

A Knowledge Perspective on Big Data by Joining Enterprise Modeling and Data Analyses

Hans-Georg Fill
Research Group Knowledge Engineering
University of Vienna
Vienna, Austria
Email: hans-georg.fill@univie.ac.at

Florian Johannsen
Chair of Business Engineering
University of Regensburg
Regensburg, Germany
Email: florian.johannsen@wiwi.uni-regensburg.de

Abstract—In this paper we discuss how knowledge management can contribute to the analysis of big data by joining enterprise modeling methods with data analyses. The goal of this approach is to enable the seamless interaction and exchange of information between knowledge-oriented representations as provided by enterprise modeling on the one hand and methods for analyzing data on the other hand. For the realization of the approach we revert to techniques of metamodeling. These permit to describe the necessary extensions of enterprise modeling methods and implement them as IT-based tools using metamodeling platforms. For evaluating the feasibility of our approach we describe a generic implementation using the ADOxx metamodeling platform and the R toolkit. In addition, we discuss the application to a use case from the area of business process improvement and the according implementation within the ADOxx-based RUPERT tool.

Keywords—enterprise modeling, knowledge management, data analysis, metamodeling

I. INTRODUCTION

With the upcoming of the phenomenon of big data, businesses are today confronted with the challenge of gaining insights into the large amounts of data generated by their own and external IT systems. This includes not only data from traditional ERP systems but with increasing intensity also data from social media applications such as Facebook, Twitter, or YouTube [1], [2], as well as from services and sensors in end-user products [3], [4]. The insights gained from this data then need to be considered for managerial decisions [5].

On the technical side, the difficulty lies in the particular features of big data which make it difficult to handle. This includes not only the volume, i.e. the sheer amount of data at the scale of millions, billions or even trillions of records, but also the high velocity of the generation of data, its variety in terms of different types and data sources that need to be aligned, its lack of veracity and structure, as well as the difficulty of extracting it from different systems and technical platforms [6], [7].

Although the area of data mining has developed a number of sophisticated techniques for analyzing big data [8], the

major challenge lies in the alignment of insights gained from these analyses with business practices and management decisions [9], [5]. In particular, this task is highly demanding from a cognitive side due to the multi-dimensional nature of the solution space. In order to take well-founded decisions, decision makers not only need to take into account the results of the data analyses but also the according domain expertise [10], [5], [8]. At the same time they must consider further knowledge aspects such as the strategic orientation of a company, the enterprise culture, the legal environment, the structure of its business processes, the intellectual capital and capabilities of the organization, as well as the skills and motivation of its employees to implement the decisions derived from the data analyses [11], [12], [5], [10].

In knowledge management, techniques of enterprise modeling serve as an approach for reducing complexity and enabling users in dealing with sophisticated interrelationships between knowledge and technical systems [13], [14]. Through transforming the requirements on a business level to formalized and technology-oriented views [15], enterprise modeling supports the representation, communication, and analysis of knowledge in organizations [16], [17]. This concerns for example knowledge about business models, the legal environment, business processes, or strategic goals, which resides in the heads of an organization's staff. In technology-oriented knowledge management, enterprise modeling is realized in the form of according IT-based tools [14], [18], [19]. This permits to establish interfaces to IT systems for processing the model information, e.g. using simulations [20]. However, potential synergies between enterprise modeling and technology-oriented knowledge management often remain unexploited as corresponding initiatives run separately within companies [21]. By reverting to the techniques of metamodeling [22], structured approaches are available for designing conceptual modeling languages, expressing them using formal specifications, and realizing them as IT-based implementations [23], [24]. In addition, metamodeling also permits to transfer the results of these elaborations into technical implementations using metamodeling platforms [20], [25].

For addressing the challenge of aligning insights gained from big data with domain knowledge, we thus propose an approach for joining enterprise modeling and data analyses. The goal is to add a knowledge perspective to data analyses through a conceptual and technical integration. Thereby, we follow the standard principles of engineering information systems in terms of analysis, design, evaluation, and diffusion [26]. In the following, we will therefore briefly analyze foundations for describing the approach in section II. Subsequently, in section III we will present the design of a framework for discussing the possible options for an alignment of enterprise modeling and data analyses based on metamodeling. Thereafter, the framework is evaluated in section IV. The evaluation is conducted on the one hand through an implementation on a metamodeling platform and on the other hand by applying the approach to a use case in the area of business process improvement. Finally, related work regarding our approach will be discussed in section V. The paper concludes with an outlook on further research on this topic as well as limitations.

II. FOUNDATIONS

For clarifying the terms and concepts used in the domains of enterprise modeling and metamodeling, we will briefly describe the components of modeling methods, the constituents of metamodels, and their IT-based realization.

According to a widely-used framework proposed by Karagiannis and Kühn [27], [20], modeling methods are composed of a modeling technique and mechanisms and algorithms. The modeling technique comprises the modeling language with its syntax, semantics, and notation, as well as a modeling procedure for specifying how to apply the modeling language to generate results [28], [29], [30]. The mechanisms and algorithms are further divided into generic, specific, and hybrid types. Generic mechanisms and algorithms are applicable to any modeling language, whereas specific mechanisms and algorithms only to particular modeling languages instead. Such of the hybrid type can be configured for several modeling languages.

When further defining the components of a modeling language and in particular the syntax, it has to be distinguished between the abstract syntax and the concrete syntax. The abstract syntax can be regarded as a kind of template or stencil that specifies the elements of a modeling language and how they can be combined to derive valid statements in that language. The abstract syntax is often also denoted as the *metamodel* of a modeling language [29]. The concrete syntax is used to define a concrete model instance. This can be done using graphical or textual notations or also both at the same time. In graphical modeling tools, the concrete syntax is typically realized in the form of a visual language and a corresponding data format such as a relational database or as XML-files [20].

For specifying the abstract syntax it has to be reverted to some kind of metalanguage. This can either be one of the standard computer science languages developed for such purposes, e.g. EBNF, or specialized languages that specifically target the domain of metamodeling. The abstract syntax of these metalanguages is then denoted as a *meta-metamodel*. Meta-metamodels are often tied to so-called metamodeling platforms [27]. The purpose of these software platforms is to interpret and execute the definitions of the modeling languages expressed in the metalanguages. In particular they cover aspects such as the automatic generation of graphical model editors, the provision of persistency mechanisms, or analysis and simulation functionalities. Apart from proprietary approaches for such platforms [25], more recently also an implementation-independent metalanguage has been proposed by Visic et al. [22]. Apart from the definition of modeling languages for different metamodeling platforms, this metalanguage also permits to describe additional functionalities of modeling methods such as algorithms or modeling procedures.

III. FRAMEWORK FOR ALIGNING ENTERPRISE MODELING AND DATA ANALYSES

The alignment of knowledge-oriented views on an enterprise with insights gained from data analyses can be conducted from various directions. The framework we present in the following takes a generic perspective by using metamodels as a common basis. Thereby, guiding principles are established that can then be instantiated for several approaches. Before describing the framework, we introduce two fundamental assumptions. The first assumption targets the constituents of metamodels and the second assumption the access to explicit human knowledge in models and data via metamodeling.

A. Assumptions

For the purpose of this paper we regard a simplified version of metamodels. In particular, we assume that metamodels consist only of model types, classes and relations, and attributes. Thereby, model types contain a number of classes and relations with classes and relations comprising a number of attributes. Model types can be instantiated to model instances, i.e. when a concrete diagram for representing a model is created. Classes represent structural entities that can be instantiated in a model instance. Relations are used to connect instances of classes or instances of classes and instances of model types. We further assume that relations may also span across different model instances. Attributes assign values to classes and relations. They may be of types such as integer, string, or float or of specialized types, e.g. for specifying the graphical representation of classes and relations. This view corresponds to the formalization given in [23], [31] and suffices for describing our approach.

The second assumption that we make for our framework concerns the uniform access to knowledge and data. In the field of modeling the part of the 'real-world' that is being represented in models is also denoted as the 'system under study' [32]. In our case, the systems that we target are *knowledge* in the form of the relevant domain expertise held by domain experts as well as *data* that need to be analyzed for providing additional support for decision making.

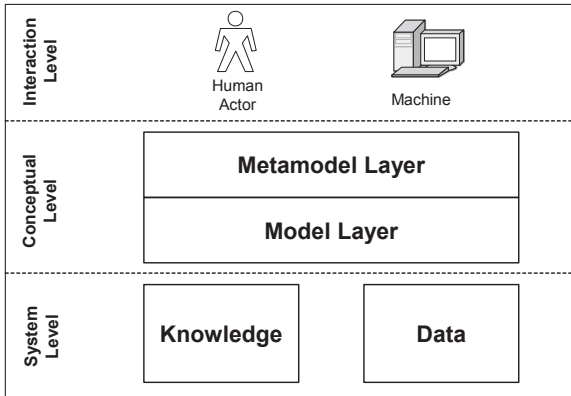


Figure 1. Uniform Access to Knowledge and Data

For both systems we assume that they can be accessed on a conceptual level via models that are based on meta-models. In the case of knowledge this may either be explicit knowledge that can be encoded using a modeling language. Alternatively, models may also represent information *about* knowledge, for example by using approaches such as knowledge source, asset, or application maps that define where knowledge resides in an organization and how it is structured and used [33]. Similarly, in the case of data, not necessarily all data and the according methods for processing need to be completely transferred into models. Rather, the models enable the access to the data, e.g. by providing references to analysis methods and according meta-data. On the level of interaction both the knowledge and the data can thus be accessed by human actors and machines in the same way – see also figure 1.

B. Metamodel-based Alignment

Based on these assumptions, the actual alignment of the access to knowledge and data can then be conducted by reverting to the level of metamodels. In particular we regard how the *representation and analysis* of knowledge and data can be conducted by using the constituents of metamodels. Hereby, we identified three types of granularity that can be applied to representation and analyses as shown in figure 2.

Considering the case of knowledge representation and analysis at first, the types of granularity follow common practices used in the area of formal metamodeling [24], [20], [23]. Therefore, considering knowledge representation

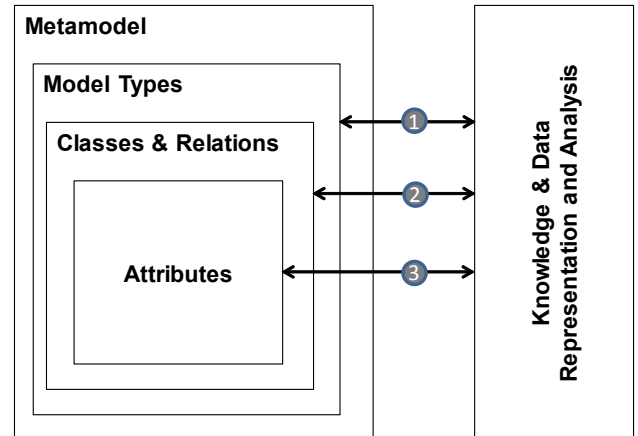


Figure 2. Types of Granularity for Aligning Knowledge & Data Representations and Analyses: (1) Model Types, (2) Classes and Relations, (3) Attributes

and analysis on the level of model types corresponds to identifying knowledge domains and knowledge management methods that can be treated separately. For example, such a knowledge domain may be the information and control flow aspects of a business process that can be represented through a model type corresponding to the BPMN standard¹. Or, the knowledge management method can be set to the identification of knowledge sources in an organization, which could be represented through a model type for knowledge source maps according to Eppler [33].

On the level of classes and relations, structural and behavioral aspects of knowledge can be represented and analyzed. Thereby, it can be chosen whether only structural or behavioral aspects shall be included or if they are combined. At the same time also the aspects of user/machine interaction need to be considered for this type of granularity.

For example, it may be beneficial in some cases to represent certain knowledge aspects just by one class that can be instantiated multiple times, e.g. in the case of process maps. Whereas in other cases, knowledge aspects need to be represented in greater detail to permit fine-grained interaction, e.g. when a user needs to depict the flow of information between activities in a business process. Regarding the analysis, formal, semi-formal, and informal analysis methods can be represented using classes and relations. A good example for this has been presented by Leutgeb et al. where informal, semi-formal, and formal rules for processing knowledge are specified [34], [35].

The most detailed type of granularity comes in the form of attributes. Here it needs to be chosen how quantitative and textual information that belongs to classes and relations is included in the models. As has already been mentioned above, this also concerns the graphical representation of classes and

¹See <http://www.omg.org/spec/BPMN/2.0/> last accessed 05-06-2015

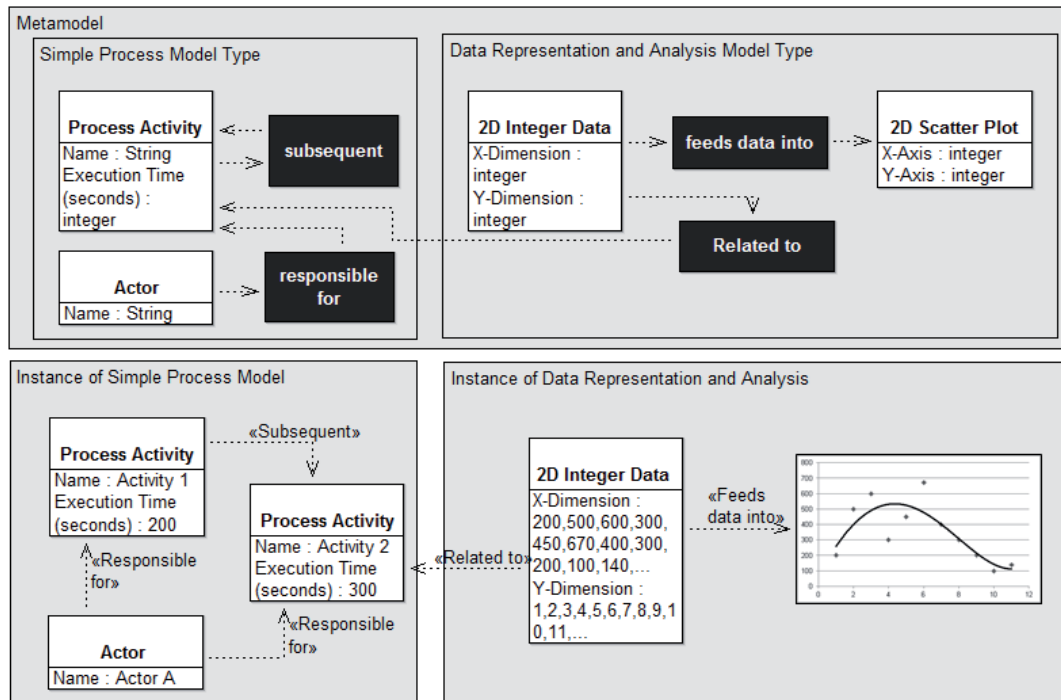


Figure 3. Sample Instance of the Framework Showing: a Simple Process Model Type and a Data Representation and Analysis Model Type on the Metamodel Level; an Instance of a Simple Process Model and an Instance of Data Representation and Analysis. Relations are Shown as Dark Rectangles, Classes are Shown with Their Assigned Attributes, Instance Relationships between the Metamodel and the Model Instances have been omitted.

relations, which is defined using graphical attributes [20]. Thereby, also dynamic aspects can be considered, e.g. by using dynamic visualizations for depicting the current state of attributes [30], [36].

For the data part similar considerations can be made using the constituents of metamodels. Based on the motivation as described in the introduction, the focus is put on the aspect of data representation and analyses here. In essence, there are two directions from which it can be chosen. The first direction is to use separate model types for representing data and the according methods for their analysis. The second direction is to combine both aspects in joint model types. Whereas the latter option eases the alignment of the representation of data and the analyses working on them, the first option provides greater flexibility in terms of decoupling the analysis part from the representation part.

Concerning classes and relations, data and their analyses can be represented in multiple ways in models. For the representation of data it can be reverted to well-known and largely established modeling approaches such as the entity-relationship approach [37] or the adaptable database design methodology [38] as well object-oriented approaches such as proposed by Trujillo et al. [39]. On the side of analyses it can be reverted to data analysis models such as used in visual database querying languages [40] or specialized approaches for the visual querying of big data [41].

By reverting to the attributes, data representations and

analyses can be further detailed by additional information. Examples would be meta information such as the ownership and provenance of certain data entities, the validity of data entities in terms of given time periods, or technical information about the access to data entities and their processing. Again, also the visual representation of the results of data analyses can be specified here.

C. Sample Instantiation of the Framework

To make the above characterized framework more graspable, we illustrate its application in the following with a sample instantiation. This will be followed by the evaluation of the framework in the form of an implementation and the application to a use case.

The sample shown in figure 3 contains a highly simplified metamodel with two model types. The first model type provides elements for representing knowledge about activities and responsible actors in business processes. The second model type provides elements for representing 2D integer data and a relation for feeding this data into scatter plots for conducting visual analyses.

Below the metamodel two model instances are depicted: first, an instance of a simple process model and second an instance for a data representation and analysis model. Whereas the process model refers to explicit domain knowledge on business processes, the data representation and analysis model complements this by adding the data perspective.

Thus, the model on the right can be used to represent data on process activities, which stems for example from ERP systems. By using the relation 'Related to', which is defined on the metamodel level, the reference to the process data can be linked to a specific process activity. By assigning the 2D data to an instance of the scatter plot class with the relation 'feeds data into', a visual analysis of the underlying data becomes possible. At the same time, the data analysis is complemented with additional information from the side of domain knowledge, e.g. by investigating the actor responsible for that particular process activity.

Although the sample is highly simplified, it clearly shows the central idea of the framework for aligning knowledge and data. When complementing for example the classes and relations in the data representation and analysis model type with other entities, e.g. for representing higher dimensional data or semi-structured data, more complex data structures can be represented. Similarly, on the analysis side, also more sophisticated methods for big data analyses can be represented, e.g. as discussed in information visualization or data mining [8], [30]. The same applies also to the knowledge side. It can be easily imagined to replace the simple process model with a full-fledged metamodel for business processes, e.g. as defined by the BPMN standard.

As already mentioned above from a technical point of view, the data does not necessarily need to be added in the model instances. Rather, the classes used for representing the data may contain references to the storage systems used for big data, e.g. by referring to Hadoop's distributed file system (HDFS)². This also applies to the part of analyses where the actual algorithms for processing the data may not be represented on the level of models but rather by referencing some external implementation, e.g. by using the Hadoop MapReduce framework³ or specialized tools for the statistical analysis of big data [42]. This will be shown in the subsequent evaluation section.

IV. EVALUATION

For the evaluation of our approach we revert to the guidelines set up in the fields of design-oriented information systems research and engineering [43], [44], [26]. In particular we will do this in two ways. At first, we will describe how the framework described above can be technically implemented in order to demonstrate the feasibility of the approach. Second, we will make use of a descriptive evaluation method by constructing a detailed scenario based on a use case to demonstrate the utility of the approach [44], [26].

²See https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html last accessed 08-06-2015

³See https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html last accessed 08-06-2015

A. Implementation

A first implementation of the framework for aligning enterprise modeling and data analyses has been conducted by using the freely available ADOxx metamodeling platform together with the open-source statistical analysis tool R [20], [45]. Metamodeling platforms permit to easily implement modeling methods by providing functionalities for the specification of modeling languages and mechanisms and algorithms [27]. Thereby, little or no programming effort is necessary to specify the syntax and graphical notation of a modeling language. Similarly, for the realization of algorithms such platforms often provide domain specific languages (DSL) with specialized constructs for interacting with the contents of models and for defining user interfaces. The chosen ADOxx metamodeling platform not only provides DSLs for the specification of modeling languages, their graphical representation, as well as for model-based algorithms. It also comes with advanced persistency features for automatically storing models and data in a relational database. In addition, ADOxx provides a number of interfaces for interacting with external tools and services including an XML and a SOAP interface as well as built-in functionalities for invoking external applications.

Statistical analysis tools are today available as commercial software tools or as open source platforms. They are specifically targeted towards the efficient processing of data for applying statistical analyses. This includes not only the calculation of different types of statistical measures and coefficients but also the generation of a variety of statistical diagrams and visualizations to inspect the data. The chosen R tool is available as open source and provides a DSL for interacting with data, for specifying analyses, and for creating visualizations [45]. A major advantage of R are the large number of available packages for adding functionality to the base platform. In particular, several packages exist for enabling the processing of big data with R [46], [42]. A recent example is RHadoop, which contains five packages for linking R with Apache Hadoop and enabling functionalities in R such as Hadoop MapReduce tasks and for interacting with the Hadoop distributed file system (HDFS) or the distributed database HBase⁴. Apart from the processing of large volumes of data, there exist also R-packages for text mining, e.g. tm⁵, machine learning, e.g. RWeka⁶, or packages offering analyses of high-dimensional data using information visualizations, e.g. parallel coordinates plots with the MASS package⁷. Thus, R seems well suited not only to conduct a large variety of data analyses but also for

⁴See <https://github.com/RevolutionAnalytics/RHadoop/wiki> last accessed 05-06-2015

⁵<http://tm.r-forge.r-project.org/> last accessed 05-06-2015

⁶<http://cran.r-project.org/web/packages/RWeka/index.html> last accessed 05-06-2015

⁷<http://cran.r-project.org/web/packages/MASS/index.html> last accessed 05-06-2015

scaling analysis approaches in order to deal with big data.

The architecture for the implementation using ADOxx and R is shown in figure 4. Thereby, only a subset of the components of the ADOxx architecture are presented – for a complete overview we refer to [20]. ADOxx is a Windows-based application that accesses a relational database. The modeling subsystem (CORE) provides an abstraction layer between the database and the application components. For the coupling we used the External Coupling component of ADOxx that allows to write algorithms in the domain specific ADOscript language. ADOscript provides APIs for interacting with external applications. This functionality was used to create R scripts based on information entered by users in the models as well as files in CSV format, which contained the data stored in the models. In this way, the metamodel shown in the example in figure 3 can be used to create 2D integer data residing in the CSV files as well as R configuration scripts originating from the instantiation of the scatterplot class.

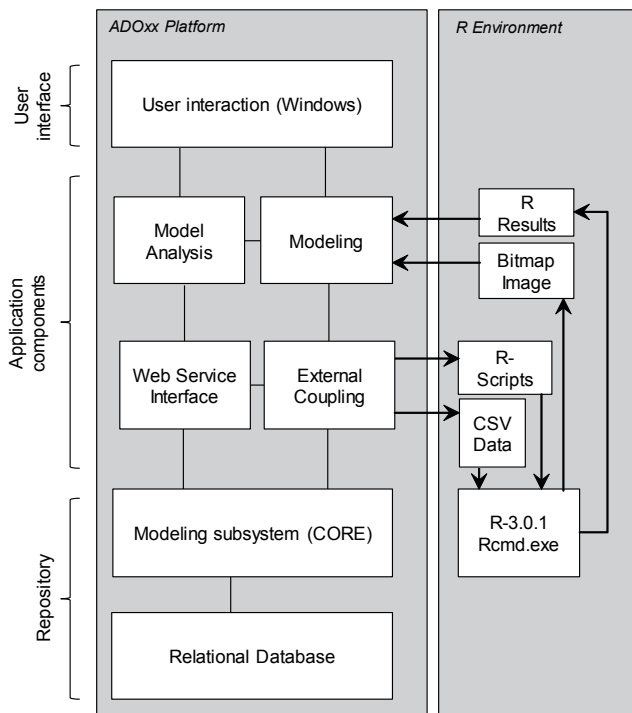


Figure 4. Architecture for Coupling ADOxx and R

With this information, the R environment was invoked programmatically. The result of the calculations in R was handed back to the modeling component and shown in the corresponding model instances. Additionally, bitmap images are generated with R, e.g. to create the actual scatterplot representation. For the example at hand, these bitmap images were complemented on the side of R with further statistical results such as Pearson and Spearman correlation coefficients and trendlines. The evaluation of the

approach for assessing its technical feasibility could thus be positively completed. Possible extensions include the use of specific big data packages for R as well as an underlying Hadoop infrastructure. As these extensions can however be realized based on the R platform used in our architecture, the interaction with the ADOxx platform would not have to be changed.

B. Use Case: Business Process Improvement

For a further evaluation, it will be discussed how the approach can be applied to a use case in the field of business process improvement (BPI). The approach was used to complement the ADOxx-based RUPERT modeling toolkit⁸ which implements a roadmap for conducting BPI projects [47], [48], [49]. In the following we describe the application of the prototype for a BPI project previously conducted at an automotive bank. The project focused on the improvement of the "end-of-terms" (EOT) process, which is of central importance for the company. Several parties cooperate within the process. The process is triggered each time a customer's leasing contract for a car ends. In that case, the customer returns the car to a predefined car dealer. The car is then inspected regarding any damages before its terminal value is determined. A car return protocol (CRP) is established by the car dealer who sends it to the automotive bank in a subsequent step. Based on this information the final customer bill is calculated by the automotive bank. The process ends with the customer meeting the bill.

In the preceding year, more than 45.000 ending leasing contracts were processed by the automotive bank. About 18% of all cases were transferred to the claims management department. For dealing with these contracts in detail, a huge amount of additional work was required from the company's side resulting in long processing times. This led to a high incidence of customer complaints. Each month, the company received about 450 complaints on long processing times, unsatisfactory customer service or calculation errors in final bills amongst others. Hence, a large quantity of customer-related data was to be stored and processed by the company's operational IT systems.

To optimize the current process performance and to restore customer satisfaction, management set up a BPI project to eliminate the problems as mentioned. The BPI project was organized in form of two workshops with operational staff from all parties involved in the process participating for eliciting the relevant knowledge. In a first workshop, the project goals were determined referring to the 'Voice-of-customer (VOC)' as well as 'Voice-of-business (VOB)' statements. These represent the verbally uttered customer expectations as well as employees' requirements on the process [50]. The main goals were to reduce the process

⁸RUPERT is freely available at <http://www.omilab.org/web/rupert/home> last accessed 05-06-2015

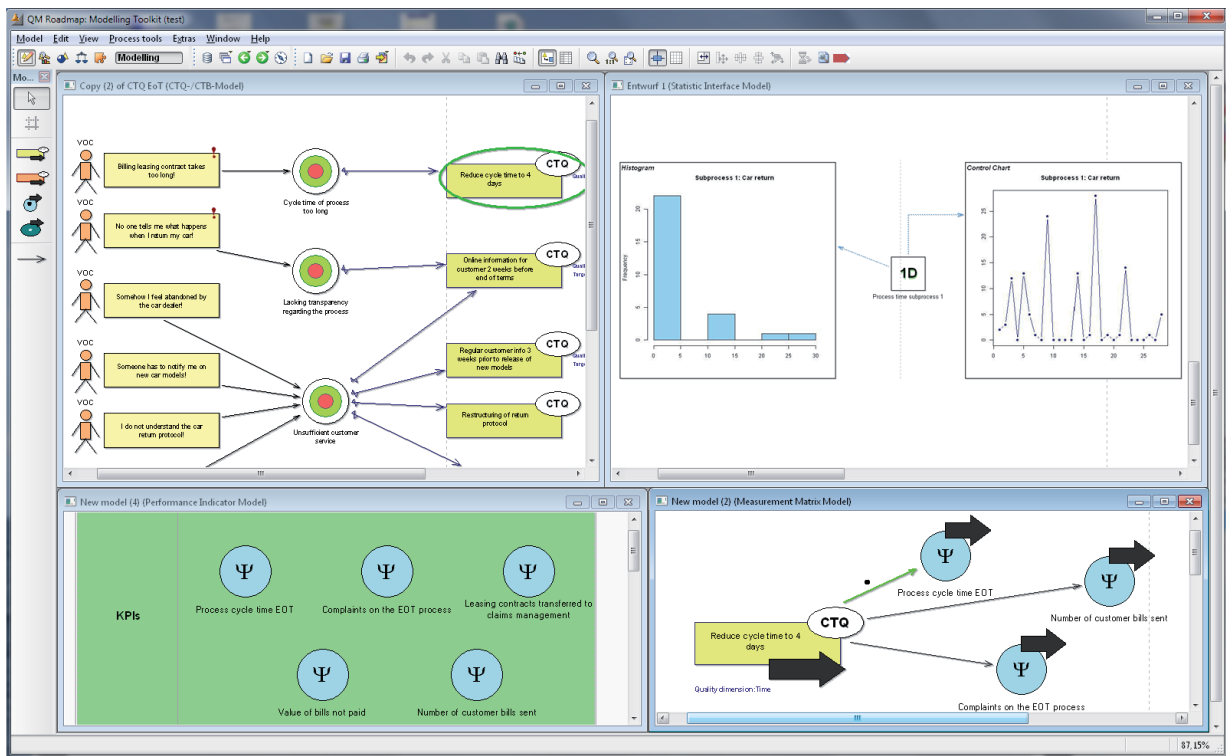


Figure 5. Screenshot of the Tool Showing the CTQ/CTB, Data Analysis, KPI, and Measurement Matrix Model Types (from left to right, top to bottom)

cycle times and to resolve the conflict of interests between the car dealer, the customer and the automotive bank.

Key performance indicators (KPIs) for measuring the level of goal achievement were then formulated by the workshop participants. These comprised (1) the cycle time of the process resp. its subprocesses, (2) the monthly distribution of complaints on the EOT process, (3) the share of ending leasing contracts transferred to the company’s claims management in each month, (4) the average value of bills not paid yet as well as (5) the number of customer bills sent to vehicle keepers by the automotive bank in the preceding year amongst others.

After this first workshop, measurement data to calculate the KPI values were collected. The data was retrieved from reports or extracted from IT systems as corresponding MS Excel files. The information was analyzed by using descriptive statistics to provide a decent base for subsequently identifying improvement opportunities. The files generated in the project were transformed into the CSV format and could be directly fed into the RUPERT modeling toolkit enabling an automatic data analysis.

The analysis of the data in the project provided new insights on problems of the current as-is process. For example, based on the data, those process steps causing delays in billing the leasing contracts could be identified. A series of process instances was extracted from the system for that purpose. Considering the time stamps of documents archived

in the database (e.g., customer bill), the cycle times for each subprocess could be determined. Subsequently, correlation analysis, histograms and control charts helped to uncover process variances in processing times, providing valuable indicators for process weaknesses.

Figure 5 shows an excerpt of the data and the analysis using our prototype (upper right model). The data was imported via a CSV interface into the ADOxx platform. By using the coupling to the underlying R platform the data can potentially be interpreted immediately using different types of statistical analyses and charts. The procedure of analyzing process performance data can be tremendously facilitated that way. Further, the conceptual model types used for systematically deriving the knowledge on the KPIs in accordance with the project goals are visualized. The project goals were determined with the help of the CTQ-/CTB-Model (Critical-to-Quality / Critical-to-Business) (upper left model), whereas the KPIs were defined via the Performance Indicator Model (lower left model). The assignment of KPIs to project goals for measuring goal achievement was done in the Measurement Matrix model (lower right model).

The results from the data analysis were presented and discussed at the beginning of the second workshop of the BPI initiative. Based on these results, opportunities for process improvement were hence derived. These comprised for example the introduction of Tablet PCs for vehicle inspectors to avoid media discontinuity during the calculation of a

car's terminal value. In total, the project led to cost savings of several million euros while the number of customer complaints could be significantly reduced at the same time.

The integration of a knowledge perspective in the form of different types of enterprise models with data analysis functionalities is thus a successful approach for supporting a real-world scenario. In particular, the joint interaction with knowledge and data in the same environment makes it easier for users to take into account relevant domain knowledge and data analyses for their decisions.

V. RELATED WORK

Although to the best of our knowledge we have so far not come across an approach that takes a similar perspective for integrating enterprise modeling and data analyses, there are some previous works that go in similar directions. Moody and Kortink have proposed an approach for the alignment of enterprise models and dimensional data models [51]. Starting from enterprise data models in the form of entity-relationship diagrams, they propose a method for transitioning to the design of data warehouses and data marts. Although this method provides a linkage between enterprise models and analyses on the data side, the approach is restricted to the data perspective of enterprise models and does not take into account knowledge aspects. In addition, their approach is unidirectional in the sense that it only covers a top-down perspective. It does not permit to regard the enterprise knowledge and the data analysis perspective in several consecutive iterations.

To a certain extent approaches in the area of semantic annotation of conceptual models are related to our approach. In this field it has been discussed how e.g. business process models can be automatically complemented with services based on semantic annotations [52], [53], [54]. These approaches are similar to ours in the way that they also add a data perspective to conceptual models. By adding semantic annotations, information from semantic schemata can be represented in the models and then processed using reasoning and queries. However, the focus is not on statistical data analyses but rather on making conceptual models executable with services.

From the perspective of visualization, approaches have been discussed that integrate information visualizations in the space of conceptual models [55], [30]. Thereby, also certain types of data analyses become possible. In contrast to the approach presented in this paper, statistical analyses had not been included in these approaches.

In the field of BPI, techniques and conceptual model types exist that relate KPIs or measurement data to results or process visualizations created in corresponding projects. The 'value-stream-map' for example strives to capture all information, material and activity flows of a business process to determine its current performance [56], [57]. In addition, relevant KPIs and measurement data can be noted in the

diagram (e.g. average processing time of an activity) indicating which activities are 'value-adding' and which are 'non-value-adding' [56]. Further, the 'measurement selection matrix' visualizes KPIs developed in a BPI project to measure the degree of project goal achievement and of fulfilling customer needs [56]. However, these diagrams types are created on a conceptual level. The statistical data analyses are to be performed separately and are not a constitutive element of the techniques.

VI. CONCLUSION AND OUTLOOK

In this paper we presented a framework for joining enterprise modeling and data analyses. The goal of this framework is to integrate the knowledge of domain experts as laid down in the form of enterprise models with data analyses. For this purpose we reverted to concepts of metamodeling that can be similarly applied to the area of enterprise modeling as well as data representation and analysis. With the presented technical architecture the feasibility of the approach could be illustrated, in particular also for enabling large scale data analyses as discussed today in the field of big data.

Integrating knowledge aspects in big data analysis potentially has a considerable impact for practice and academia. From the perspective of industry, the large quantities of enterprise models that have been created in the past may thus come to new life by serving as a basis for gaining insights through big data. In this way, previously uncoordinated, ad-hoc analyses of data can now be aligned with existing business knowledge. This adds both to the value of data analytics and of enterprise models. In terms of research opportunities, the presented approach can act as a foundation for investigating further enterprise modeling methods and their alignment with data analytics. Based on the large number of enterprise modeling methods that exist today this will require considerable research activities.

By describing a use case in the area of business process improvement, also the utility of the approach could be positively evaluated in a setting characterized by a huge quantity of customer data to be processed. As a limitation, the applicability of the approach was only shown for the use case as described above so far. In order to generalize the approach from an academic perspective, it needs to be applied to further use cases in practice and also in different industries. In addition, usability studies of the RUPERT modeling toolkit are still in progress. Whereas the first results seem very promising, a final judgement on its usability cannot be done yet.

As next steps we plan the further evaluation of the approach in user studies, practical projects and the further development of the technical architecture. In particular we will explore to what extent the technical architecture will be able to cope with big data stored in models and when

external storage systems will need to be used. This will be evaluated in detail in a range of experiments and simulations.

REFERENCES

- [1] A. Nazir, S. Raza, and C.-N. Chuah, "Unveiling Facebook: A Measurement Study of Social Network Based Applications," in *IMC'08*. ACM, 2008.
- [2] W. Fan and M. Gordon, "The Power of Social Media Analytics," *Communications of the ACM*, vol. 57, no. 6, pp. 74–81, 2014.
- [3] A. Musa, A. Gunasekaran, Y. Yusuf, and A. Abdelazim, "Embedded devices for supply chain applications: Towards hardware integration of disparate technologies," *Expert Systems with Applications*, vol. 41, pp. 137–155, 2014.
- [4] W. Brenner, D. Karagiannis, L. Kolbe, H. Krüger, H.-J. Lambert, L. Leifer, J. Leimeister, H. Oesterle, C. Petrie, H. Plattner, G. Schwabe, F. Übernickel, R. Winter, and R. Zarnekow, "User, Use & Utility Research : The digital user as new design perspective in Business & Information Systems Engineering," *Business & Information Systems Engineering*, vol. 6, no. 1, pp. 55–61, 2014.
- [5] A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review*, vol. October 2012, pp. 59–68, 2012.
- [6] T. Lukoianova and V. Rubin, "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?" in *Workshop on Advances In Classification Research Online*, vol. 24, 2014, pp. 4–15.
- [7] S. Kaisler, F. Armour, J. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in *46th Hawaii International Conference on System Sciences*. IEEE, 2013.
- [8] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data Mining with Big Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2014.
- [9] M. Schermann, H. Krcmar, H. Hemsén, V. Markl, C. Buchmüller, T. Bitter, and T. Hoeren, "Big Data - An Interdisciplinary Opportunity for Information Systems Research," *Business & Information Systems Engineering*, vol. 6, no. 5, pp. 261–266, 2014.
- [10] T. Le Dinh, T. Rickenberg, H.-G. Fill, and M. Breitner, "Enterprise Content Management Systems as a Knowledge Infrastructure: The Knowledge-Based Content Management Framework," *International Journal of e-Collaboration*, vol. 11, no. 3, pp. 49–70, 2015.
- [11] K. Dalkir, *Knowledge Management in Theory and Practice*. Oxford, UK: Elsevier Butterworth-Heinemann, 2005.
- [12] M. Nemetz, "A Meta-Model for Intellectual Capital Reports," in *Proceedings of the 6th International Conference on Practical Aspects of Knowledge Management*, D. Karagiannis and U. Reimer, Eds. Springer, 2006.
- [13] M. Fox and M. Grüninger, "On Ontologies and Enterprise Modelling," in *International Conference on Enterprise Integration Modelling Technology 97*. Springer, 1997.
- [14] R. Maier, *Knowledge Management Systems: Information and Communication Technologies for Knowledge Management*. Springer, 2004, vol. Second Edition.
- [15] H. Demirkan, R. J. Kauffman, J. A. Vayghan, H.-G. Fill, D. Karagiannis, and P. Maglio, "Service-oriented technology and management: Perspectives on research and practice for the coming decade," *Electronic Commerce Research and Applications*, vol. 7, no. 4, pp. 356–376, 2008.
- [16] J. Mylopoulos, "Conceptual Modeling and Telos," in *Conceptual Modelling, Databases and CASE: An Integrated View of Information Systems Development*, P. Loucopoulos and R. Zicari, Eds. Wiley, 1992, pp. 49–68.
- [17] Y. Wand and R. Weber, "Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda," *Information Systems Research*, vol. 13, no. 4, pp. 363–376, 2002.
- [18] S. Smolnik, F. Teuteberg, and O. Thomas, *Semantic Technologies for Business and Information Systems Engineering: Concepts and Applications*. IGI Press, 2012.
- [19] M. Koivulahti-Ojala and T. Kakola, "Design, Implementation, and Evaluation of a Virtual Meeting Tool-Based Innovation for UML Technology Training in Global Organizations," in *HICSS'2012*. IEEE, 2012.
- [20] H.-G. Fill and D. Karagiannis, "On the Conceptualisation of Modelling Methods Using the ADOxx Meta Modelling Platform," *Enterprise Modelling and Information Systems Architectures*, vol. 8, no. 1, pp. 4–25, 2013.
- [21] R. McAdam, J. Antony, M. Kumar, and S. A. Hazlett, "Absorbing new knowledge in small and medium-sized enterprises: A multiple case analysis of Six Sigma," *International Small Business Journal*, vol. 32, no. 1, pp. 81–109, 2014.
- [22] N. Visic, H.-G. Fill, R. Buchmann, and D. Karagiannis, "A Domain-specific Language for Modeling Method Definition: from Requirements to Grammar," in *Ninth International Conference on Research Challenges in Information Science*. IEEE, 2015.
- [23] H.-G. Fill, T. Redmond, and D. Karagiannis, "FDMM: A Formalism for Describing ADOxx Meta Models and Models," in *Proceedings of ICEIS 2012, Wroclaw, Poland*, L. Maciaszek, A. Cuzzocrea, and J. Cordeiro, Eds., vol. 3, 2012, pp. 133–144.
- [24] D. Karagiannis, H.-G. Fill, P. Hoeffler, and M. Nemetz, "Metamodeling: Some Application Areas in Information Systems," in *UNISCON*, R. Kaschek and et al., Eds. Springer, 2008, pp. 175–188.
- [25] H. Kern, A. Hummel, and S. Kühne, "Towards a Comparative Analysis of Meta-Metamodels," in *SPLASH '11 Workshops Proceedings*. Portland, USA: ACM, 2011, pp. 7–12.
- [26] H. Österle, J. Becker, U. Frank, H. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E. Sinz, "Memorandum on design-oriented information systems research," *European Journal of Information Systems*, vol. 20, pp. 7–10, 2011.

- [27] D. Karagiannis and H. Kühn, "Metamodeling Platforms," in *Third International Conference EC-Web 2002 Dexa 2002*, K. Bauknecht, A. Min Tjoa, and G. Quirchmayr, Eds. Springer, 2002, p. 182.
- [28] H.-G. Fill, "On the Conceptualization of a Modeling Language for Semantic Model Annotations," in *Advanced Information Systems Engineering Workshops, CAiSE 2011*, C. Salinesi and O. Pastor, Eds. Springer, 2011, pp. 134–148.
- [29] D. Harel and B. Rumpe, "Meaningful Modeling: What's the Semantics of "Semantics"?" *IEEE Computer*, vol. October 2004, pp. 64–72, 2004.
- [30] H.-G. Fill, *Visualisation for Semantic Information Systems*. Gabler, 2009.
- [31] H.-G. Fill, T. Redmond, and D. Karagiannis, "Formalizing Meta Models with FDM: The ADOxx Case," in *Enterprise Information Systems*, J. Cordeiro, L. Maciaszek, and J. Filipe, Eds. Springer, 2013, pp. 429–451.
- [32] T. Kühne, "Matters of (meta-) modeling," *Software and Systems Modeling*, vol. 5, pp. 369–385, 2006.
- [33] M. Eppler, "Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases," in *HICSS'2001*. IEEE, 2001.
- [34] A. Leutgeb, W. Utz, R. Woitsch, and H.-G. Fill, "Adaptive Processes in E-Government - A Field Report about Semantic-based Approaches from the EU-Project FIT," in *ICEIS'07. INSTICC*, 2007, pp. 264–269.
- [35] A. Leutgeb, W. Utz, R. Woitsch, M. Kossowski, and H.-G. Fill, "Agile Business Process Management in Public Administrations using Business Rules - A Realisation Approach within EU-Project FIT," in *TED Conference on e-Government*. Trauner Verlag, 2007, pp. 66–76.
- [36] H.-G. Fill and D. Karagiannis, "Semantic Visualization for Business Process Models," in *Int. Workshop on Visual Languages and Computing*. Grand Canyon, USA: Knowledge Systems Institute, 2006, pp. 168–173.
- [37] P. P.-S. Chen, "The Entity-Relationship Model-Toward a Unified View of Data," *ACM Transactions on Database Systems*, vol. 1, no. 1, pp. 9–36, 1976.
- [38] N. Roussopoulos and T. Yeh, "An Adaptable Methodology for Database Design," *IEEE Computer*, vol. 17, no. 5, pp. 64–80, 1984.
- [39] J. Trujillo, M. Palomar, J. Gomez, and I. Song, "Designing Data Warehouses with OO Conceptual Models," *IEEE Computer*, vol. 34, no. 12, pp. 66–75, 2001.
- [40] A. Morris, A. Abdelmoty, B. El-Geresy, and C. Jones, "A Filter Flow Visual Querying Language and Interface for Spatial Databases," *GeoInformatica*, vol. 8, no. 2, pp. 107–141, 2004.
- [41] A. Soyulu, M. Giese, E. Jimenez-Ruiz, E. Kharlamov, D. Zheleznyakov, and I. Horrocks, "OptiqueVQS Towards an Ontology-based Visual Query System for Big Data," in *5th Int. Conference on Management of Emergent Digital EcoSystems*. ACM, 2013, pp. 119–126.
- [42] A. Ohri, "Using R with Data and Bigger Data," in *R for Cloud Computing*, A. Ohri, Ed. Springer, pp. 193–215.
- [43] Royal Academy of Engineering, *Philosophy of Engineering Volume 1 of the proceedings of a series of seminars held at The Royal Academy of Engineering*, 2010.
- [44] A. Hevner, S. March, J. Park, and S. Ram, "Design Science in Information System Research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
- [45] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: <http://www.R-project.org>
- [46] Y. Song, G. Alatorre, N. Mandagere, and A. Singh, "Storage Mining: Where IT Management Meets Big Data Analytics," in *International Congress on Big Data*. IEEE, 2013, pp. 421–422.
- [47] F. Johannsen and H.-G. Fill, "Supporting Knowledge Elicitation and Analysis for Business Process Improvement through a Modeling Tool," in *International Conference on Business Informatics 2015*. AIS, 2015.
- [48] —, "RUPERT: A modelling tool for supporting business process improvement initiatives," in *Proceedings of DESRIST'2014*. Springer, 2014, pp. 418–422.
- [49] —, "Codification of Knowledge in Business Process Improvement Projects," in *European Conference on Information Systems (ECIS'2014)*. AIS, 2014.
- [50] S. Pande, P. P. Neuman, R., and R. Cavanagh, R., *The Six Sigma Way: How GE, Motorola, and other top companies are honing their performance*. New York: McGraw-Hill, 2000.
- [51] D. Moody and M. A. Kortink, "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design," in *Int. Workshop on Design and Management of Data Warehouses*. CEUR, 2000.
- [52] M. Hepp, F. Leymann, J. Domingue, A. Wahler, and D. Fensel, "Semantic business process management: A vision towards using semantic web services for business process management," in *ICEBE 2005*. IEEE, 2005, pp. 535–540.
- [53] H.-G. Fill, "Using Semantically Annotated Models for Supporting Business Process Benchmarking," in *BIR'2011 Conference*. Springer, 2011, pp. 29–43.
- [54] H.-G. Fill and P. Burzynski, "Integrating Ontology Models and Conceptual Models using a Meta Modeling Approach," in *11th International Protégé Conference*, 2009.
- [55] H.-G. Fill and P. Höfferer, "Visual Enhancements of Enterprise Models," in *Multikonferenz Wirtschaftsinformatik 2006*, F. Lehner, H. Nösekabel, and P. Kleinschmidt, Eds. GITO Verlag, 2006, pp. 541–550.
- [56] M. George, D. Rowlands, M. Price, and J. Maxey, *Lean Six Sigma Pocket Toolbox*. New York: McGraw-Hill, 2005.
- [57] J. P. Womack and D. T. Jones, *Lean Thinking*. New York: Simon & Schuster, 1996.