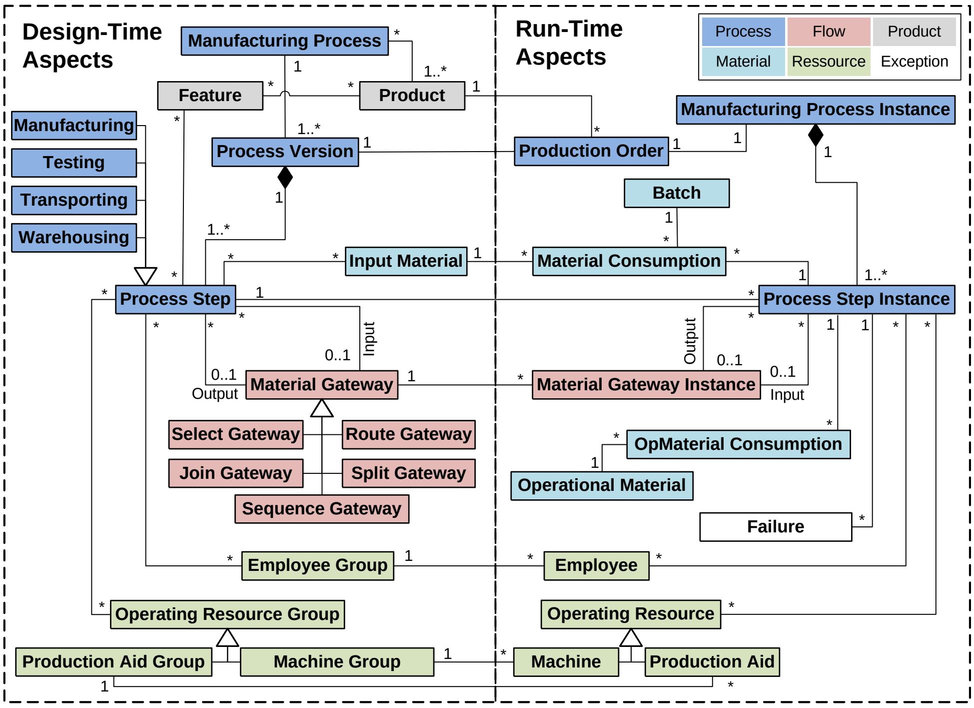

Figure 2: Main components of the holistic process meta model.

especially rules for knowledge-based process planning.

A manufacturing process comprises several *process steps*, that is, all steps necessary to produce the specified product. In order to analyze the evolution of a manufacturing process over time, different *process versions* can be defined, which comprise individual compositions of process steps. According to (Erlach, 2011), (Zor *et al.*, 2011), we differentiate *manufacturing* steps, comprising the actual manufacturing and assembly of parts, *testing* steps, which refer to quality control activities in a process, *transporting* steps, covering the movement of parts between different steps, and *warehousing* steps, referring to stock-keeping. Process steps are further associated with three types of resource groups, namely *operating resource groups* comprising *machine groups* and *productions aid groups*, as well as *employee groups*. These groups define requirements for the actual resources selected during process execution and control, e. g., specific machines, tools and workers. *Input material* refers to products and parts as external input of process steps, e. g., for assembly operations. It defines necessary material properties and amounts as described in the work plan.

Regarding the process flow, that is, the connection of process steps and the modeling of different paths, we exclusively focus on the flow of material as done in value stream design (Erlach, 2011). Thus, we omit additional control flow aspects for the sake of understandability. Moreover, we model the flow of material using material gateways and refine the concept in (Zor *et al.*, 2011) as follows: Two process steps are always connected by a *material gateway*. The first and the last step of a process have no input gateway or no output gateway respectively. Moreover, we differentiate five types of gateways: The *sequence gateway* defines a simple sequential passing of material from one process step to the other. The *route gateway* represents a diversion point in the material flow, i. e., one out of several possible subsequent process steps has to be chosen according to a defined condition. As a counterpart, the *select gateway* refers to a selection of one out of several preceding process steps. The *split gateway* creates parallel flows of material with a condition defining how the material is split up. The *join gateway* again joins parallel material flows.

4.1.2 Run-Time Aspects

The run-time perspective focuses on the execution of single instances of a manufacturing process which are initiated by a *production order*. The latter defines the customer as well as various order details like batch size. Instantiation refers to process execution and control and comprises the detailed planning of resources and materials. That is, individual *ma-*

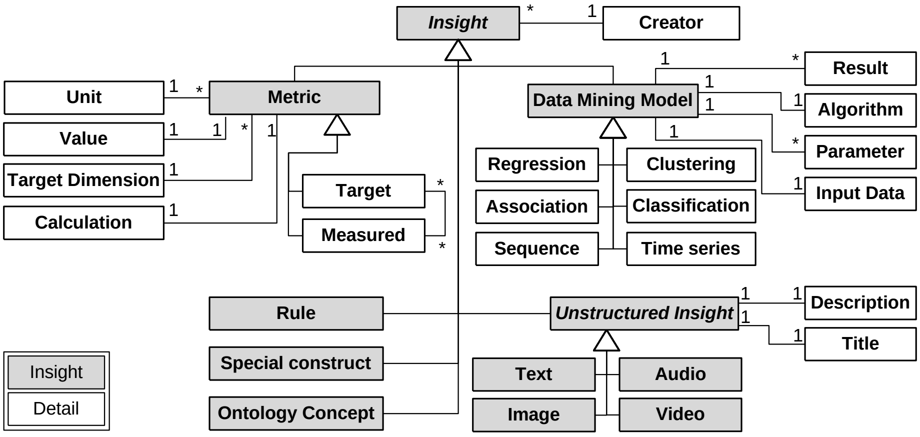

Figure 3: Main components of the insight model.

chines, *production aids* and *employees* are selected for process execution and are therefore associated to a *process step instance* which in turn belongs to a *manufacturing process instance*. Moreover, *material consumption* associates the actual *batch* of input material processed in a step instance.

In addition, there are elements which are not modeled at design-time, especially *failures*, which may occur during process execution, and the consumption of *operational material*. The latter refers to external input material which is consumed in a process step but does not become part of the product itself, e. g., oil or electricity.

4.2 Insight Model

Figure 3 shows the main components of the repository's insight model. In the following, we focus on metrics, data mining models and unstructured insights as major types of insights in manufacturing.

In general, an *insight* is associated with a *creator* referring to the employee who created the insight. This enables the integration of the MKR with existing yellow page systems for community-based knowledge management by linking the creator with its entry in the yellow page system.

Metrics primarily comprise the actual *value* and the *unit* of measurement, e. g., seconds or kilos. Moreover, they are organized in general *target dimensions* of manufacturing, especially time, quality, flexibility and cost (Kaushish, 2010). For example, lead time and adherence to delivery dates are metrics belonging to the time dimension. The *calculation* defines the formula as well as the meaning of the metric itself. Besides, we differentiate two types of metrics: *Target* metrics define values in terms of thresholds to be achieved during process execution, e. g., the maximum lead time of a process, whereas *measured* metrics comprise the actual recorded value. Thereby, measured metrics are associated with one or more target metrics with the latter defining, e. g., maximum or minimum values.

With respect to *data mining models*, we differentiate six major types, namely *regression* models, *classification* models, *association* models, *time series* and *sequences*. For a detailed description, we refer to (Han *et al.*, 2012). Each model is generated by a certain algorithm, e. g., a classification tree can be generated by the C4.5 algorithm, and algorithm-specific parameters, e. g., whether tree pruning is activated, are stored as well. Moreover, the *input data* that is used as a source for the algorithm is specified using a predicate filter which is evaluated over the repository's data. Further, the repository allows to store application *results* of data mining models, e. g., when a regression model is applied for predicting a metric. Yet, we assume this to be useful only in special cases, e. g., for compliance reasons. *Unstructured insights* have no predefined components and are thus generally descripted by a *title* and a textual *description*.

4.3 Insight Association

In general, insights can be associated with all components of the process meta model whereas one insight can be associated with multiple components and vice versa. In the following, we detail the association of metrics, data mining models and unstructured insights with respect to the major meta model components for a process-oriented browsing of insights, namely processes, process steps and resources (see Table 1). These associations are to be seen on a conceptual level independent of the im-

Table 1: Association of insights with process meta model components.

Insights / Meta Model Components	**Target Metrics**	**Measured Metrics**	**Data Mining Models**	**Un-structured Insights**
Design-Time Components	x	x	x	x
Run-Time Components	-	x	-	o

x / ○ / - Insights fully/partially/not associated

plementation, e. g., whether they may be enforced using application logic or database constraints.

With respect to metrics, *target metrics* are solely associated with design-time components like the version of the manufacturing process and operating resource groups as they define values to be achieved. *Measured metrics* are generated during process execution, e. g., the actual lead time of a process instance is measured. Thus, they are associated with corresponding run-time components. However, measured metrics may be aggregated over several run-time components representing values on the design-time level as well, e. g., the average lead time of a selected manufacturing process.

Data mining models describe patterns and relationships of a set of run-time elements, e. g., a clustering of process instances of a selected manufacturing process. Thus, they are solely associated with design-time elements.

Unstructured insights are generally associated with both design-time and run-time components like certain machines or entire machine groups. Yet, with respect to processes and process steps, unstructured insights are solely associated with the corresponding design-time components in order to reuse them across all process instances and step instances.

5 IMPLEMENTATION ISSUES

In this section, we analyze the characteristics of the data in the MKR and discuss different storage architectures. Moreover, we present a prototypical implementation of the MKR.

5.1 Data Characteristics

A storage-oriented analysis of the conceptual data model presented in Section 4 reveals several kinds of data that have to be stored. In the following, we characterize these different kinds of data as a basis for the development of a storage architecture for the MKR. Thereby, we focus on selected types of insights, namely, metrics, data mining models and unstructured insights, as major types of insights in manufacturing process management. Thus, there are four kinds of data to be stored:

- *Data concerning the manufacturing process and metrics*: This comprises data related to all components of the process meta model as well as on *corresponding* metrics. Thus, the data are well structured and can be very large in volume, especially with respect to process instance data. Moreover, they have to be efficiently accessed by analytical applications, in particular data mining tools, in order to generate data mining models.

- *Data concerning data mining models*: These data have to allow for a universal representation of data mining models as well as associated parameters in order to exchange them with external data mining tools for model evaluation and application.

- *Data concerning unstructured insights*: These data are semi-structured or unstructured and may comprise large volumes of multimedia data and text. The latter should be searchable whereas the former is primarily stored for manual exchange by the user.

- *Data concerning associations:* These data are structured and refer to the association of insights and components of the process meta model as outlined in Section 4.3. These data have to facilitate a flexible association, even if insights and meta model components are stored in different systems.

5.2 Storage Architectures

With respect to the above data characteristics, relational database technology constitutes the starting point of a storage architecture for the MKR to store *data on processes and metrics* in a multidimensional warehouse structure. This mature technology is suitable here because it handles huge amounts of structured data in a scalable and universally accessible way.

Regarding *data mining models*, there are two major references for their specification and exchange: The Predictive Model Markup Language (PMML) (Data Mining Group, 2013) is an XML-based format to specify data mining models in a semi-structured and vendor-independent way. Besides, the Common Warehouse Meta Model (CWM) (Poole *et al.*, 2003) and its data mining package define a general meta model for data mining models. Both approaches define the structure of the actual mining model, e. g., a decision tree, as well as parameters used to generate it, e. g., pruning settings. Yet, in contrast to the CWM data mining model, PMML is supported by a wide range of commercial and open source data mining tools and thus represents the first choice to store data mining models in a semi-structured format in the MKR.

Hence, semi- and unstructured data on unstructured insights and data mining models have to be stored and associated with structured data on processes and metrics in the MKR. As mentioned, relational database technology is suitable to store these data on processes and metrics. Taking this into account, there are two major storage architecture variants for the MKR:

- In the *relational-only architecture*, additional features of relational database management systems are exploited to store semi- and unstructured data together with structured data on processes and metrics in the relational database. That is, PMML files are stored as XML data and binary large objects (BLOB) and character large objects (CLOB) are used to store unstructured insights. Associations are directly realized as foreign key relationships between tuples representing insights and process components in the database. Moreover, full-text search capabilities of relational database management systems are employed to make use of text in unstructured insights like PDF documents.

- In the *extended architecture*, semi- and unstructured data are stored separately from the relational database in a Content Management System (CMS) (Kampffmeyer, 2007). A CMS allows for the central storage and management of content items, which are accessed by object identifiers. The latter are used to realize the association of insights and process components based on mapping tables. These tables combine primary keys of process components with object identifiers of corresponding insights.

For a comparison of these architecture variants, we refer to three major criteria: *the handling of insights*, the *realization of associations* between insights and process components as well as *maintenance* issues (see Table 2).

In view of the *handling of insights*, the extended architecture profits from advanced functions of a CMS. Apart from simple full-text search capabilities as in the relational-only architecture, a CMS typically provides text recognition functions as well as versioning concepts for content items. Moreover, it allows for a workflow-oriented handling of content items and thus eases the reuse and sharing of insights in workflow-based processes (Kampffmeyer, 2007).

Table 2: Comparison of architecture variants.

	Relational-only Architecture	**Extended Architecture**
Handling of Insights	-	+
Realization of Associations	+	-
Maintenance	+	-

Regarding the *realization of associations* between insights and process components, the relational-only architecture allows for a simple implementation using foreign key constraints. In contrast, the extended architecture requires additional efforts to ensure consistency of associations, e. g., to make sure that all affected associations are deleted if a corresponding content item in the CMS is removed.

With respect to *maintenance* issues, the relational-only architecture reduces maintenance efforts as existing database procedures, e. g., for backup and recovery, can be seamlessly applied to data on insights. In contrast, the extended architecture requires the maintenance of two separate storage systems.

To conclude, we opt for the relational-only architecture to implement the MKR as it eases the association of insights and process components and reduces maintenance efforts.

5.3 Prototypical Implementation

Our prototypical implementation is based on the work of (Vetlugin, 2012) and is carried out as part of our Advanced Manufacturing Analytics platform for the data-driven optimization of manufacturing processes. The platform comprises data mining use cases for continuous process improvement (Gröger *et al.*, 2012b) and makes use of a manufacturing-specific process warehouse, the Manufacturing Warehouse (Gröger *et al.*, 2012a).

The technical architecture of our prototype is based on the relational-only architecture discussed above and is shown in Figure 4. We implemented a simplified version of the MKR's conceptual data model as a relational schema in an IBM DB2 database. To this end, we extended the schema of the Manufacturing Warehouse with respect to the holistic process meta model and selected insights. The schema is oriented towards a relational snowflake schema to realize a multidimensional structure of the holistic process meta model with minimum redun-

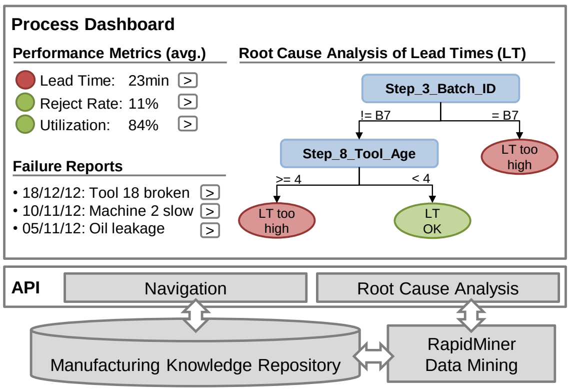

Figure 4: Technical architecture of the prototype.

dancy in the dimension tables. The schema comprises process step instances as central facts with metrics like lead time of a process step. Process details, e. g., the process the step belongs to, as well as employed resources in a step are treated as dimensions. This structure enables a multidimensional analysis of process execution data in the MKR, e. g., using Online Analytical Processing (OLAP).

Moreover, we defined an API implemented in Java which comprises two services: *Navigation* enables both browsing of the MKR's contents, e. g., to view insights associated with a certain process step, and uploading of new contents, e. g., photos. *Root cause analysis* is a data mining use case described in (Gröger *et al.*, 2012b) and focuses on the analysis of metric deviations using decision trees, e. g., for production supervisors identifying influence factors for high lead times as shown in Figure 4.

Decision trees are stored as insights in the MKR whereas RapidMiner is used as an open source data mining tool to derive the decision trees. These API services can be accessed by applications using the MKR.

6 EVALUATION

In this section, we provide a qualitative evaluation of the MKR as well as a technical proof of concept. We first evaluate the holistic process meta model and the insight model including MKR's support for analysis and insight generation. Next, we show that the MKR satisfies the whole range of process-oriented information needs and thus enables a holistic process management. Finally, for a technical proof of concept, we employ our prototype of the MKR in an exemplary case in the automotive industry.

6.1 Evaluation of the Holistic Process Meta Model

To evaluate the holistic process meta model (see Section 4.1), we analyze its universal applicability by showing that it covers all types of discrete manufacturing processes. In addition, we compare it with existing manufacturing meta models and show that it provides a sound basis for insight association and universal process representation in the MKR.

6.1.1 Evaluation of Universal Applicability

According to (Buzacott *et al.*, 2013), there are three general types of processes in discrete manufacturing, namely *mass manufacturing processes, series manufacturing processes* and *one-piece manufacturing processes* (see Figure 5). These types differ in their scale of production, their organization and their market orientation. In the following, we briefly describe these types and analyze how the holistic process meta model covers them.

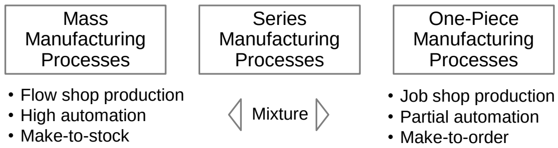

Figure 5: Types of discrete manufacturing processes.

Mass manufacturing processes focus on the production of large quantities of highly standardized products. The organization follows a flow shop production concept with a high degree of automation. That is, workers are mainly in charge of controlling machines which are sequentially connected by automated transportation steps. Production is decoupled from demand by a make-to-stock approach. To represent a flow shop production, the holistic process meta models allows for the modelling of a flow of *manufacturing steps* and *transportation steps* connected by *sequence gateways*. Moreover, various *operating resources* can be modeled to represent the high degree of automation. Besides, *warehousing steps* can be employed to represent make-to-stock aspects. Thus, the holistic process meta model supports the modelling of mass manufacturing process.

One-piece manufacturing processes focus on the production of single customized products. These processes are organized according to a job shop production layout with a high degree of flexibility and only partial automation. There is a significant amount of manual work whereas functionally similar workplaces are grouped together in the factory. Routing between these workplaces is complex as it may vary for each product. One-piece manufacturing follows the make-to-order principle with no significant stock keeping. The holistic process meta model allows for the representation of a job shop production layout by modelling various *material gateways* to represent flexible routings between *manufacturing steps* and *transportation steps*. Moreover, a *production order* models individual orders of customers whereas *features* allow to represent the customization of the ordered *product.* Hence, one-piece manufacturing processes can be represented by the holistic process meta model, as well.

Series manufacturing processes focus on the production of different but related products in predefined lot sizes and represent a hybrid form between one-piece manufacturing and mass manufacturing. These processes are based on a combination of flow shop production and job shop production depending on the lot size. Thereby, a middle to high degree of automation and temporary stock keeping are typical. As stated above, the holistic process meta model allows to represent both flow shop production and job shop production as well as a combination using additional *material gateways*. Besides, *production orders*, *products* and *features* can be modelled to represent series information, e. g., the number of goods and the product variant to be produced. Thus, series manufacturing is covered by the holistic meta model, too.

To sum up, the holistic process meta model supports the modelling of all general types of discrete manufacturing processes. Individual manufacturing processes in industry practice can be seen as derived or hybrid forms of these types (Buzacott *et al.*, 2013) and are thus supported by the meta model, as well. This confirms the universal applicability of the holistic process meta model and thus the *generality of the MKR for discrete manufacturing*.

6.1.2 Comparison of Meta Models

To evaluate the holistic process meta model with respect to existing process meta models in manufacturing, we did a qualitative comparison against the requirements defined in Section 3. For the comparison (see Table 3), we chose the ISA-95-1 process meta model (International Society of Automation, 2000) as it represents a common standard implemented in manufacturing execution systems. Moreover, we selected the Virtual Factory (VF) Data Model (Terkaj *et al.*, 2012) as it integrates a wide

Table 3: Comparison of process meta models.

	Holistic Process Meta Model	ISA-95-1 Model	Virtual Factory Data Model	Value Stream Design Model
Universal Applicability in Discrete Manufacturing	+	+	+	+
Integration of Design-Time and Run-Time Aspects	+	-	-	-
Combination of Process View and Ressource View	-	+	+	-
Support for Process Evolution	+	-	+	-
Model Simplicity for Insight Association	+	-	-	+

range of industrial and scientific manufacturing meta models. In addition, we chose the value stream design (VSD) model (Erlach, 2011) because value stream design is a typical method used to document manufacturing processes.

All these models are *universally applicable for discrete manufacturing* processes without focusing on specific branches or industries. Yet, only our holistic model *integrates design-time aspects and run-time aspects*, that is, information on process planning and process execution. We consider this an important point for a holistic knowledge management because the combined analysis of process execution information, e. g., resulting from machine data or occurred failures, as well as process planning information enables the generation of novel insights (Kemper *et al.*, 2013). For instance, a data-mining-based root cause analysis of metric deviations as described in Section 5.3 can reveal new knowledge for process improvement. Neither the ISA-95-1 model, nor the VF model nor the VSD model support the explicit modelling of process execution information.

Regarding the *combination of a process view and a resource view*, our meta model as well as the VSD model allow for the modelling of process flow aspects using gateways as well as basic resource information, e. g., on machines and production aids. However, detailed specifications of resources, e. g., regarding maintenance requirements, are not covered by these models. In contrast, the ISA-95-1 model as well as the VF model provide an additional resource view with details on all types of resources.

The tracking of the *process evolution* is fully supported by our meta model and the VF model. Both models include versioning concepts and keep track of the adaption of process models. This is important to support continuous knowledge-driven process improvement by reusing insights and evaluating their improvement impact over time. The VSD model does not focus on tracking process adaptions and the ISA-95-1 model only supports versioning of selected parts of the model without tracking changes of the entire process model over time.

The *model simplicity* refers to the number of elements and the structure of the model with respect to the comprehensibility for the end user. We consider model simplicity an important factor as it reduces the barriers for the collection and reuse of insights by the user which stimulates knowledge management. The VF model is comparatively complex as it integrates various meta models, e. g., on products, processes and resources, and comprises several abstraction layers. Similarly, the ISA-95-1 model comprises multiple generic definitions on processes and resources, e. g., abstract resource requirements are matched with actual resource capabilities. In contrast, the VSD model is designed for a simple modelling of manufacturing processes with a core list of process elements. Our holistic meta model refines and extends these elements without referring to generic views or definitions.

All in all, the qualitative comparison reveals that only the holistic process meta model fully supports the integration of design-time and run-time aspects and provides both model simplicity and support for process evolution. Although the resource view is only basically represented, a coarse-grained association of resource-related insights is possible with the holistic process meta model. Hence, it provides a *sound basis for insight association and universal process representation in the MKR*.

6.2 Evaluation of the Insight Model and MKR's Analysis Support

In the following, we evaluate the insight model (see Section 4.2) in combination with the MKR's analysis support and show that the MKR provides a comprehensive basis for the generation and storage of analysis results and insights of major data analytics systems (see Figure 6).

According to (Kemper *et al.*, 2010), there are four general types of data analytics for knowledge generation in business intelligence: *free data exploration*, *OLAP*, *reporting* and *model-based analytics*. The analysis results of these systems represent insights and thus have to be covered by the insight model to store them in the MKR. Moreover, the MKR as a whole should support the use of these data analytics for the generation of new insights.

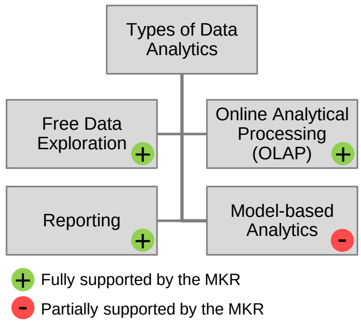

Figure 6: Types of data analytics and support by the MKR.

Free data exploration refers to the direct search and browsing of insights in the MKR by the user. There is no direct generation of new analysis results. Free data exploration rather provides the basis for further analytics by identifying needs for new analyses, e. g., failure reports which require further root cause analyses. The MKR fully supports free data exploration by navigation features (see Section 5.3).

OLAP comprises the multidimensional analysis of metric-oriented information (Pendse and Creeth, 1995). Metrics represent analysis facts and dimensions constitute views on these facts, e. g., analyzing the lead time of certain process steps. The MKR fully supports process-oriented OLAP analyses because (1) the insight model defines metrics and their relationships and (2) the MKR makes use of a multidimensional warehouse structure with these metrics as facts and elements of the process meta model as dimensions (see Section 5.3).

Reporting systems focus on the textual, graphical or diagram-oriented documentation of metric-information in reports. Reports constitute semi-structured or unstructured text documents and are thus covered as text insights in the insights model. Moreover, the multidimensional structure of the MKR with metrics as central facts supports the use of reporting systems, which are typically employed on multidimensional data warehouses.

Model-based analytics comprise data mining approaches (Han *et al.*, 2012) and expert systems (Giarratano and Riley, 2005). The former refer to the broad range of data mining techniques and models, e. g., clustering and classification. The latter mainly comprise case-based, model-based and rule-based approaches. With respect to the insight model, the six major types of data mining models are covered explicitly and further types may be added flexibly by inheritance. Formal rules are supported by the insight model, as well. In contrast, formal cases and formal models are not directly supported by the insight model due to their heterogeneity in different case-based und model-based applications. That is, they have to be represented as unstructured textual insights in order to incorporate them in the MKR. Considering the generation of data mining models, the MKR, with its multidimensional structure, fully supports the use of data mining tools like RapidMiner (see Section 5.3). Yet, the use of specific case-based, rule-based or model-based applications may require application-specific adaptions of the MKR's data structure due to missing standards for expert systems.

To sum up, the MKR supports both the generation and storage of analysis results from reporting and OLAP applications as well data mining systems and includes free data exploration. The key enablers are the insight model in combination with a multi-dimensional structure based on the process meta model which comprises process model data and process execution data for analysis purposes. With respect to expert systems, formal rules, models and cases can be stored in the MKR. Yet, the use of the MKR as a data basis for expert systems to generate new insights requires additional application-specific adaptions.

6.3 Evaluation of the MKR for Knowledge Management

On the basis of the above evaluation of the holistic process meta model and the insight model, we show that the realization of the models in the MKR enables a holistic process knowledge management by satisfying the whole range of process-oriented information needs in manufacturing.

In our previous work (Gröger *et al.*, 2013), we identified four general types of process-oriented information, namely *process context, process performance, process documentation* and *process communication*. In the following, we describe these information needs and analyze how the MKR satisfies them. Table 4 shows for each information need whether it is satisfied using insights or meta model components of the MKR.

Table 4: Information needs satisfied by the MKR.

	Process Meta Model	Insights
Process Context	X	
Process Performance	X	X
Process Documentation		X
Process Communication		X

- *Process context* refers to the structure and the status of the overall process and its underlying resources, e. g., machines, as well as the goods to be produced. The MKR's process meta model comprises all information relevant for the process context: Process steps and material gateways represent the structure and employees and operating resources provide information on process resources both from a design-time and a run-time point of view. Information on the product and its features is available, as well.

- *Process performance* alludes to information about the effectiveness and efficiency of the process and its resources. All information relevant for process performance is provided by insights, especially metrics and data mining models, as well as information about material consumption in the meta model.

- *Process documentation* refers to information to support the execution of a process, e. g., work instructions, as well as information for process improvement, especially improvement suggestions. Process documentation can be represented as special kinds of unstructured insights which may comprise text, audio or video supported by the MKR.

- *Process communication* covers information exchanged between employees participating in the process, especially text, video or audio messages. These can be treated as corresponding insights and are therefore supported by the MKR, too.

To conclude, the MKR satisfies the whole range of process-oriented information needs in manufacturing and thus enables a holistic knowledge management. The MKR consolidates knowledge across the entire process lifecycle and facilitates sharing amongst various target groups of users. Moreover, the MKR enables the cross-correlation of different types of knowledge like failure reports, metrics and data mining models to support the discovery of new insights for process improvement.

6.4 Case Example and Technical Proof of Concept

The technical proof of concept is based on the application of the prototype of the MKR (see Section 5.3) in an exemplary case in the automotive industry, that is, the mass production of steel springs for car motors as described in (Erlach, 2011). The manufacturing process consists of several sequential steps for winding, tempering and shot peening of springs and involves various machines like winding machines.

For our technical proof of concept, we modelled the manufacturing process according to the holistic process meta model whereas we used the typical model constructs to represent mass manufacturing processes as described in Section 6.1.1. On this basis, we identified attributes of resources and process steps, e. g., winding speed of winding machines, and generated corresponding process model and process execution data to populate the MKR with instance data. Thereby, we generated data on 100.000 executions of the manufacturing process and calculated metric values, e. g., for lead times and quality rates.

With respect to insights, we did several root cause analyses on lead times using process execution data in the MKR and deduced corresponding decision trees as data mining models which were stored in the MKR. Moreover, we stored exemplary machine manuals, photos and reports as JPEG and PDF files representing unstructured insights in the MKR.

Considering an application on top of the MKR, we implemented a knowledge-based *process dashboard* on an Android tablet pc addressing both workers on the factory shop floor and production supervisors (see Figure 4). The dashboard is based on our requirements analysis described in (Gröger *et al.*, 2013) and represents an application using the MKR and its API to provide mobile access to different kinds of process knowledge in different application scenarios. For instance, workers can get information on best practices and work instructions as well as upload photos and reports of manufacturing failures. Besides, production supervisors can correlate metrics and failure reports and execute root cause analyses.

Based on our test system (Windows Server 2008 R2, Core i7-2620M@2,7 GHz, 8 GB RAM) and data on 100.000 process instances, the MKR proved to provide acceptable system performance for interactive usage in typical application scenarios of the dashboard described in (Gröger *et al.*, 2013).

This technical proof of concept demonstrates the fundamental feasibility and applicability of the MKR combined with suitable applications like the dashboard. The MKR proved to provide the facilities for insight generation, storage and reuse based on data of a realistic manufacturing process.

7 CONCLUSION AND FUTURE WORK

In this article, we introduced the Manufacturing Knowledge Repository, a holistic repository facilitating process knowledge management in manufacturing. It consolidates structured and unstructured knowledge, e. g., metrics, data mining models and text documents, and can be used by various applications. We presented the conceptual data model including a holistic process meta model and an insight model and discussed different storage architectures. We did a qualitative evaluation of the data models and presented a technical proof of concept based on a prototypical implementation in a case example.

With respect to future work, our goal is to implement an alternative storage architecture for the MKR and to investigate novel analytics on top of the MKR. That is, on the one hand, we are going to implement the extended architecture introduced in this article. This architecture seams promising to us as it exploits the functionality of a Content Management System for the workflow-based reuse and distribution of insights in business processes. On the other hand, we are going to examine novel analytics which combine structured and unstructured knowledge to generate new insights, e. g., combining data mining on process execution data and text analytics on unstructured text documents.

ACKNOWLEDGEMENTS

The authors would like to thank the German Research Foundation (DFG) for financial support of this project as part of the Graduate School of Excellence advanced Manufacturing Engineering (GSaME) at the University of Stuttgart.

REFERENCES

Ackoff, R.L. (1989), “From Data to Wisdom”, *Journal of Applied Systems Analysis*, Vol. 16, pp. 3–9.

Brown, M.G. (1996), *Keeping score. Using the right metrics to drive world-class performance*, American Management Association, New York.

Buzacott, J.A., Corsten, H., Gössinger, R. and Schneider, H.M. (2013), *Production planning and control. Basics and concepts*, Oldenbourg, München.

Chapman, C.B. and Pinfold, M. (2001), “The application of a knowledge based engineering approach to the rapid design and analysis of an automotive structure”, *Advances in Engineering Software*, Vol. 32 No. 12, pp. 903–912.

Data Mining Group (2013), "PMML 4.1 Standard", http://www.dmg.org/v4-1/GeneralStructure.html (accessed 06.02.13).

Davenport, T.H. and Prusak, L. (2000), *Working knowledge. How organizations manage what they know*, Harvard Business School Press, Boston.

Economist Intelligence Unit (2007), *Knowledge management in manufacturing*, London.

Erlach, K. (2011), *Value stream design. The way to lean factory*, Springer, Berlin.

Fischer, U., Stokic, D. and Beckmann, T. (2000), "Management of Corporate Knowledge for Process Improvement in Manufacturing Companies", in Stanford-Smith, B. and Kidd, P.T. (Eds.), *E-business: Key Issues, Applications and Technologies*, IOS Press, Amsterdam, pp. 721–727.

Giarratano, J.C. and Riley, G. (2005), *Expert systems. Principles and programming*, 4th ed., Thomson Course Technology, Boston.

Giovannini, A., Aubry, A., Panetto, H., Dassisti, M. and Haouzi, E.H. (2012), "Knowledge-Based System for Manufacturing Sustainability", in *Proc. of the IFAC Symposium on Information Control Problems in Manufacturing (INCOM) 2012*, Elsevier, pp. 1333–1338.

Goossenaerts, J., Arai, E., Shirase, K., Mills, J.J. and Kimura, F. (2005), "Enhancing Knowledge and Skill Chains in Manufacturing and Engineering", in Arai, E., Goossenaerts, J., Kimura, F. and Shirase, K. (Eds.), *Knowledge and skill chains in engineering and manufacturing*, Springer, London, pp. 1–10.

Gröger, C., Schlaudraff, J., Niedermann, F. and Mitschang, B. (2012a), "Warehousing Manufacturing Data. A Holistic Process Warehouse for Advanced Manufacturing Analytics", in *Proc. of the Int. Conference on Data Warehousing and Knowledge Discovery (DaWaK) 2012*, Springer, Berlin, pp. 142–155.

Gröger, C., Niedermann, F. and Mitschang, B. (2012b), "Data Mining-driven Manufacturing Process Optimization", in *Proc. of the World Congress on Engineering (WCE) 2012*, Newswood, Hong Kong, pp. 1475–1481.

Gröger, C., Hillmann, H., Hahn, F., Mitschang, B. and Westkämper, E. (2013), "The Operational Process Dashboard for Manufacturing", in *Proc. of the CIRP Conference on Manufacturing Systems (CMS) 2013*, Elsevier, pp. 205–210.

Han, J., Kamber, M. and Pei, J. (2012), *Data Mining. Concepts and Techniques*, 3rd ed., Morgan Kaufmann, Waltham.

International Society of Automation (ISA) (2000), *Enterprise-Control System Integration. Part 1: Models and Terminology, Standard ISA-95-1*.

Kampffmeyer, U. (2007), *ECM - Enterprise Content Management*, Project Consult, Hamburg.

Kaushish, J.P. (2010), *Manufacturing processes*, 2nd ed., Prentice Hall India, New Delhi.

Kemper, H.-G., Baars, H. and Lasi, H. (2013), "An Integrated Business Intelligence Framework. Closing the Gap Between IT Support for Management and for Production", in Rausch, P., Sheta, A.F. and Ayesh, A. (Eds.), *Business Intelligence and Performance Management*, Springer, London, pp. 13–26.

Kemper, H.-G., Baars, H. and Mehanna, W. (2010), *Business Intelligence*, 3rd ed., Vieweg, Wiesbaden.

Kiritsis, D. (1995), "A review of knowledge-based expert systems for process planning. Methods and problems", *International Journal of Advanced Manufacturing Technology*, Vol. 10 No. 4, pp. 240–262.

Lemaignan, S., Siadat, A., Dantan, J.-Y. and Semenenko, A. (2006), "MASON: A Proposal For An Ontology Of Manufacturing Domain", in *Proc. of the IEEE Workshop on Distributed Intelligent Systems (DIS) 2006*, IEEE, Los Alamitos, pp. 195–200.

Ma, Z., Wetzstein, B., Anicic, D., Heymans, S. and Leymann, F. (2007), "Semantic Business Process Repository", in *Proc. of the Workshop on Semantic Business Process and Product Lifecycle Management (SBPM) 2007*, CEUR Proceedings, Innsbruck, pp. 92–100.

Mazumdar, S., Varga, A., Lanfranchi, V., Petrelli, D. and Ciravegna, F. (2012), "A Knowledge Dashboard for Manufacturing Industries", in *Proc. of the European Semantic Web Conference Worksops (ESWC) 2011*, Springer, Berlin, pp. 112–124.

Monauni, M. and Foschiani, S. (2013), "Agility in Production Networks - Classification, Design and Configuration", in *Proc. of. the Int. Conference on Production Research (ICPR) 2013*.

Morgan, T. (2002), *Business rules and information systems. Aligning IT with business goals*, Addison-Wesley, Boston.

Niedermann, F., Schwarz, H. and Mitschang, B. (2011), "Managing Insights - A Repository for Process Analytics, Optimization and Decision Support", in *Proc. of the Int. Conference on Knowledge Management and Information Sharing (KMIS) 2011*, SciTePress, Paris, pp. 424–429.

Pendse, N. and Creeth, R. (1995), *The OLAP Report. Succeeding with on-line analytical processing. Volume 1*, Business Intelligence.

Poole, J., Chang, D., Tolbert, D. and Mellor, D. (2003), *Common warehouse metamodel*. Wiley, New York.

Shah, J.J. and Mäntylä, M. (1995), *Parametric and feature based CAD/CAM. Concepts, techniques, applications*, John Wiley & Sons, New York.

Terkaj, W., Pedrielli, G. and Sacco, M. (2012), "Virtual Factory Data Model", in *Proc. of the Int. Workshop on Ontology and Semantic Web for Manufacturing (OSEMA) 2012*, CEUR, Aachen, pp. 29–43.

Vetlugin, A. (2012), *A process insight repository supporting process optimization*, Master Thesis, University of Stuttgart.

Zor, S., Leymann, F. and Schumm, D. (2011), "A Proposal of BPMN Extensions for the Manufacturing Domain", in *Proc. of the CIRP Conference on Manufacturing Systems (CMS) 2011*, pp. 1–6.