

*Henning Baars (Hrsg.)*

**Innovationsstränge in der  
Business-Intelligence-Forschung**

*Fünfter Workshop „Business Intelligence“*

der GI-Fachgruppe Business Intelligence  
in Zusammenarbeit mit dem Institut für Wirtschaftsinformatik  
der Technischen Universität Bergakademie Freiberg

am 12. Juni 2013 in Freiberg

**Tagungsband**



## Vorwort zum 5. Workshop „Business Intelligence“

**Der fünfte Workshop „Business Intelligence“** der gleichnamigen GI-Fachgruppe hatte erneut das Ziel, innovative Forschungsansätze und Forschungsergebnisse aus dem Bereich der Business Intelligence (BI) zu präsentieren, zu diskutieren und in Bezug zu setzen. Vorgestellt wurden neben originären Forschungsergebnissen bewusst auch Zwischenergebnisse aus Forschungsprojekten (*Research in Progress*), konkrete Forschungsideen sowie neue methodische Herangehensweisen.

Dieser Tagungsband enthält vier der angenommenen und auf dem Workshop vorgestellten Beiträge mit sehr unterschiedlichen methodischen und inhaltlichen Schwerpunkten:

Julian Ereth (Universität Stuttgart) und Daniel Dahl (Bayer Business Services) gehen in ihrem Beitrag auf den Themenkomplex „Business Intelligence in the Cloud“ ein. Die Autoren plädieren für eine Dekomposition von BI-Lösungen in BI-Services und eine gezielte Berücksichtigung der jeweiligen Abhängigkeiten zwischen den Services. Der Ansatz wird anhand von zwei Szenarios illustriert.

Marcus Hofmann, Robert Krawatzek und Frieder Jacobi (TU Chemnitz) diskutieren in „Entwicklung eines Software-Prototyps zur automatischen Erstellung nutzerspezifischer ETL-Dokumentation“ die Einsatzmöglichkeiten gestaltungsorientierter Forschung im BI-Bereich am Beispiel eines Ansatzes zur automatisierten Dokumentationserstellung. Aus ihren Erfahrungen leiten sie einige wertvolle Empfehlungen für die Durchführung designorientierter Forschungsprojekte ab.

Safwan Sulaiman, Jorge Marx Gómez (Universität Oldenburg) und Joachim Kurzhöfer (Lufthansa Systems) schlagen in „Business Intelligence Systems Optimization to Enable Better Self-Service Business Users“ vor, die Nutzungsverläufe von Power Usern und Analysten zu speichern, auszuwerten und darauf aufbauend über eine Vorschlagskomponente anderen Nutzern Empfehlungen für Analysen zu geben. Ziele sind die Diffusion von Analysewissen sowie die Förderung einer Self-Service BI.

Der Artikel „Zur methodischen Vorbereitung von Data-Mining-Projekten unter Verwendung von CRISP-DM im Kontext diskreter Produktionsprozesse“ von Uwe Wieland, Marco Fischer (TU Dresden) zeigt Anforderungen für eine Erweiterung zum Vorgehensmodell CRISP-DM für die Domäne diskreter Produktionsprozesse auf. Diesen wird mit der Entwicklung eines Prozessmetamodell als Sprachbeschreibung, eines darauf aufbauenden Prozessmodells sowie durch Hinweise zur Modellnutzung begegnet.

---

## Inhalt

Julian Ereth, Daniel Dahl: Fundamentals for a Service-based evaluation concept for Business Intelligence in the Cloud	1
Marcus Hofmann, Robert Krawatzek, Frieder Jacobi: Entwicklung eines Software-Prototyps zur automatischen Erstellung nutzerspezifischer ETL-Dokumentation: Ein Fallbeispiel für gestaltungsorientierte, problemzentrierte Forschung	21
Safwan Sulaiman, Jorge Marx Gómez, Joachim Kurzhöfer: Business Intelligence Systems Optimization to Enable Better Self-Service Business Users	35
Uwe Wieland, Marco Fischer: Zur methodischen Vorbereitung von Data-Mining-Projekten im Kontext diskreter Produktionsprozesse	47

# **Business Intelligence in the Cloud: Fundamentals for a Service-based Evaluation Concept**

*Julian Ereth<sup>1</sup>  
Dr. Daniel Dahl<sup>2</sup>*

<sup>1</sup>*University of Stuttgart*  
<sup>2</sup>*Bayer Business Services GmbH*

## **Abstract**

*Organizations can achieve many benefits by the integration of Cloud computing approaches in their Business Intelligence (BI) environments. However, the integrated and complex BI architectures make the integration of this new technology difficult. Since the Cloud is no universal remedy, it would be useful to have an evaluation model to decide when the use of Cloud BI is appropriate. This paper constitutes the first steps towards such a model by identifying and structuring consequences of Cloud computing on Service-based BI architectures. For this purpose, a comprehensive literature review and a case study in cooperation with Bayer, a large German company, were conducted. The results are illustrated on the basis of real life scenarios which indicate that an approach that considers the great number of complex interdependencies in BI architectures is required.*

## **1 Introduction**

Recently there has been a lot of talk about Cloud computing. Many people consider this new technology to have the potential to transform a large part of the IT industry. (Armbrust et al., 2010; Fenn & LeHong, 2011) While various sectors already use this new technology successfully (techconsult GmbH, 2012; Wallraff & Weber, 2012), the integration in more demanding areas such as Business Intelligence still holds many challenges (Baars & Kemper, 2010; Reyes, 2010; Willem & Jakobus, 2010).

Business Intelligence (BI) characterizes holistic approaches to build and run an integrated management support infrastructure. (Baars & Kemper, 2008; Moss & Atre, 2003) The wide range of this task raises the need for a great number of heterogeneous systems. (Kimball & Ross, 2002; McKnight, 2007a; Mohammed, Altmann, & Hwang, 2010) Over time, these systems often evolve into integrated and highly complex BI architectures.

(McKnight, 2007b) Moreover, new technologies like Cloud computing permanently pose new challenges to these infrastructures. In order to master the problems arising from this situation, organizational concepts like the idea of dedicated BI governance structures (Meredith, 2008) and BI service management (BISM) (Horakh, Baars, & Kemper, 2008; Kashanchi & Toland, 2006) emerged.

In an organizational context, the Cloud can be regarded as a new outsourcing alternative. (Baars & Kemper, 2010; Böhm, Leimeister, Riedl, & Krcmar, 2011) But apart from well-known issues with outsourcing, the Cloud comes with additional challenges that have to be taken into account when integrating this new technology into proprietary BI architectures. (Baars & Kemper, 2010; Willem & Jakobus, 2010) Problems often associated with Cloud-based BI such as vague data confidentiality, the fear of a vendor lock-in or simply technological limits indicate that not every service is appropriate to be obtained from the Cloud. (Armbrust et al., 2010; Chow et al., 2009; Reyes, 2010) It is therefore essential to be clear about the Cloud suitability of individual components and services in order to include Cloud Computing into a BI architecture successfully. Thus, the goal of this paper is to gather and structure impacts that have to be considered when planning the use of Cloud computing in the sophisticated domain of Business Intelligence. This represents the preliminary work in an on-going larger research process with the ambition to create an applicable framework for assessing the Cloud suitability of a BI service.

The paper will have the following course. Firstly, basic terms and concepts will be introduced and discussed on the basis of related literature. This part will mainly cover the basics of Cloud computing and IT service management as well as their adaption to the sector of Business Intelligence. Next, the methodical approach used in the research will be illustrated and the conducted case study will be introduced. Subsequently, the influences of Cloud BI derived from the reviewed literature and the conducted interviews, will be discussed in more detail to structure the problem further and carve out challenges. Afterwards, two real life scenarios that were developed in cooperation with an industrial partner will illustrate the major issues which mainly originate from the complexity and the interdependencies in a BI architecture. Finally, the results and next steps will be summarized in a concluding appraisal.

## **2 Overview of fundamental concepts and related topics**

Since the combination of Cloud computing and BI service management is a rather new topic, the following section clarifies elementary terms and discusses the basic ideas behind related concepts and their adaption to the domain of Business Intelligence.

## 2.1 Cloud computing

There are numerous varying understandings of the term Cloud computing depending on the point of view. (Vaquero, Rodero-Merino, Caceres, & Lindner, 2008) While the Cloud is often just seen as "infinite computing resources available on demand" (Armbrust et al., 2010, p. 51), there are also more sophisticated approaches which constitute Cloud computing as a new paradigm in IT sourcing that increases agility and enables new business models. (Baars & Kemper, 2010; Hayes, 2008; Willem & Jakobus, 2010) A widely used definition describes Cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (...) that can be rapidly provisioned and released" (Mell & Grance, 2009, p. 2). The services offered via the Cloud usually differ in between the supply of fundamental computing resources (IaaS) (Bundesverband Informationswirtschaft, 2009; Mell & Grance, 2009; Rodero-Merino et al., 2010) up to the provision of complex platform (PaaS) (Mell & Grance, 2009; L. Wang et al., 2008; Youseff, Butrico, & Da Silva, 2008) or application software (SaaS) (Bundesverband Informationswirtschaft, 2009; Kern, Lacity, & Willcocks, 2002; Mell & Grance, 2009).

Frequently associated advantages with Cloud computing are cost savings through Usage-Based Pricing models and the possibility to scale resources on demand. (Armbrust et al., 2010; Grossman, 2009; Hayes, 2008; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011) It is another common argument that the Cloud simplifies the IT and therefore improves the level of quality and increases agility. (Armbrust et al., 2010; Hayes, 2008; Reeves et al., 2009) Conversely, the use of Cloud computing also raises significant challenges such as a lack of Cloud trust and security, insufficient availability, legal issues or the fear of a vendor lock-in. (Armbrust et al., 2010; Dillon, Wu, & Chang, 2010; Hayes, 2008; Takabi, Joshi, & Ahn, 2010) Furthermore, in many cases the integration of a Cloud solution may not make sense for economic reasons or simply can't be realized due to technological limits. (Armbrust et al., 2010; Dillon et al., 2010; Voorsluys, Broberg, & Buyya, 2011)

## 2.2 BI services

BI service management (BISM) is an adaption of IT service management (ITSM), which emerged from the increasing complexity of IT systems in order to implement a more coherent approach to improve flexibility and manageability of IT environments (Van Haren Publishing, 2007; Winniford, Conger, & Erickson-Harris, 2009). Therefore, it can be summarized as "defining, managing, and delivering IT services to support business goals and customer needs" (Winniford et al., 2009, p. 153), whereat the mentioned service term represents a combination of technology, people and processes. (Best Management

Practice, 2012; Van Haren Publishing, 2007) This definition clarifies the more holistic organizational character of service management and thus clearly distinguishes it from a purely technical service-oriented architecture (SOA) approach (Krafzig, Banke, & Slama, 2005; Papazoglou, 2003). Furthermore, service management frameworks used in practice, like ITIL® or COBIT, usually include additional concepts to design, monitor and maintain services. (Unger, 2011; Van Haren Publishing, 2007; Winniford et al., 2009)

BI service management structures and allocates technological and organizational elements of BI solutions to individual services. (Kemper, Baars, & Horakh, 2007; Van Haren Publishing, 2007) To depict such a BI service the framework shown in figure 1 can be used. It has been derived in the context of research on BI outsourcing to unbundle the complex structures of contemporary BI Solutions and slice them into well-defined services. (Kemper et al., 2007) For this purpose, the following three dimensions are defined.

**Tool:** BI services usually use various subsystems (tools) to fulfill their functions like data ware houses, analyzing software or front ends. In order to classify these components, it is useful to refer to existing BI frameworks. (Kemper et al., 2007)

**Business specificity:** This dimension depicts the closeness of a service to the business. The following layers can be distinguished to clarify the specification of this dimension. (Kemper et al., 2007; Kern et al., 2002)

- *Hardware:* The provision and running of the relevant computing resources such as storage and other equipment necessary to operate one or more BI components.
- *Tools:* Involves BI software like ETL tools or Data Warehouse software.
- *Templates:* Represent preconfigured applications and prearranged contents that can be adapted to individual needs (e.g. reference models from software vendors).
- *Content:* This layer is related to the actual business semantics and therefore contains concrete business content. An example is the filled instance of a data cube.

**Life cycle:** The last dimension to be considered is the phase of the application life cycle. In most cases this dimension can be divided into development and operation since those are the most significant stages of a service. However, if necessary, this dimension can be refined to any appropriate granularity. (Kemper et al., 2007)



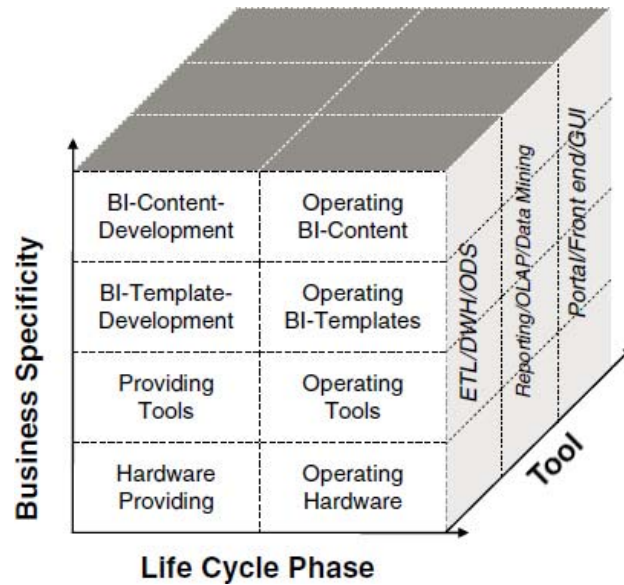


Figure 1: BI service framework – Kemper et al 2007

To enable a better coverage of customer needs and to achieve a higher scope of efficiency, compositions of BI services are delivered as holistic BI solutions which contribute to the user’s business processes. (Akkermans et al., 2004; Horakh et al., 2008) The combination and the multiple usages of BI services lead to interdependencies between particular components, activities and processes as well as responsibilities of BI solutions.

### 2.3 Applications of Cloud computing in a Service-based BI environment

This section describes various ways how Cloud computing can be integrated into a Business Intelligence architecture which mainly differ in their scope and the related complexity. (Baars & Kemper, 2010) Table 1 outlines different scenarios that will be further described below.

Scenario	Characteristics
<b>Add-on functionality</b>	Isolated, short-term, few impacts
<b>Tool integration</b>	Integrated, long-term, high impacts
<b>Solution provision</b>	Complex structures, high impacts

Table 1: Scenarios for Cloud BI

The most straightforward idea is to obtain particular *add-on functionality* from the Cloud. These add-ons deliver particular functional blocks for a temporary use. (Baars & Kemper,

2010; Marston et al., 2011) Due to its additional and short-term character, this scenario usually has a low impact on a given BI service environment and is therefore relatively risk free. (Baars & Kemper, 2010)

A more long-term oriented approach is the integration or replacement of a tool with a Cloud-based alternative. This scenario can provide many advantages but depending on the type of the integrated tool it can also cause far reaching implications for the existing BI services. (Baars & Kemper, 2010; Cullen & Willcocks, 2003) One major issue here, especially for the domain of BI with its close collaboration between systems and the high number of dependencies, is the lack of standards for the combination of cloud and non-cloud parts (Rochwerger et al., 2009; Vaquero et al., 2008).

An expansion of this idea is a solution provision where a provider supplies a complete end to end system in the Cloud. Although this concept is much more complex to integrate in a proprietary architecture, it might be suited for special purposes e.g. for temporary systems which have to be set up rapidly. (Baars & Kemper, 2010)

### **3 Methodology**

Due to the relatively new nature of the subject Cloud computing in the domain of Business Intelligence, a qualitative design with an exploratory character was used for this research. (Stebbins, 2001; Yin, 2010)

The ambition of this work is to examine the impacts of Cloud computing on Service-based BI architectures as figure 2 outlines. The study therefore tries to identify and analyze influencing factors and interdependencies in this context.

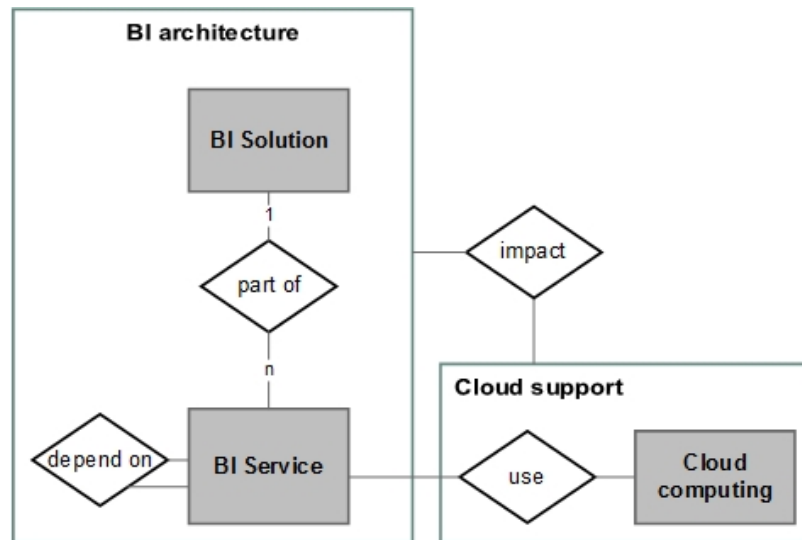


Figure 2: Conceptual framework

In the beginning, a comprehensive literature review (Auerbach & Silverstein, 2003; Yin, 2010) was conducted to get an overview of relevant topics and current research subjects. This review contained a semantic analysis of approximately 100 scientific papers, white-papers and market reports mainly from scientific databases like ACM, IEEE Xplore, SpringerLink and JSTOR as well as from other public accessible sources. Based on this theoretical foundation, a workshop and two group interviews with business intelligence experts at Bayer AG, a large German pharmaceutical company, were arranged to get a practical view on the subject.

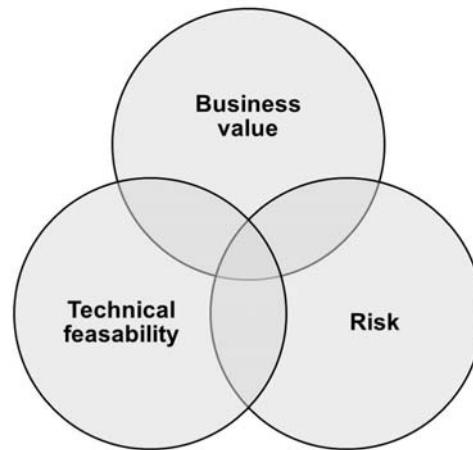
The gathered information and an additional analysis of the existing BI infrastructure in the cooperating company were used to structure challenges and consequences of Cloud-based Business Intelligence. Since it turned out that the interdependencies of systems and services play an essential role in the Cloud BI context, a further in-depth interview was conducted and multiple real life scenarios were developed to examine the related issues more precisely.

#### 4 Structuring of decision criteria for Cloud-based Business Intelligence

Since the usage of Cloud-based services in Business Intelligence can be treated as a special type of outsourcing (Baars & Kemper, 2010; Böhm et al., 2011), it seems appropriate to review established IT outsourcing approaches and afterwards carve out the special challenges with regard to Cloud computing in Business Intelligence services.

In order to justify an IT outsourcing decision, basically two main conditions have to be met: the technical feasibility has to be given and the gained business value, achieved through outsourcing, in this particular case over the Cloud, have to predominate the associated challenges and risks. (Akomode, Lees, & Irgens, 1998; Böhm et al., 2011; Cullen & Willcocks, 2003; Loh & Venkatraman, 1995)

Numerous heterogeneous factors and dependencies have to be considered in order to cope with these questions. Consequently, many established evaluation concepts in this context use Multi-Criteria Decision Making Methods with hierarchies to prioritize relevant criteria in order to decrease complexity of the problem. (Araz, Ozfirat, & Ozkarahan, 2007; Menzel, Schönherr, Nimis, & Tai, 2011; J.-J. Wang & Yang, 2007) According to these methods, the relevant Cloud decision criteria for BI services that were gathered from the literature review and conducted interviews, can be generalized by the following superior categories (Figure 3).



*Figure 3: Superior categories of Cloud impact*

**Business value:** The economic perspective is an essential pillar to assess the Cloud suitability of a BI service. (Armbrust et al., 2010; Marston et al., 2011; Mohammed et al., 2010; Skilton & Director, 2010) This category summarizes all criteria that affect the business value in a positive or negative way. Many decision criteria in this class are very obvious e.g. the reduction of costs or the increasing focus on key competencies. But there are also criteria which are more complicated to identify, for instance, the enabling of competitive advantages and new business models. A key problem here is that an assessment solely based on output measures is of limited value for understanding the whole range of IT impacts. (Mooney, Gurbaxani, & Kraemer, 1996) It is thus essential to consider the organizational dependencies and the long-range effects on the business value to constitute a realistic evaluation. (Mohammed et al., 2010; Mooney et al., 1996)

**Technical feasibility:** The technical feasibility is a basic condition, required to integrate the Cloud in an existing environment. Straightforward, the technical suitability of individual components can be verified by matching hardware and software requirements as well as technical indicators, such as data transfer volume or availability on Service Level Agreements of Cloud providers. (Buyya, Yeo, & Venugopal, 2008; Deb, 2010; Rimal, Jukan, Katsaros, & Goeleven, 2011)

However, despite this rather intuitive approach, the more profound issue is to assess the integration complexity of a Cloud-based component in an existing BI architecture. The sophisticated interdependencies in proprietary BI architectures (Horakh et al., 2008; McKnight, 2007b) and the lack of standards for the technical interoperability of Cloud services (Rochwerger et al., 2009; Vaquero et al., 2008) make integration to one of the major issues. The complexity manifests itself with factors such as the number of connected systems or the degree of individualization (e.g. through non-proprietary code). (Baars & Kemper, 2010; Deb, 2010)

**Risk:** Risk can be defined as “the possibility of loss or injury” (Boehm, 1991, p. 33) and plays an important role in any IT outsourcing decision. (Aubert, Patry, & Rivard, 1998; Earl, 1996; Hirschheim, Heinzl, & Dibbern, 2009) There are several approaches to assess the risk of IT outsourcing. They often consider hidden costs, contractual difficulties, Service Debasement and the loss of organizational competencies. (Aubert et al., 1998; Hirschheim et al., 2009)

In relation to Cloud BI, especially issues in security and trust are frequently referred to as the main risk factors. (Armbrust et al., 2010; Voorsluys et al., 2011) Regulatory and legal issues such as data privacy or legal liability are also often mentioned in this context. (Armbrust et al., 2010; Catteddu, 2010; Voorsluys et al., 2011) These risks mainly depend on factors like the confidentiality of the handled data or security requirements of the affected system. Other major concerns are the fear of a vendor lock in, which follows from the lack of standards and interoperability in Cloud computing (Marston et al., 2011; Meredith, 2008) and other technical risks like insufficient availability or unpredictable performance (Armbrust et al., 2010). In order to handle these risks the requirements and the available offers have to be aligned through service level agreements. (Buyya et al., 2008; Takabi et al., 2010)

A closer examination of the criteria in these three classes reveals many overlaps and interdependencies. For instance, a greater technical complexity directly affects the business value through increasing costs and higher technical and economic risks. Or similarly, gained business value through a more agile infrastructure strongly depends on the technical feasibilities of individual components as well as the BI architecture itself.

### **Complexity of BI services and its influence on the Cloud suitability**

Preservative common denominators of the discussed issues are the high complexity of BI architecture and the strong interdependencies between the involved BI services. This observation indicates that the complexity of the surrounding BI environment significantly exerts influence on the Cloud suitability of particular BI services. This complexity mainly results from highly integrated multi-layer Business Intelligence architectures needed to provide a holistic management support. (Baars & Kemper, 2008; McKnight, 2007b) In order to implement this enterprise-wide approach, these architectures are interwoven with all sorts of transactional systems through all hierarchy levels and business functions. (Chamoni & Gluchowski, 2004; Horakh et al., 2008; Moss & Atre, 2003) This integrated structure entails numerous interdependencies between BI solutions and particular BI services which have to be considered when integrating Cloud-based systems. Furthermore, the organizational dependencies in such Service-based environments have to be considered. For instance, new components have to be integrated in existing operating structures like change or incidence management. This correlation between the complexity of a system and the ability to use a Cloud approach is also illustrated by the real life scenarios in the following section.

## **5 Inspected scenarios**

In the course of the conducted case study, several scenarios based upon real life cases from the Business Intelligence environment of the involved industrial partner were analyzed. Below, two of these scenarios are used to illustrate the previously discussed issues.

### **Scenario 1: A temporary Social Media Analytics platform**

This case originated from an innovation project with the aim to gather relevant marketing information from social networks and online communities. The unique character of social media raised the need for special tools. For instance, data- and text-mining algorithms e.g. sentiment analysis, were required to gain knowledge from unstructured content. Furthermore, the large size of data sets raised additional performance challenges. (Barbier & Liu, 2011; Kaplan & Haenlein, 2010) The use of a Cloud-based solution in this scenario seems

likely since many providers offer a short-term provision of the needed functionality with a scalable environment.

Based on the previous classification, this case represents a typical *add-on* scenario where special functions are obtained from the Cloud. Due to the inexistence of proprietary tools and the isolated short-term use in a project, the usage of a Cloud-based solution in this case would have little impact on the existing BI architecture; hence, there is no need for an extensive integration process. In addition, the fact that there are barely interdependencies with existing BI systems entails that the focus of a Cloud evaluation in this case is mainly on local decision criteria like economic and technical factors of the system.

### **Scenario 2: Cloud support in a complex best-of-breed BI solution**

The second scenario is based upon a large proprietary BI solution. Due to the wide range of activities, the system over time has evolved into a complex best-of-breed solution; hence, it contains multiple subsystems across all conceptual levels and is highly integrated into the Service-based BI architecture as well as into operational systems.

In consideration of the large scale, this system provides a wide range of *tool replacement* and *add-on* possibilities across all conceptual layers. For instance, services on the data layer can benefit from on demand scaling of computing resources to handle big data volumes or load peaks. But also more sophisticated applications like a short-term SaaS provision of particular user interfaces on the access layer are imaginable.

Since the services all interplay in a holistic BI solution, the manifold interdependencies in this architecture have to be considered when it should be enriched with Cloud-based components. As previously outlined, trust issues are a major concern when Cloud providers are involved. Consequently, data confidentiality has to be ensured across all internal and especially external systems in this end-to-end solution. The strong collaborations between the various systems in the case complicate this issue further.

A change in the given BI architecture can moreover provoke far-reaching organizational impacts. It must be guaranteed that services obtained from the Cloud provide the quality and performance that users and other depending services require. Furthermore, a seamless integration into the solution life cycle must be possible to ensure a long-term integration. Therefore, an integration in given service management structures such as change and incident management is absolutely essential.

### Comparison of the scenarios

These two scenarios illustrate the various circumstances and connected problems that could occur when integrating Cloud approaches into BI architectures.

The obvious difference between the cases is the level of complexity. While scenario one has an isolated, short-term character and comes with nearly no integration effort. The second case, in contrast, deals with a large and highly integrated BI solution. The integration effort in this sophisticated environment is rather large since the manifold interdependencies between the affected services have to be considered. Moreover, the long-term replacement of a component in this holistic approach comes with the need of an organizational integration in existing BI governance structures.

## 6 Conclusion and next steps

The closer examination of the impacts of Cloud Computing on Service-based Business Intelligence environments has shown that many factors have to be considered to reasonably assess the use of a Cloud-based solution. Apart from common issues with Cloud BI, like security and trust, especially the complex structure of BI architectures raises the need of a more coherent perspective to evaluate the Cloud suitability of particular BI services. The previously discussed scenarios illustrate, that such an approach have to particularly consider the dependencies between BI services and the resulting technical and organizational integration complexity.

The next step in the research process towards a Service-based evaluation concept for Business Intelligence in the Cloud is to substantiate the discussed issues and to operationalize the decision criteria from this initial work in a prototype framework that considers the sophisticated interdependencies in a BI architecture. Additionally, further application scenarios have to be developed to verify the use of the framework in practice.

## 7 References

- Akkermans, H., Baida, Z., Gordijn, J., Peiia, N., Altuna, A., & Laresgoiti, I. (2004). Value Webs: using ontologies to bundle real-world services. *Intelligent Systems, IEEE*, 19(4), 57-66.
- Akomode, O Joseph, Lees, Brian, & Irgens, Christopher. (1998). Constructing customised models and providing information to support IT outsourcing decisions. *Logistics Information Management*, 11(2), 114-127.



- Araz, C., Ozfirat, P. M., & Ozkarahan, I. (2007). An integrated multicriteria decision-making methodology for outsourcing management. *Computers & Operations Research*, 34(12), 3738-3756. doi: DOI 10.1016/j.cor.2006.01.014
- Armbrust, Michael, Fox, Armando, Griffith, Rean, Joseph, Anthony D., Katz, Randy, Konwinski, Andy, . . . Zaharia, Matei. (2010). A view of cloud computing. *Commun. ACM*, 53(4), 50-58.
- Aubert, B.A., Patry, M., & Rivard, S. (1998). *Assessing the risk of IT outsourcing*. Paper presented at the System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conference on.
- Auerbach, C.F., & Silverstein, L.B. (2003). *Qualitative Data: An Introduction to Coding and Analysis*: New York University Press.
- Baars, Henning, & Kemper, Hans-Georg. (2010). *Business Intelligence in the Cloud?* Paper presented at the PACIS.
- Baars, Henning, & Kemper, Hans-George. (2008). Management Support with Structured and Unstructured Data - An Integrated Business Intelligence Framework. *Inf. Sys. Manag.*, 25(2), 132-148.
- Barbier, Geoffrey, & Liu, Huan. (2011). Data Mining in Social Media *Social Network Data Analytics* (pp. 327-352).
- Best Management Practice. (2012). Best Management Practice portfolio: common glossary of terms and definitions.
- Boehm, Barry W. (1991). Software risk management: principles and practices. *Software, IEEE*, 8(1), 32-41.
- Böhm, Markus, Leimeister, Stefanie, Riedl, Christoph, & Krcmar, Helmut. (2011). Cloud Computing – Outsourcing 2.0 or a new Business Model for IT Provisioning? In F. Keuper, C. Oecking & A. Degenhardt (Eds.), *Application Management* (pp. 31-56): Gabler.
- Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (2009). *Cloud Computing - Evolution in der Technik, Revolution im Business. BITKOM Leitfaden, Oktober 2009*.
- Buyya, Rajkumar, Yeo, Chee Shin, & Venugopal, Srikumar. (2008). *Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities*. Paper presented at the High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on.
- Catteddu, Daniele. (2010). *Cloud Computing: benefits, risks and recommendations for information security*: Springer.
- Chamoni, Peter, & Gluchowski, Peter. (2004). Integrationstrends bei Business-Intelligence-Systemen - Empirische Untersuchung auf Basis des Business Intelligence Maturity Model. *Wirtschaftsinformatik*, 46(2), 119-128.
- Chow, Richard, Golle, Philippe, Jakobsson, Markus, Shi, Elaine, Staddon, Jessica, Masuoka, Ryusuke, & Molina, Jesus. (2009). *Controlling data in the cloud: outsourcing computation without outsourcing control*. Paper presented at the Proceedings of the 2009 ACM workshop on Cloud computing security.
- Cullen, S., & Willcocks, L.L. (2003). *Intelligent It Outsourcing: Eight Building Blocks to Success*: Elsevier Butterworth-Heinemann.
- Deb, Brijesh. (2010). Assess enterprise applications for cloudmigration. *White Paper, IBM developer works*.
- Dillon, T., Wu, Chen, & Chang, E. (2010). *Cloud Computing: Issues and Challenges*. Paper presented at the Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on.

- Earl, Michael J. (1996). The risks of outsourcing IT. *Sloan management review*, 37, 26-32.
- Fenn, Jackie, & LeHong, Hung. (2011). Hype cycle for emerging technologies, 2011. *Gartner, July*.
- Grossman, R.L. (2009). The Case for Cloud Computing. *IT Professional*, 11(2), 23-27.
- Hayes, Brian. (2008). Cloud computing. *Commun. ACM*, 51(7), 9-11.
- Hirschheim, R.R.A., Heinzl, A., & Dibbern, J. (2009). *Information Systems Outsourcing: Enduring Themes, Global Challenges, and Process Opportunities*: Springer Berlin Heidelberg.
- Horakh, Thomas A., Baars, Henning, & Kemper, Hans-Georg. (2008). *Mastering Business Intelligence Complexity - A Service-Based Approach as a Prerequisite for BI Governance*. Paper presented at the AMCIS.
- Kaplan, Andreas M., & Haenlein, Michael. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59 - 68.
- Kashanchi, Ramisa, & Toland, Janet. (2006). Can ITIL contribute to IT/business alignment? An initial investigation. *WIRTSCHAFTSINFORMATIK*, 48(5), 340-348.
- Kemper, Hans-Georg, Baars, Henning, & Horakh, Thomas A. (2007). *Business Intelligence Outsourcing - A Framework*. Paper presented at the ECIS.
- Kern, Thomas, Lacity, Mary Cecelia, & Willcocks, Leslie. (2002). *Netsourcing: renting business applications and services over a network*: FT Press.
- Kimball, Ralph, & Ross, Margy. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (2nd ed.). New York, NY, USA: John Wiley & Sons, Inc.
- Krafzig, D., Banke, K., & Slama, D. (2005). *Enterprise SOA: Service Oriented Architecture Best Practices*: Prentice Hall Professional Technical Reference.
- Loh, Lawrence, & Venkatraman, N. (1995). *An Empirical Study of Information Technology Outsourcing: Benefits, Risks, and Performance Implications*. Paper presented at the ICIS 1995 Proceedings.
- Marston, Sean, Li, Zhi, Bandyopadhyay, Subhajyoti, Zhang, Juheng, & Ghalsasi, Anand. (2011). Cloud computing - The business perspective. *Decision Support Systems*, 51(1), 176-189.
- McKnight, W. (2007a). Moving Business Intelligence to the Operational World, Part 1. *DM REVIEW*, 17(8), 28.
- McKnight, W. (2007b). New Age Data Warehousing. *DM REVIEW*, 17(11), 49.
- Mell, Peter, & Grance, Tim. (2009). The NIST Definition of Cloud Computing.
- Menzel, Michael, Schönherr, Marten, Nimis, Jens, & Tai, Stefan. (2011). (MC2)2: A Generic Decision-Making Framework and its Application to Cloud Computing. *CoRR*, abs/1112.1851.
- Meredith, Rob. (2008). *Information Technology Governance and Decision Support Systems*. Paper presented at the CDM.
- Mohammed, AshrafBany, Altmann, Jörn, & Hwang, Junseok. (2010). Cloud Computing Value Chains: Understanding Businesses and Value Creation in the Cloud. In D. Neumann, M. Baker, J. Altmann & O. Rana (Eds.), *Economic Models and Algorithms for Distributed Systems* (pp. 187-208): Birkhäuser Basel.
- Mooney, John G., Gurbaxani, Vijay, & Kraemer, Kenneth L. (1996). A process oriented framework for assessing the business value of information technology. *SIGMIS Database*, 27(2), 68-81.

- Moss, L.T., & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*: Pearson Education.
- Papazoglou, M.P. (2003). *Service-oriented computing: concepts, characteristics and directions*. Paper presented at the Web Information Systems Engineering, 2003. WISE 2003. Proceedings of the Fourth International Conference on.
- Reeves, Drue, Blum, Dan, Watson, Richard, Creese, Guy, Blakley, Bob, Haddad, Chris, Lewis, Jamie. (2009). *Cloud Computing: Transforming IT*. *Burton Group publication, Utah USA, 84047-84169*.
- Reyes, Eumir P. (2010). *A systems Thinking approach to business intelligence solutions based on cloud computing*. Massachusetts Institute of Technology.
- Rimal, BhaskarPrasad, Jukan, Admela, Katsaros, Dimitrios, & Goeleven, Yves. (2011). Architectural Requirements for Cloud Computing Systems: An Enterprise Cloud Approach. *Journal of Grid Computing*, 9(1), 3-26.
- Rochwerger, Benny, Breitgand, David, Levy, Eliezer, Galis, Alex, Nagin, Kenneth, Llorente, Ignacio M, . . . others. (2009). The reservoir model and architecture for open federated cloud computing. *IBM Journal of Research and Development*, 53(4), 4-1.
- Rodero-Merino, Luis, Vaquero, Luis M., Gil, Victor, Gal'a, n, Ferm'i,n, Font'a, n, Javier, Montero, Rub'e, n S., & Llorente, Ignacio M. (2010). From infrastructure delivery to service management in clouds. *Future Gener. Comput. Syst.*, 26(8), 1226-1240.
- Skilton, Mark, & Director, Capgemini. (2010). Building Return on Investment from Cloud Computing. *White Paper, The Open Group*.
- Stebbins, R.A. (2001). *Exploratory Research in the Social Sciences*: SAGE Publications.
- Takabi, H., Joshi, J.B.D., & Ahn, Gail-Joon. (2010). Security and Privacy Challenges in Cloud Computing Environments. *Security Privacy, IEEE*, 8(6), 24-31.
- techconsult GmbH. (2012). IT-Cloud-Index 2012. *techconsult GmbH Q4 2012*.
- Unger, Carsten. (2011). *Entwicklung eines Rahmenkonzepts für das Management von Business-Intelligence-Systemen*. (Hochschulschrift), Eul, Lohmar. Retrieved from <http://d-nb.info/1016666071/04>
- Van Haren Publishing. (2007). *It Service Management: An Introduction*: Van Haran Publishing.
- Vaquero, Luis M., Rodero-Merino, Luis, Caceres, Juan, & Lindner, Maik. (2008). A break in the clouds: towards a cloud definition. *SIGCOMM Comput. Commun. Rev.*, 39(1), 50-55.
- Voorsluys, William, Broberg, James, & Buyya, Rajkumar. (2011). Introduction to Cloud Computing *Cloud Computing* (pp. 1-41): John Wiley & Sons, Inc.
- Wallraff, Bruno, & Weber, Mathias. (2012). Cloud Monitor 2012. *KPMG Studie*.
- Wang, Jian-Jun, & Yang, De-Li. (2007). Using a hybrid multi-criteria decision aid method for information systems outsourcing. *Computers & Operations Research*, 34(12), 3691 - 3700.
- Wang, Lizhe, Tao, Jie, Kunze, M., Castellanos, A.C., Kramer, D., & Karl, W. (2008). *Scientific Cloud Computing: Early Definition and Experience*. Paper presented at the High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on.
- Wang, MingXue, Bandara, Kosala Yapa, & Pahl, Claus. (2010). *Process as a Service*. Paper presented at the IEEE SCC.
- Willem, Thompson, & Jakobus, van der Walt. (2010). Business intelligence in the cloud. *SA Journal of Information Management*, 12(1).

- Winniford, MaryAnne, Conger, Sue, & Erickson-Harris, Lisa. (2009). Confusion in the Ranks: IT Service Management Practice and Terminology. *Information Systems Management*, 26(2), 153-163.
- Yin, R.K. (2010). *Qualitative Research from Start to Finish*: Guilford Publication.
- Youseff, L., Butrico, M., & Da Silva, D. (2008). *Toward a Unified Ontology of Cloud Computing*. Paper presented at the Grid Computing Environments Workshop, 2008. GCE '08.

# **Entwicklung eines Software-Prototyps zur automatischen Erstellung nutzerspezifischer ETL-Dokumentation: Ein detailliertes Fallbeispiel für gestaltungsorientierte, problemzentrierte Forschung**

*Marcus Hofmann, Robert Krawatzek, Frieder Jacobi*

*Technische Universität Chemnitz*

*Fakultät für Wirtschaftswissenschaften, Professur Wirtschaftsinformatik II*

## **Abstract**

*Die Durchführung von rigoroser und relevanter gestaltungsorientierter Forschung (engl. Design Science Research, DSR) gewinnt in der Wirtschaftsinformatik international an Akzeptanz und Bedeutung. Der vorliegende Beitrag verfolgt das Ziel, das Verständnis von DSR in der Wirtschaftsinformatik zu schärfen und DSR-Erfahrungen mit der Community zu teilen. Dafür wird ein Fallbeispiel, das ein umfangreiches, mit DSR durchgeführtes Forschungsprojekt zum Gegenstand hat, detailliert beschrieben und anschließend kritisch diskutiert.*

## **1 Einleitung**

Die gestaltungsorientierte Herangehensweise zur Problemlösung findet in der deutschen Wirtschaftsinformatik seit jeher Anwendung (Österle et al., 2010) und wird zunehmend im angelsächsischen Raum befürwortet (Chen, 2011). Die von Hevner et al. (2004) vorgeschlagenen Leitsätze sowie der von Peffers et al. (2007) angegebene Prozess für die Durchführung von rigoroser und relevanter gestaltungsorientierter Forschung (engl. Design Science Research, DSR) verhalf dem gestaltungsorientierten Problemlösungsansatz zur internationalen Akzeptanz.

Sowohl Hevner et al. (2004) als auch Peffers et al. (2007) nutzen in ihren Veröffentlichungen abgeschlossene und veröffentlichte Forschungsprojekte, um die Umsetzbarkeit und die Anwendung ihrer entwickelten Ergebnisse zu demonstrieren. Hevner et al. geben hierfür drei Beispiele an und veranschaulichen jeweils rückwirkend, wie diese die sieben

vorgeschlagenen DSR-Leitsätze befolgt haben. Peffers et al. demonstrieren ebenfalls rückwirkend für die vier möglichen Forschungseinstiegspunkte („problemzentriert“, „zielstellungszentriert“, „entwurfs- und entwicklungszentriert“ sowie „initiiert durch Auftraggeber/Kontext“) die Anwendung der DSR-Forschungsaktivitäten mithilfe eines Fallbeispiels. Der Einsatz von Fallbeispielen<sup>1</sup> veranschaulicht die Anwendung von DSR und erhöht das DSR-Verständnis (Hevner et al., 2004; Peffers et al., 2007).

Die von Peffers et al. (2007) präsentierte DSR-Demonstration leistet einen Beitrag zur Steigerung des DSR-Verständnisses, geht aber aufgrund der nachträglich zugeordneten DSR-Forschungsaktivitäten sowie der kompakten Darstellung der Fallbeispiele (jedes Forschungsprojekt wird auf maximal drei Seiten erläutert) aus Sicht der Autoren nicht genügend in die Tiefe und öffnet somit einen Interpretationsspielraum.

Um das DSR-Verständnis weiter zu schärfen, präsentiert der vorliegende Beitrag daher zunächst ein detailliertes DSR-Fallbeispiel, in dem ein abgeschlossenes Forschungsprojekt präsentiert wird, welches den DSR-Prozess vollständig durchlaufen hat (Kapitel 2). Die Struktur des Fallbeispiels orientiert sich an dem Vorgehen von Peffers et al. (2007). Abschließend folgt eine kritische Diskussion, in der dargelegt wird, inwiefern das Vorgehen des präsentierten Forschungsprojektes anzupassen ist, um die Rigorosität des DSR-Forschungsergebnisses zu steigern (Kapitel 3).

## **2 DSR-Fallbeispiel „Computer-Aided Data Warehouse Engineering“**

Das Forschungsprojekt „Computer-Aided Data Warehouse Engineering“ (CAWE) der Technischen Universität Chemnitz hatte mit einer Laufzeit von 3 Jahren (August 2010 bis Juli 2013) die Entwicklung eines modellgetriebenen Vorgehens zur Unterstützung des Lebenszyklus von Business-Intelligence-Systemen (BI-Systemen) zum Forschungsgegenstand. Ein wesentlicher Schwerpunkt der Forschungsarbeit lag in der Entwicklung eines Software-Prototyps zur automatischen Erzeugung von Dokumentation für Extraktions-, Transformations- und Ladeprozesse (ETL-Prozesse) innerhalb bestehender BI-Systeme (Re-Dokumentation). Ausschlaggebend für die Entwicklung dieses Prototyps war eine zu Projektbeginn durchgeführte Umfrage (Gluchowski et al., 2011; Hofmann et al., 2012), welche einen entsprechenden Bedarf identifizierte. Das CAWE-Projekt ist demnach ein Fallbeispiel für DSR-Forschung mit einem problemzentrierten Einstiegspunkt in die Forschungsaktivitäten.

---

<sup>1</sup> Im Rahmen dieses Beitrages wird der Begriff „Fallbeispiel“ für die Bezeichnung von beispielhaften Anwendungsfällen (engl. Use Cases) verwendet und ist vom Begriff „Fallstudie“ (engl. Case Study) abzugrenzen.

Dieses Kapitel beschreibt das Vorgehen und die Ergebnisse des Projektes anhand der von Peffers et al. (2007) bezüglich des DSR-Prozesses vorgeschlagenen Aktivitäten.

## 2.1 Aktivität 1: Problemidentifikation und Motivation

### *Problemidentifikation*

Eine in 2011 von CAWE durchgeführte Umfrage unter 119 Unternehmen hat ergeben, dass BI-Systeme gar nicht oder nur unzureichend dokumentiert werden (Gluchowski et al., 2011; Hofmann et al., 2012). BI-Architekturkomponenten werden lediglich von 36,1% der Befragten dokumentiert. Für den Fall, dass diese dokumentiert werden, betrifft das zu 79,1% die Berichte und Reports. Das Schlusslicht bilden ETL-Prozesse mit einer Dokumentationshäufigkeit von 53,5%.

Ferner wurde gezeigt, dass die Dokumentation häufig nur einmalig in der Entwurfs- oder der Entwicklungsphase der Systeme erfolgt. Das Aufwand-Nutzen Verhältnis für die Erstellung von Dokumentation wurde von den befragten Unternehmen ungünstig bewertet. So schätzten lediglich 48% der Umfrageteilnehmer den durch Dokumentation gestifteten Nutzen höher ein als den dadurch verursachten Aufwand. Als Gründe dafür wurden z.B. (I) der hohe personelle Aufwand bei der Dokumentationserstellung, (II) die zu geringe Haltbarkeit (sinkende Aktualität vorhandener Dokumentation im Zeitverlauf) sowie (III) die fehlende strukturelle und inhaltliche Eignung der Dokumentation für unterschiedliche Nutzergruppen genannt.

### *Motivation*

Mit Hilfe der Literatur motivierte CAWE die Relevanz von ETL-Dokumentationen, in dem es den möglichen Nutzen für verschiedene Adressatengruppen herleitete (Jacobi et al., 2012). Da die ETL-Prozessmodellierung bis zu 80% der Entwicklungszeit in Data-Warehouse-Projekten in Anspruch nimmt (Greenfield, 1996), ist die Dokumentation von ETL-Prozessen – welche zudem laut Umfrage am seltensten dokumentiert werden – von besonderem Interesse. Von einer ETL-Dokumentation profitieren *Entwickler* und *Unternehmen* wie folgt.

Entwickler erhalten Informationen über den aktuellen technischen Aufbau des Systems, wodurch das Umsetzen von Änderungen und Erweiterungen erleichtert wird.

Unternehmen profitieren von stets aktueller Dokumentation durch Kosteneinsparungen, die sich u.a. aus folgenden Punkten ergeben:

- durch eine verbesserte Qualität und daraus resultierender gewonnener Zeit, die bisher benötigt wurde, um mangelhaft dokumentierte Systeme zu verstehen (Chikofsky & Cross, 1990),
- neue Mitarbeiter bekommen die Möglichkeit, sich schneller in die Systeme einzuarbeiten, um diese produktiv nutzen zu können und
- Expertenwissen liegt in expliziter Form vor, wodurch beim Ausscheiden von Wissensträgern keine mit hohem Aufwand zu füllende Wissenslücke entsteht.

## **2.2 Aktivität 2: Beschreibung der Zielstellungen**

Im Zuge der Ausarbeitung der Problemstellung identifizierte CAWE folgende Punkte, die dafür verantwortlich sind, dass Unternehmen ETL-Prozesse nicht bzw. nur unzureichend dokumentieren: (I) zu hoher personeller Aufwand, (II) zu geringe Haltbarkeit und (III) fehlende strukturelle und inhaltliche Eignung für verschiedene Nutzergruppen.

Aus diesen wurden die Zielstellungen abgeleitet, welche eine nutzenbringende Dokumentation zu erfüllen hat:

- (I) kostengünstige Erstellung,
- (II) hochwertige Qualität und
- (III) nutzerspezifische Struktur und Inhalt.

Bei einer Analyse des Softwaremarktes wurde kein Werkzeug gefunden, das eine umfassende Dokumentation von ETL-Prozessen unterstützt und dabei die genannten Zielstellungen abdeckt (Jacobi et al., 2012). Daher verfolgte CAWE das Forschungsziel, einen Software-Prototyp (im Sinne des DSR-Artefakttyps „Instanziierung“ (March & Smith, 1995; Hevner et al, 2004)) zu entwickeln, der die Dokumentation von ETL-Prozessen unter Berücksichtigung der hergeleiteten Zielstellungen unterstützt.

## **2.3 Aktivität 3: Entwurf und Entwicklung**

### ***Entwurf einer Lösung ausgehend von den Zielstellungen***

Nachdem Gluchowski & Kurze (2010) bereits gezeigt haben, dass der Aufwand für die Erstellung von Dokumentation für multidimensionale Datenstrukturen (Spezifika der Datenhaltungskomponenten) durch den Einsatz von automatisierten Dokumentationsprozessen um bis zu 75% gesenkt werden kann, wurde auf eine automatisierte Lösung zur Erfüllung von Anforderung I des Forschungsziels gesetzt.



Anschließend blieb zu klären, durch welche Eigenschaften sich eine qualitativ hochwertige Dokumentation (Anforderung II) auszeichnet. Nach Wallmüller (Wallmüller, 2001) ist eine hochwertige Dokumentation durch folgende acht Merkmale gekennzeichnet:

- *„Änderbarkeit*: Eignung von Dokumenten zur Ermittlung aller von einer Änderung betroffenen Dokumententeile und zur Durchführung der Änderung.
- *Aktualität*: Übereinstimmung der Beschreibung des Programms in der Dokumentation mit dem jeweils geltenden Zustand des Programms.
- *Eindeutigkeit*: Eignung von Dokumenten zur unmissverständlichen Vermittlung von Information an jeden Leser.
- *Identifizierbarkeit*: Eindeutige Ansprechbarkeit der Teile von Dokumenten, die Angaben zu einem abgegrenzten Sachverhalt geben, die den Leser interessieren.
- *Normkonformität*: Erfüllung der für die Erstellung von Dokumenten geltenden Vorschriften und Normen.
- *Verständlichkeit*: Eignung von Dokumenten zur erfolgreichen Vermittlung der darin enthaltenen Informationen an einen sachkundigen Leser.
- *Vollständigkeit*: Vorhandensein der für den Zweck der Dokumentation notwendigen und hinreichenden Informationen [hinreichende Informationen liegen dann vor, wenn ein Dokument so wenig Informationen wie möglich, aber dennoch alle zur Erfüllung des Dokumentzwecks nötigen Informationen beinhaltet; Anm. d. A.].
- *Widerspruchsfreiheit*: Nichtvorhandensein von einander entgegenstehenden Aussagen im Dokument.“

Um Anforderung III des Forschungsziels im Sinne einer geeigneten Präsentation des Dokumentationsinhalts zu erfüllen, empfiehlt (Wallmüller, 2001) verschiedene, an den Zielgruppeninteressen ausgerichtete, Instanzen der Dokumentation zu erstellen.

Da nicht alle von Wallmüller vorgeschlagenen Dokumentationsmerkmale in Abhängigkeit von der Nutzergruppe verschieden ausgeprägt sein müssen, wurde in Forward & Lethbridge (2002) und Krawatzeck et al. (2011) eine Unterteilung der Merkmale von hochwertiger Dokumentation in zielgruppenabhängige (nach Forward & Lethbridge (2002) als Dokumentattribute bezeichnet) und zielgruppenunabhängige Merkmale (so genannte Erstellungsprozessattribute (Krawatzeck et al., 2011)) vorgenommen. Demnach zählen die Merkmale *Eindeutigkeit*, *Verständlichkeit* und *Vollständigkeit* in die Rubrik der zielgruppenabhängigen Attribute (Dokumentattribute), wohingegen die Qualitäts-

merkmale *Änderbarkeit*, *Aktualität*, *Identifizierbarkeit*, *Normkonformität* und *Widerspruchsfreiheit* unabhängig von individuellen Bedürfnissen sind.

Die zu dokumentierenden ETL-Prozesse beinhalten bereits einen großen Teil der für die Erstellung einer Dokumentation nötigen Informationen in Form von Quelltext (Forward & Lethbridge, 2002). Für den Bereich Data Warehouse Engineering angepasste modellgetriebene Softwareentwicklungsansätze (Kurze, 2011; Mazón & Trujillo, 2008) beinhalten zudem – über die im Quelltext enthaltenen Informationen hinaus – weitere für die Erstellung von nutzerspezifischer Dokumentation erforderliche Informationen, die aus Metadaten über die verschiedenen Anwendungen gewonnen und in maschinenlesbaren Modellen abgelegt werden. Durch die automatisierte Erzeugung von ETL-Dokumentation direkt aus den vorliegenden Modellen werden bis auf Normkonformität alle Erstellungsprozessattribute erfüllt (Krawatzek et al., 2011):

- *Änderbarkeit* und *Aktualität*: die Automatisierung erlaubt es, Systemänderungen jederzeit durch eine Neugenerierung der Dokumentation mit minimalem Personalaufwand zu veröffentlichen und damit aktuell zu halten,
- *Identifizierbarkeit*: durch die verfügbaren Metadaten ist es möglich, Zusammenhänge zwischen einzelnen Objekten, die im Quelltext u.U. nicht abbildbar sind, in der Dokumentation wiederzugeben und
- *Widerspruchsfreiheit*: durch die automatische Erzeugung von Dokumentation aus Modellen wird in der Dokumentation genau der Grad an Widerspruchsfreiheit abgebildet, der in den Modellen vorliegt.

Des Weiteren erlaubt die Verwendung eines generischen Dokumentenformatstandards (bspw. DITA (Darwin Information Typing Architecture) (DITA, o.J.) oder DocBook (DocBook, o.J.)) als zentrales Zwischenformat bei der Erstellung der Zielausgabeformate die Einhaltung der Struktur- und Inhaltsvorgaben der ISO-IEC 26514 für Benutzerdokumentation (International Organization for Standardization, 2008), und erfüllt damit das in (Krawatzek et al., 2011) nicht betrachtete Erstellungsprozessattribut Normkonformität.

Um die Dokumentation nutzerspezifisch zu generieren, müssen die bereits vorhandenen Modelle um eine Konfigurationsmöglichkeit bei der Verknüpfung von Informationen erweitert werden. Diese Konfiguration ist in das modellgetriebene Vorgehen eingebettet, was im Ergebnis dazu führt, dass sie ebenfalls in Form von Modellen vorliegt. Über diesen Konfigurationsansatz wird es möglich, die bereitgestellten Inhalte (beeinflusst die Dokumentattribute Vollständigkeit und Eindeutigkeit), das Layout und das Ausgabeformat (beeinflussen Verständlichkeit) der Dokumentation nutzerspezifisch anzupassen.

Die dargestellten Überlegungen und Designentscheidungen (Automatisierung durch Anwendung der modellgetriebenen Softwareentwicklung, Verwendung eines Dokumentenformatstandards und Schaffen von Konfigurationsmöglichkeiten) führten zum Entwurf eines Architektur-Frameworks (ein Modell im Sinne der DSR (March & Smith, 1995; Hevner et al., 2004)), welches die Architektur der gesuchten Lösung beschreibt (vgl. Abbildung 1). Für eine detaillierte Beschreibung der einzelnen Komponenten des Architektur-Frameworks sei der interessierte Leser auf (Jacobi et al., 2012) verwiesen.

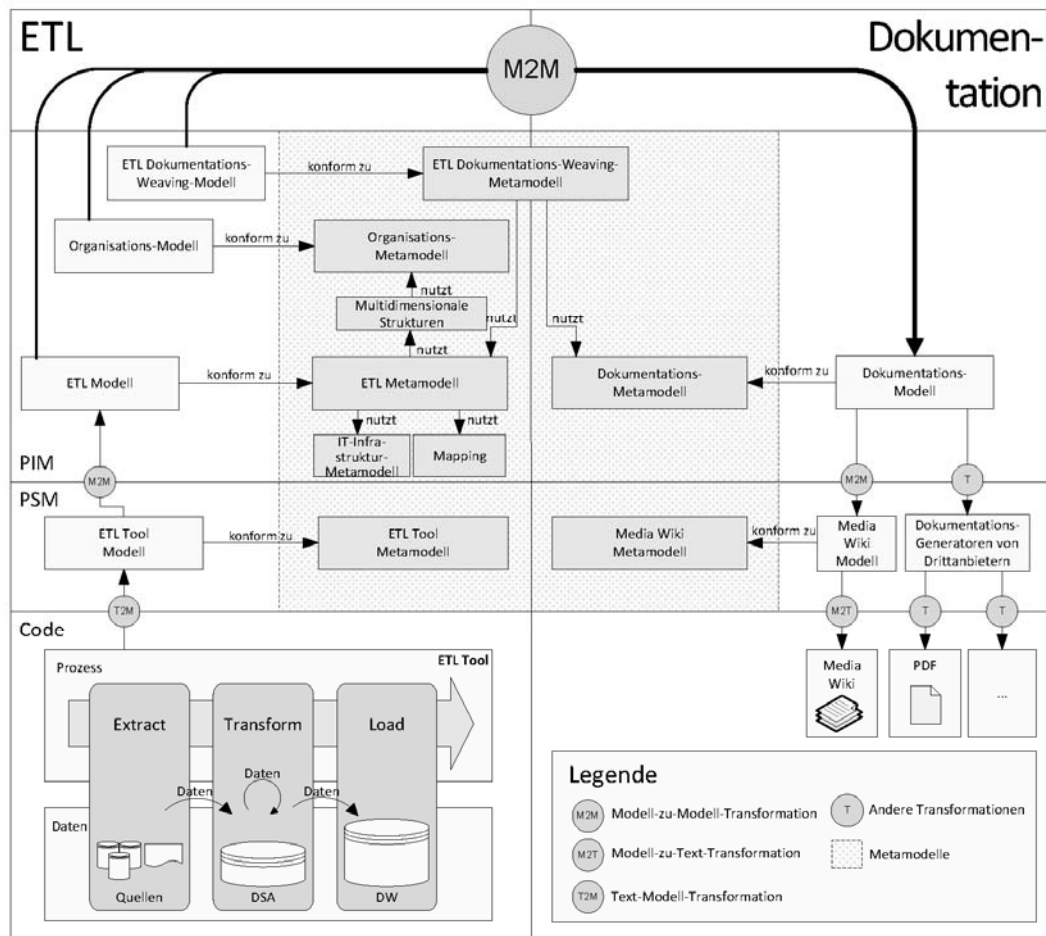


Abbildung 1: Architektur-Framework zur automatischen Erstellung nutzerspezifischer ETL-Dokumentation (Jacobi et al., 2012)

### Entwicklung der Lösung

Das in Abbildung 1 dargestellte Framework wurde mit Hilfe des Eclipse Modeling Frameworks in Form des Software-Prototypen „DW Documenter“ implementiert.

## 2.4 Aktivität 4: Demonstration

Zur Demonstration des DW Documenters wurden für verschiedene Plattformen kleinere, von Praxispartnern bereitgestellte ETL-Prozesse dokumentiert. So konnte nachgewiesen werden, dass die entworfene Lösung in der Lage ist (Krawatzek et al., 2012):

- ETL-Prozesse automatisiert zu dokumentieren,
- detaillierte Informationen über Variablen sowie deren Lese- und Schreibzugriffe anzugeben (vgl. Abbildung 2, links),
- detaillierte Informationen über die bearbeiteten Felder einer Aktivität anzugeben,
- vollständige Lineage- & Impact Analyse pro Feld (vgl. Abbildung 2, rechts) durchzuführen und
- verschiedene Ausgabeformate wie Wiki-Seiten und PDF-Dokumente (vgl. Abbildung 2 und 3) zu erzeugen.

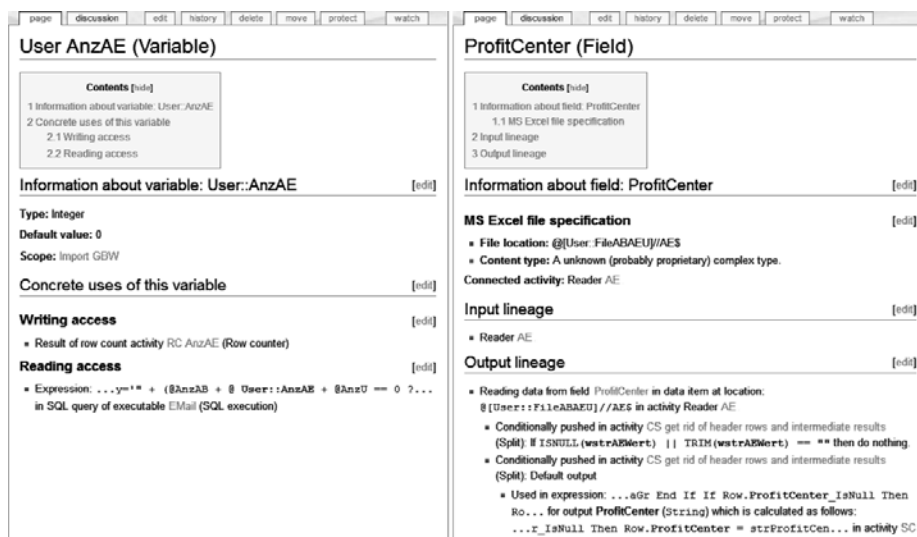


Abbildung 2: Mit DW Documenter erzeugte Dokumentation eines mit SSIS erstellten ETL-Prozesses im MediaWiki-Format (Krawatzek et al., 2012), (Hyperlinks sind heller dargestellt).

## 2.5 Aktivität 5: Evaluation

Eine Möglichkeit zur Evaluation von Artefakten besteht nach Hevner et al. (2004) darin, den erzeugten Nutzen der Artefakte mit dem Nutzen von anderen Artefakten, die dasselbe Problem lösen, zu vergleichen. Eine vergleichende Evaluation des DW Documenters (genauer: des entworfenen Architektur-Frameworks) mit anderen Softwareprodukten, welche im Rahmen der Analyse des Softwaremarktes identifiziert wurden (vgl. Aktivität

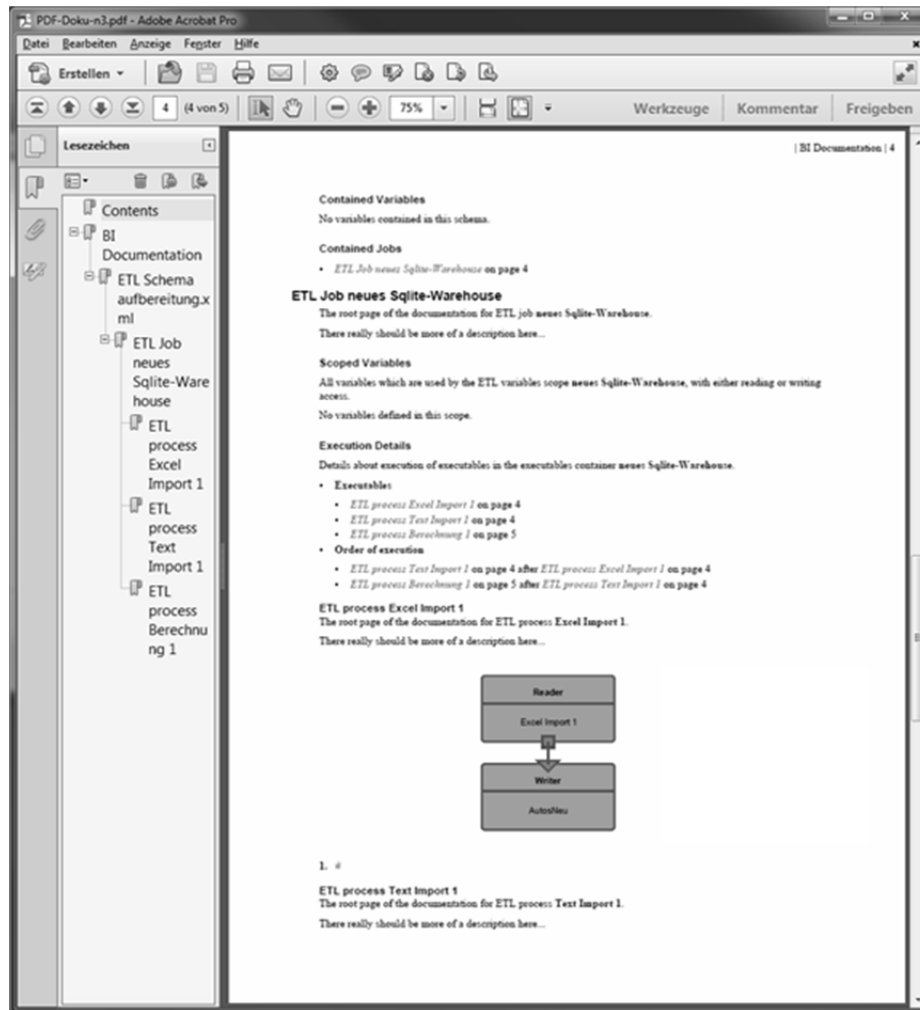


Abbildung 3: Mit DW Documenter erzeugte Dokumentation eines mit  $n^3$  DataWarehouseBuilder erstellten ETL-Prozesses im PDF-Format inklusive grafischer Notation für die logische ETL-Prozessbeschreibung (Hyperlinks sind heller dargestellt).

2), war nicht möglich. Zielführend war hingegen ein Vergleich der durch die Softwarelösungen erzeugten Zielartefakte (ETL-Dokumentation). Die Grundlage für den Vergleich bilden die in Aktivität 2 definierten und in Aktivität 3 spezifizierten Zielstellungen:

- I. Erstellung von kostengünstiger Dokumentation,
- II. Einhaltung der acht Merkmale von qualitativ hochwertiger Dokumentation (Dokument- und Erstellungsprozessattribute) und
- III. Bereitstellung von Möglichkeiten zur nutzerspezifischen Konfiguration.

Durch die Einschränkung auf automatisierte Lösungsansätze zur Dokumentationserstellung unterstützen alle zur Untersuchung herangezogenen Softwarewerkzeuge – inkl. des

DW Documenters – eine kostengünstige Erstellung von Dokumentation (Zielstellung I) und erfüllen die Erstellungsprozessattribute (erster Aspekt von Zielstellung II).

Allerdings konnte der DW Documenter durch eine tiefgehende Metadatenanalyse signifikante Verbesserungen im Bereich der Dokumentattribute (zweiter Aspekt von Zielstellung II) erzielen.

Weiterhin konnte festgestellt werden, dass der Prototyp DW Documenter als einziges Werkzeug die Möglichkeit zur nutzerspezifischen Konfiguration unterstützt (Zielstellung III). Durch diese lassen sich nicht nur die Bedürfnisse unterschiedlicher Zielgruppen abdecken, sondern auch nutzerspezifische Verbesserungen im Bereich der Dokumentattribute erzielen:

- Konfiguration des Inhalts (beeinflusst *Vollständigkeit* und *Eindeutigkeit*),
- Konfiguration des Layouts (beeinflusst *Verständlichkeit*) und
- Konfiguration des Ausgabeformats (beeinflusst *Verständlichkeit*).

Die dargestellte Argumentation bezüglich der Zielattribute macht deutlich, dass der DW Documenter den untersuchten Dokumentationslösungen im Bereich ETL überlegen ist.

## **2.6 Aktivität 6: Kommunikation der (Zwischen-)Ergebnisse**

Um der Bedeutung der einzelnen DSR-Aktivitäten in Hinblick auf ein relevantes Gesamtergebnis Rechnung zu tragen (Gregor & Hevner, 2013), wurden die Zwischenergebnisse wie folgt in der wissenschaftlichen Gemeinschaft publiziert und frühzeitig zur Diskussion gestellt:

- Aktivität 1: Gluchowski et al. (2011), Hofmann et al. (2012),
- 2 und 3: Jacobi et al. (2012) und
- Aktivität 4: Krawatzek et al. (2012).

Das vorliegende Paper fasst die Ergebnisse der DSR-Aktivitäten als Fallbeispiel zusammen und ergänzt die bisherigen Veröffentlichungen um die Aktivität 5 „Evaluation“.

Neben der Veröffentlichung innerhalb der wissenschaftlichen Gemeinschaft wurden die Projektergebnisse in Form eines Fachvortrags auf einer praxisorientierten Konferenz (Hofmann et al, 2011) sowie im Rahmen eines Messebesuches (CeBIT‘2012) der Praxis kommuniziert.

### 3 Kritische Diskussion

Die detaillierte Beschreibung des DSR-Fallbeispiels CAWE demonstriert, dass der DSR-Prozess nach Peffers et al. (2007) erfolgreich umgesetzt werden kann. Trotz erfolgreichem Durchlauf des DSR-Prozesses inkl. Kommunikation der Ergebnisse beinhaltet das CAWE-Beispiel Verbesserungspotential, welches im Folgenden diskutiert und in zukünftigen DSR-Projekten berücksichtigt werden kann.

Im Rahmen der dritten Aktivität „Entwurf einer Lösung“ sollte zur Darstellung von Modellen (betrifft im CAWE-Beispiel das Architektur-Framework) auf standardisierte Sprachen zurückgegriffen werden. Durch die Verwendung von standardisierten Sprachen wird ein gemeinsames Grundverständnis sichergestellt und die eigentliche Lösung kann auf einer standardisierten Basis (beispw. mit Gutachtern) diskutiert werden. Ist zur Darstellung eines spezifischen Problems bzw. einer spezifischen Lösung keine standardisierte Sprache vorhanden, sollten zunächst die dafür notwendigen Konstrukte (DSR-Artefakttyp) eigenständig entwickelt werden (identifizierte Forschungslücke; Anwendung von DSR möglich).

Findet die Demonstration von DSR-Ergebnissen in Zusammenarbeit mit Praxispartnern statt, empfiehlt es sich, diese als Interviews zu führen und zu dokumentieren. Zeigt sich im Rahmen dieser Demonstration, dass Zielstellungen noch verfeinert werden müssen, dienen die dokumentierten Gespräche als Nachweis für die notwendigen Anpassungen. Ferner können diese nach dem erneuten Durchlauf eines „Generate/Test Cycle“ (vgl. Hevner et al., 2004) als Motivation für das Publizieren von verbesserten DSR-Ergebnissen genutzt werden. Im Rahmen der Demonstration des DW Documenters wurde beispielsweise der Bedarf an Metadatenanalysen für Lineage- und Impact-Angaben identifiziert (verbessert „Vollständigkeit“ einer ETL-Dokumentation) und im nächsten Iterationsschritt im Architektur-Framework berücksichtigt. Die Publikation des verbesserten Frameworks war aufgrund fehlender Transparenz nicht möglich, da die Praxisanforderungen nicht dokumentiert wurden und somit nicht als Motivation für eine erneute Iteration angegeben werden konnten.

Des Weiteren empfiehlt es sich, bereits im Rahmen der Aktivität 2 Metriken zu definieren, welche eine Messung der aufgestellten Zielstellungen erlauben. Diese Metriken ermöglichen es, eine quantitative und somit eindeutige Evaluation (Aktivität 5) durchzuführen. Ein Ausweichen auf die im CAWE-Beispiel angewendete argumentative Evaluation ist somit hinfällig.

Abschließend sei auf die Publikation von Gregor & Hevner (2013) verwiesen, welche für die erfolgreiche Kommunikation von DSR-Ergebnissen (Aktivität 6) ein „DSR know-

ledge contribution framework“ sowie ein „DSR communication schema“ bereitstellt. Diese vielversprechenden Kommunikationshinweise wurden erst nach Abschluss des CAWE-Projektes veröffentlicht und konnten daher nicht berücksichtigt werden.

## 4 Anmerkungen

Das „Computer-Aided Data Warehouse Engineering“ (CAWE) Projekt, in dessen Rahmen der vorliegende Beitrag entstanden ist, wird mit Mitteln des ESF und des Freistaates Sachsen gefördert.

## 5 Literaturverzeichnis

- Chen, H.: Editorial: Design science, grand challenges, and societal impacts. *Journal ACM Transactions on Management Information Systems*. 2, 1 (2011).
- Chikofsky, E.J., Cross, J. H., I.: Reverse engineering and design recovery: a taxonomy. *IEEE Software*. 7, 1, pp. 13–17 (1990).
- DITA, <http://www.oasis-open.org/committees/dita>.
- DocBook, <http://www.oasis-open.org/docbook>.
- Forward, A., Lethbridge, T.C.: The Relevance of Software Documentation, Tools and Technologies: a Survey. In: *Proceedings of the 2002 ACM symposium on Document engineering - DocEng'02*. pp. 26–33, ACM Press, New York, USA (2002).
- Gluchowski, P., Hofmann, M., Jacobi, F., Krawatzek, R., Müller, A.: *Business-Intelligence-Umfrage 2011: Softwaregestütztes Lebenszyklusmanagement und aktuelles Dokumentationsgeschehen für Business-Intelligence-Systeme*. Technische Universität Chemnitz, Chemnitz, Deutschland (2011).
- Gluchowski, P., Kurze, C.: Modellierung und Dokumentation von BI-Systemen. *CONTROLLING - Zeitschrift für erfolgsorientierte Unternehmensplanung*. 22, pp. 676–682 (2010).
- Greenfield, L.: Don't let data warehousing gotchas getcha. *Datamation*, 42(5), pp. 76–77 (1996), zitiert nach: Inmon, B.: *The Data Warehouse Budget*. *DM Review Magazine*. p. 2 (1997).
- Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Research for Maximum Impact. *Management Information Systems Quarterly*. 37, 2, pp. 337–355, (2013).



- Hevner, A.R., March, S. T., Park, J., Ram, S.: Design Science in Information Systems Research. *Management Information Systems Quarterly*. 28, 1, pp. 75–105 (2004).
- Hofmann, M., Gluchowski, P., Jacobi, F., Kurze, C.: Computer-Aided Warehouse Engineering: Dokumentation und Modellierung komplexer Data-Warehouse-Systeme. Vortrag. 11. Europäische TDWI Konferenz, München, Deutschland (2011).
- Hofmann, M., Müller, A., Jacobi, F., Krawatzeck, R.: Umfrage 2011: „Dokumentation von Business-Intelligence-Systemen“ - Ergebnisse und Auswertung. In: Tagungsband der Multikonferenz Wirtschaftsinformatik 2012. pp. 1091–1104, GITO Verlag, Berlin (2012).
- International Organization for Standardization: Systems and software engineering - Requirements for designers and developers of user documentation (ISO-IEC 26514). (2008).
- Jacobi, F., Krawatzeck, R., Hofmann, M.: Meeting the Need for ETL Documentation: A Model-driven Framework for Customizable Documentation Generation. In: Proceedings of the Americas Conference on Information Systems. (AMCIS'12). Paper 23, Seattle, USA (2012).
- Krawatzeck, R, Jacobi, F., Müller, A., Hofmann, M.: Konzeption eines Frameworks zur automatisierten Erstellung nutzerspezifischer IT-Systemdokumentationen. In: Workshop Business Intelligence 2011 (WSBI'11) der GI-Fachgruppe Business Intelligence, Business Intelligence - Impulse für die Forschung oder Impulse durch die Forschung. pp. 15–26, CEUR Workshop Proceedings, Stuttgart (2011).
- Krawatzeck, R., Jacobi, F., Hofmann M.: CAWE DW Documenter: A Model-driven Tool for Customizable ETL Documentation Generation. In: Proceedings of the 31st International Conference on Conceptual Modeling (ER'2012). Florence, Italy (2012).
- Kurze, C.: Computer-Aided Warehouse Engineering: Anwendung modellgetriebener Entwicklungsparadigmen auf Data-Warehouse-Systeme. Verlag Dr. Kovač, Hamburg (2011).
- March, S. T., & Smith, G. F.: Design and natural science research on information technology. *Decision Support Systems*. 15, 4, 251–266 (1995).
- Mazón, J.-N., Trujillo, J.: An MDA approach for the development of data warehouses. *Decision Support Systems*. 45, 1, pp. 41–58 (2008).
- Österle, H., Becker, J., Frank, U., Hess, T.; Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E. J.: Memorandum on design-oriented information systems research. *European Journal of Information Systems*, 20, 1, pp. 7–10 (2010).

Peppers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*. 24, 3, pp. 45–77 (2007).

Wallmüller, E.: Die Rolle der Dokumentation in Software-Projekten. *Software-Qualitätsmanagement in der Praxis: Software-Qualität durch Führung und Verbesserung von Software-Prozessen*. pp. 149–156 Hanser Fachbuch (2001).

# **Business Intelligence Systems Optimization to Enable Better Self-Service Business Users**

*Safwan Sulaiman<sup>1</sup>, Jorge Marx Gómez<sup>1</sup>, Joachim Kurzhöfer<sup>2</sup>*

<sup>1</sup> *Very Large Business Applications / University of Oldenburg*

<sup>2</sup> *Lufthansa Systems*

## **Abstract**

*The success of the enterprise depends heavily on its decisions. Therefore, companies use Business Intelligence (BI) systems to support managers in their decision making process. However, the acceptance and usage of BI systems by end-users (business users) don't reach the expected goals. This is because of the high complexity and the irrelevance of delivered information. Business users can't use these systems to get the right information without relying on power users. The financial and time costs behind the communication between business and power users are high. The goal of this work is to develop a new BI architecture to reduce the complexity of using BI systems and optimize their usage by business users. The proposed solution offers suggestions to business users during their navigation in BI systems. Offering suggestions helps inexperienced business users in performing complex analysis. This is done by transferring the knowledge of power users to business users.*

## **1 Introduction**

The enterprises adopt Business Intelligence (BI) systems to assist the decision makers in their decision making process. They don't only assist this process, but also they ease and improve the overall management decisions (Laudon, Laudon, & Schoder, 2010, p. 736). Nowadays, the demands for information as a production factor increase. Based on (Chamoni & Gluchowski, 2010, p. 4), the reasons why these demands are inevitable include that the internal and the external conditions of the current economic life are rapidly changing and they are often very complex. According to (Ranjan, 2009), enterprises consider information as their second important resource after their people. BI systems provide enterprises with timely and accurate information, which allows them to make deci-

sions and react quickly on customer needs and market changes. Therefore, BI systems promote the enterprises that apply them with superiority above their other counterparts. (Bain & Company, 2011) stated that the effective decision-making processes are the key to the company's success. In other words, the success of any enterprise depends heavily on its decisions and BI systems, in this context, play an important role to support the information-based decisions.

In 2011, the report conducted by (Gartner, 2012) shows the worldwide sales of BI platforms, analytical applications and performance management software. The market volume exceeded 12 billion U.S. dollars. Alone, the sales of BI platforms amounted to more than 7.7 billion U.S. dollars, with an increase of 16.3% in 2011. According to this CIO-survey, the important reasons for this strong growth is that BI and analytical systems have the highest priority for CIOs (Gartner, 2012).

However, the wide applications of BI systems in enterprises have still some drawbacks and don't fully meet the requirements of their utilization. Due to the high complexity and irrelevance of the supplied information, less than 30% of potential users can fully benefit from BI systems (Kurzechner, 2011). Moreover, despite the large investments in BI systems, (Schmaltz, 2010, p. 2) shows in his dissertation that acceptance and usage of BI systems do not often reach the expected results. Additionally, the willingness today of the decision makers to attend advanced trainings for complex BI systems is very low (Knopf & Wortmann, 2011).

This paper suggests a new approach to optimize the usage of BI systems for wide variety of users based on transferring the knowledge from the skilled users to those who are less experienced. This is done by applying tracing methods on the skilled users to extract their analysis paths, which are used afterwards as recommendation to the less experienced ones.

The next section of this paper lists the basic foundations that are necessary to make the clear distinction of BI users and applications. Section 3 states the problem that the research behind this paper addresses. The followed research methodology is then explained in the fourth section. Details of the suggested approach are then explained in section 5. A list of related works to the proposed approach is then placed in section 6. Finally, section 7 concludes the main ideas presented in this paper and shows the potential future directions that can be derived from this research.

## 2 Basic Foundations

In order to have a common understanding of the used terms in this paper, the following two sections clarify the types of BI users, provide a classification of BI applications and explain the complexity of their usage.

### 2.1 Users of the BI Systems

In the literature, there are several contributions to classify BI users. On the one hand, (Gluchowski, Gabriel, & Dittmar, 2008, pp. 105 – 107) identified three user groups from the usage of BI systems perspective. These groups are information consumers, analysts and specialists. On the other hand, (Knopf & Wortmann, 2011, p. 29) distinguished three different user profiles based on their abilities of using BI systems. These profiles are user, power user and analyst. Finally, based on style of interacting with information for decision making, (Eckerson, 2011) classified BI users into power users and casual users. This paper tries to make a consensus among the aforementioned classifications of BI users. Thereby, two factors are taken into account to classify BI users. These factors are the user ability of using BI systems and the user's informative behavior<sup>1</sup>. Accordingly, two types of BI users are identified:

1. *Power Users or Information Producers*: These are business analysts, analytical modelers and IT professionals. They have the ability to generate reports, analyze data and perform flexible navigation options in the multidimensional data models (Gluchowski et al., 2008, p. 106). Such users are considered as information producers and they generate information that can be used either for their own decision or for the decisions of business users.
2. *Business Users or Information Consumers*: These are executives, managers and operations staff. They are considered more as information consumers. The main source of their information is the information provided by the power users. They consume this information to make their decisions. Furthermore, these users can only use the predefined standard reports or dashboards that don't require any technological or methodological knowledge.

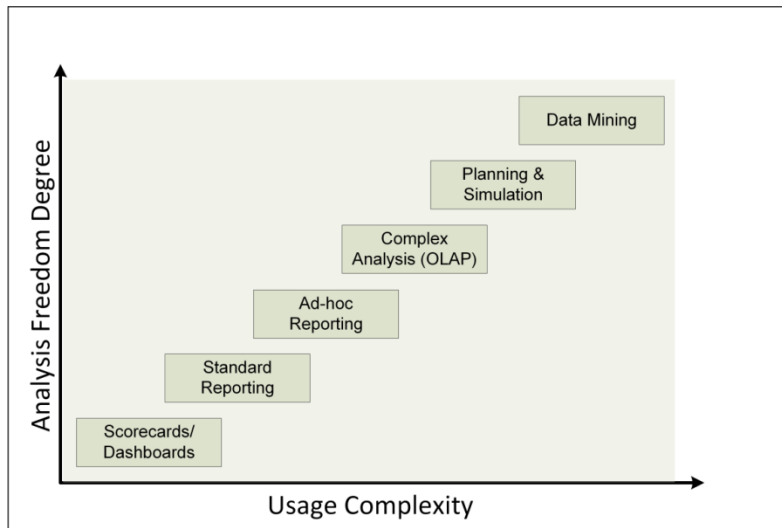
The main argument behind the aforementioned classification of BI Users is that the issues that are researched in this work are directly or indirectly related either to the power or to the business users. Based on that, we will propose an approach to transfer the power users' knowledge (skilled users) to the business users (unskilled users).

---

<sup>1</sup> What is meant by user's informative behavior is how BI user handles the information? The approach behind this paper answers this question and classifies BI users as either information consumer or information producer.

## 2.2 Business Intelligence Tools

This section gives short overview of the key BI tools. These tools cover a set of diverse requirements for different sets of user groups. The different types of these BI tools include scorecards/dashboards, ad-hoc reporting, complex analysis (OLAP) and data mining. Figure 1 illustrates the relationship between the user's analysis freedom degree and the usage complexity of these BI tools.



*Figure 1: Classes of Business Intelligence Applications (Tools) (Based on Figure 2 in (Bange, 2010, p. 141))*

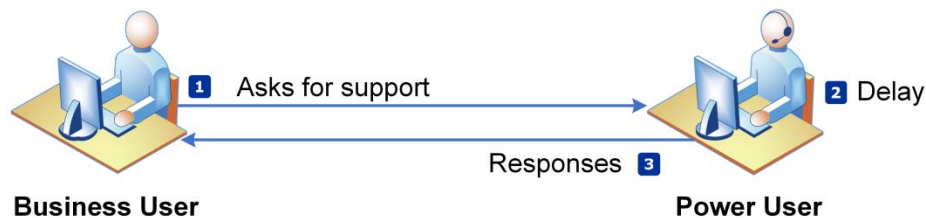
(Bange, 2010) indicated that the application usage complexity grows with the analysis freedom degree given to the user. Thus, business users can only use dashboards and standard reports, since its usage complexity is relatively lower than other BI tools. Consequently, business users are limited in their analysis freedom degree. In contrast to that, power users have the ability to use the different kinds of BI tools with higher usage complexity and more analysis freedom degree. Examples of such tools include complex analysis (OLAP) and data mining.

## 3 Research Statement

One of the main strategic business goals of any enterprise is to have a reliable reporting platform in which quality data are provided for analysis and decision making purposes. Having such platforms in production helps enterprises in overcoming traditional challenges in data integration among heterogeneous information systems and in dealing with data redundancy. Overcoming such challenges by the employment of one of these new reporting platforms makes this latter more complex due to data and application inter-

dependencies (Kulkarni, 2012). As observed by (Mertens & Krahn, 2012), the flexibility and powerfulness of BI systems let business users still face significant difficulties in carrying out ad-hoc analysis reports.

Business users normally lack crucial information to take proper decisions while using BI tools. This drawback comes from the missed or incomplete methodological and technical knowledge in using these tools. To help business users to overcome such challenge, the support of power users is needed in many cases. Figure 2 depicts the major communications between a business user and a power user. If a business user needs support to perform a specific task, which he can't handle, he sends a request to the power user. In turn, the power user processes the requests, however, with some time delay and then sends the result back to business users.



*Figure 2: Communication between business users and power user*

Based on discussions with experts, managers, and industrial stakeholders we have discovered different scenarios after performing the communications in the figure above. These scenarios include:

- Business users realize that the information sent by power users does not exactly meet their expectations.
- Business users don't understand some of the results' values.
- Business users require additional information about a specific department or organizational unit.

It is noticed that in these three scenarios, business users in most cases should send again new requests to the power users acquiring new information or asking for support to provide more meaning for a subset of the information they received. Consequently, such communication among business and power users can occur in a repeated manner or in iterations and this creates considerable extra overhead from the power users.

These scenarios have been presented again to some experts, managers in the German Lufthansa Systems company. The BI experts in this company stated that in every department or organization unit, there are one or two power users, who are responsible of supporting many business users. As a result, each power user will be overloaded with a large number of requests coming from many business users.

### **High Communication Costs**

The above-described communications have high costs on enterprises. These costs can be classified into the following two types:

- *Time costs:* Business user should always wait for processing his request by one of the busy power users. On the one hand, it is always crucial for the managers to get the information they need in the right time during the decision making process (Spahn, Kleb, Grimm, & Scheidl, 2008). On the other hand, power users must interrupt their own work to process the requests of business users instead of focusing on activities that are more valuable. Examples of such activities include developing new applications, expanding data in data warehouses from existing and new sources, improving data quality processing, or incorporating new technologies to improve performance (Imhoff & White, 2011).
- *Financial costs:* While the time runs, the value of any decision decreases. This loss can be expressed in companies in term of money. Sales will be reduced, and more cost in human resources or logistics will increase. Moreover, enterprises need excessive number of power users to cover the large amount of business users' requests. Such situations make pressure on enterprises' managers to increase their IT budget accordingly.

This kind of cost classification is just to clarify how different kinds of costs might arise. At the end, every kind of costs is going to be paid in any enterprise in form of money.

Based on the previous discussion about the communication between business and power users and the accompanying costs, enterprises need to react to such situation by reducing the number of communications between their business and power users as much as possible. This requires more innovative alternatives to the actual existing ones. The proposed approach tries to overcome this challenge by minimizing the communication between business and power users. This will indirectly decrease the costs by relying on the harvested knowledge of the existing power users. Another added value in applying such new approach is to provide business users with the information they need in a timely manner and without having a direct connection with power users (Spahn et al., 2008). One of the central requirements of business managers, as explained in (Knopf & Wortmann, 2011), is the analysis flexibility. However, as discussed before in the second section of this paper, this flexibility often leads to more usage complexity. Relying on the concepts and



ideas of self-service BI (Imhoff & White, 2011), the proposed approach tries also to encounter such situation by empowering business users with necessary knowledge to perform their analysis and create their reports. The focus here is to move business users to a higher layer in the BI's analysis freedom degree (like OLAP in Figure 1). This is done by promoting more self-service business users with minimal interactions with power users.

The major objective of this work is the conception and development of an enhanced BI system in which the knowledge of power users is extracted and transferred to business users. This new BI system should offer suggestions to business users to help them in their analysis. This work will concentrate on OLAP as a complex analysis application<sup>2</sup>.

## 4 Research Methodology

Our research follows the design science research, which has its roots in engineering science. In contrast to behavioral science research that seeks to develop and justify theories that explain human or organizational behavior, design science is fundamentally a problem-solving paradigm. It seeks extending the boundaries of human and organizational capabilities by creating new and innovative artifacts (Hevner, March, Park, & Ram, 2004). Business Intelligence as a research field of Business Information Systems is especially marked in Germany with a strong practical relevance as well as a design-oriented research discipline (Baars, 2011).

The boundary of our research will be explained based on the information systems research framework of (Hevner et al., 2004). In this framework, the problem space is defined in the environment, which includes people, organizations and their existing and planned technologies. In this context, both BI business and power users represent the different roles of the people in any organization or enterprise that use BI systems as a decision support technology. The business need, which is defined in Section 3, assures the relevance of our research. Design science addresses research through the building and evaluation of artifacts designed to meet the identified business needs. The enhanced architecture of BI Systems is considered as the resulted artifact of the research. The research rigor is achieved by appropriately applying foundations and methodologies from the existing knowledge base. To evaluate this artifact, the proposed architecture should be implemented as a proof of the concept. The last phase is the evaluation of the results prototypical implementation using some evaluation methods like the evaluation research

---

<sup>2</sup> OLAP stands for Online Analytical Processing. It is widely used in enterprises as part of business intelligence and data warehousing suits (Chamoni & Gluchowski, 2010, p. 199).

method of (Österle et al., 2010). This phase is realized via discussions with BI experts from different organizations.

## 5 Enhanced BI Architecture

This section explains the proposed approach in details. The main idea is the conception and development of a new architecture for BI systems. This architecture extends the typical existing architectures. New components are added to implement new functionalities that are not available in the existing architectures. Figure 3 illustrates the proposed architecture of BI systems.

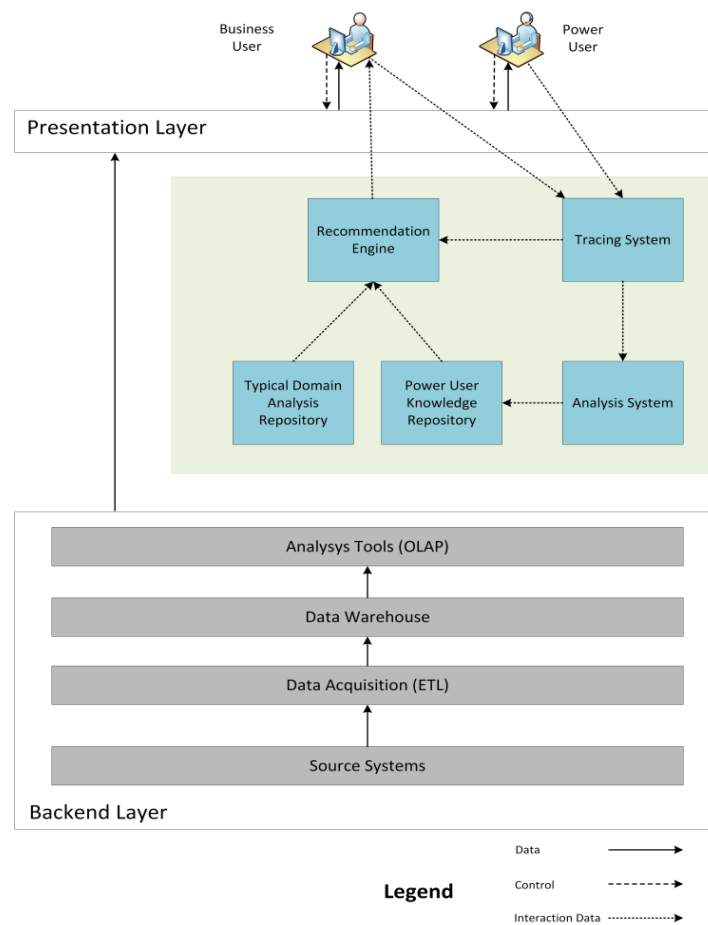


Figure 3: Enhanced BI Architecture

In the following, the main components of the depicted architecture and the main interactions between them are described. Besides the typical BI architecture's components:

source systems, ETL<sup>3</sup>, data warehouse, analysis tools and presentation layer the enhanced architecture has the following new components:

- *Tracing System*: This component is responsible of tracing the actions of the power and business users while they use a BI system. The power user's actions should be stored in a proper storage medium like a database or log file. Two important points must be taken into consideration in the design time of this component. Firstly, after applying a proper observation on power users' actions, what are the criteria to select specific actions to extract the knowledge from power users? Secondly, the representation of the actions in form of storage shall be identified to enable the application of a proper knowledge extraction algorithm to power users' actions. This data representation includes basic information, such as timestamp, unique identifier, and name or description of the action.
- *Analysis System*: This component applies a set of algorithms to the tracing data collected by the tracing system. The output of these algorithms is specific patterns that represent the knowledge of the power user (based on a predefined business needs). Each pattern represents one analysis path<sup>4</sup>. In this way, the knowledge of each power user can be collected after aggregating all his/her possible analysis paths. To do so, the intention is to use time series analysis or to develop new algorithms, if the result of the time series algorithms is not appropriate.
- *Power User Knowledge Repository*: This component stores the power users' analysis paths extracted by the analysis system. To enhance the quality of the analysis paths, they should be classified into active or inactive analysis paths based on their repetition. The analysis path is considered active, if the power user repeats it for at least five times. Otherwise, the analysis path is considered inactive and it will be ignored as long as it is not repeated more than five times.
- *Typical Domain Analysis Repository*: This component is responsible of providing recommendations to business users regarding a specific domain if the tracing system component detects information related of such domain while the user executes some actions in the BI tools. Typical domain example is the human resources domain that includes information about salaries, training costs, travelling costs, etc.

---

<sup>3</sup> ETL stands for "Extraction, Transformation, and Loading". The ETL process is the sequence of applications that extract data sets from the various sources, bring them to a data staging area, apply a sequence of processes to prepare the data for migration into the data warehouse, and actually load them (Loshin, 2012).

<sup>4</sup> Analysis path in this paper represents a specific sequence of steps accomplished by the user while interacting with a BI-System.

- *Recommendation Engine:* This component has the responsibility of offering suggestions to business users based on the stored actions in the power user's knowledge repository and typical domain analysis repository. This component has an interface with the tracing system to get information about the last steps done by the business user (the system consider now just the last three steps). Then it compares these steps with the stored analysis paths in the power user's knowledge repository. The result of this comparison should lead to one analysis path. Based on that, the recommendation system should offer the business user suggestions following the found analysis path. These suggestions will help the business user to advance in performing the analysis.

## 6 Related Works

It was explained in the article of (Baars, 2006) how to distribute BI knowledge. The analysis result and templates should be accessed from other users in the enterprise using knowledge management systems. Analysis result can avoid the double works by the same information need. Analysis template includes basic information that enables BI users performing the analysis in any context. This approach has several technical and organizational challenges. These include the requirement of combining different interfaces and formats. However, this work lacks the need to motivate users to explain and distribute her/his skills to the knowledge management system.

The approach of (Mertens & Krahn, 2012) "Knowledge based business intelligence for business user information self-service" is based on a semantic metadata layer which is capable to import and manage modeled semantic metadata. The provided metadata is supposed to be used for further analysis in order to allow the self-service business user. This approach requires an explicit deriving and modeling of analysis and domain knowledge of experts and power users. Then this knowledge has to be imported to the Analytical Information System.

### *Advantages of the proposed BI architecture over existing work:*

The first advantage of our approach is the automatic extracting of power users' knowledge in which the knowledge repository sub-system of the proposed architecture is kept always up-to-date. This knowledge extraction process was done in some of other related works like (Mertens et al.) in a manual manner. In addition, there will be minimal dependencies among business and power users. The second advantage is the real time offering of suggestions to business users. In many cases, business users are stuck in processing complex analysis in BI tools and they are in a need for some aid by power users. Embedding the knowledge of power users in the knowledge repository will enhance the

overall BI architecture to provide this knowledge in form of suggestions. This will decrease the overhead on the power users and expose implicitly their knowledge to the business users to perform complex analyses independently.

## 7 Conclusions and future work

In this paper, the focus was to introduce a new BI architecture that enables business users in getting information and performing complex analysis without interacting with power users. This is done by extracting the knowledge of power users by applying tracing method on their actions while using the BI tools. This knowledge is then offered, in form of suggestions, to business users while they try to perform complex analysis.

In our future works, the provided suggestions to the business user will be evaluated. Besides that, we will consider the issue of refining the proposed architecture. Moreover, as a proof of concept, a prototype will be implemented to show the practicability of the overall concept. For this purpose, we will concentrate on open Source BI systems to extend it to conform to the objectives of this work.

## References

- Baars, H. (2006). Distribution von Business-Intelligence-Wissen. In *Analytische Informationssysteme* (pp. 409–424). Berlin: Springer.
- Baars, H. (2011). Impulse für die Forschung oder Impulse durch die Forschung: Entwicklungsperspektiven der BI-Forschung in Deutschland. Presented at the 3. Workshop Business Intelligence der GI-Fachgruppe WI-BI.
- Bain & Company. (2011). Effektive Entscheidungsprozesse sind Schlüssel für den Unternehmenserfolg. Retrieved July 17, 2012, from [http://www.process.vogel.de/management\\_und\\_it/branchen\\_maerkte/marktbarometer/articles/301345/](http://www.process.vogel.de/management_und_it/branchen_maerkte/marktbarometer/articles/301345/)
- Bange, C. (2010). Werkzeuge für analytische Informationssysteme. *Analytische Informationssysteme*, 4, 131–156.
- Chamoni, P., & Gluchowski, P. (2010). *Analytische Informationssysteme* (4th ed.). Springer.
- Eckerson, W. (2011). The Secrets of Self-Service BI. Wayne Eckerson - BeyeNETWORK. Retrieved from [http://www.b-eye-network.com/blogs/eckerson/archives/2011/01/the\\_secrets\\_of.php](http://www.b-eye-network.com/blogs/eckerson/archives/2011/01/the_secrets_of.php)

- Gartner. (2012). Market Share: All Software Markets, Worldwide, 2011. Gartner Business Intelligence Summit. Retrieved from <http://www.gartner.com/it/page.jsp?id=1971516>
- Gluchowski, P., Gabriel, R., & Dittmar, C. (2008). Management Support Systeme und Business Intelligence (2nd ed.). Berlin Heidelberg: Springer.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Mis Quarterly*, 28(1), 75–105.
- Imhoff, C., & White, C. (2011). Self-Service Business Intelligence. TDWI Best Practices Report, Third Quarter.
- Knopf, S., & Wortmann, F. (2011). Frontends für Führungskräfte - endlich zielgruppentauglich? *Zeitschrift Hmd-Praxis Der Wirtschaftsinformatik*, 48, 282.
- Kulkarni, N. (2012). Information Management: Embrace the Future of BI: Self Service. Retrieved August 9, 2012, from <http://license.icopyright.net/user/viewFreeUse.act?fuid=MTY0OTAyMjA=>
- Kurzlechner, W. (2011). 3 Ebenen der Konsumerisierung: Gartner: BI vor radikalem Wandel. Retrieved July 19, 2012, from [www.cio.de/knowledgecenter/bi/2283277/](http://www.cio.de/knowledgecenter/bi/2283277/)
- Laudon, K. C., Laudon, J. P., & Schoder, D. (2010). *Wirtschaftsinformatik: Eine Einführung*. Pearson Education.
- Loshin, D. (2012). *Business intelligence: the savvy manager's guide*. Morgan Kaufmann Pub.
- Mertens, M., & Krahn, T. (2012). Knowledge Based Business Intelligence for Business User Information Self-Service. In S. Brüggemann & C. d' Amato (Eds.), *Collaboration and the Semantic Web: Social Networks, Knowledge Networks, and Knowledge Resources* (pp. 271–296). IGI Global.
- Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., ... Sinz, E. J. (2010). Memorandum on design-oriented information systems research. *European Journal of Information Systems*, 20(1), 7–10.
- Ranjan, J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9(1), 60–70.
- Schmaltz, M. A. (2010). *Methode zur Messung und Steigerung der individuellen Akzeptanz von Informationslogistik in Unternehmen (DISSERTATION)*. Universität St. Gallen.
- Spahn, M., Kleb, J., Grimm, S., & Scheidl, S. (2008). Supporting business intelligence by providing ontology-based end-user information self-service. In *Proceedings of the First international Workshop on ontology-Supported Business intelligence* (p. 10). Retrieved from <http://dl.acm.org/citation.cfm?id=1452567.1452577>

# **Zur methodischen Vorbereitung von Data-Mining-Projekten unter Verwendung von CRISP-DM im Kontext diskreter Produktionsprozesse**

*Uwe Wieland, Marco Fischer*

*Technische Universität Dresden  
Lehrstuhl für Wirtschaftsinformatik  
Business Intelligence Research*

## **Abstract**

*Die Analyse von Produktionsprozessen innerhalb von Data-Mining-Projekten stellt einen hohen Anspruch an die interdisziplinäre Zusammenarbeit zwischen Domänen- und Data-Mining-Experten. Die vorgeschlagene modellgestützte Methode offeriert dazu einen ersten Vorschlag, wie real-weltliche Produktionsprozesse in eine von analytischen Verfahren geprägte Welt überführt werden können. Dazu werden die Anforderungen aus beiden Bereichen sowie ein Standardvorgehen für diesen Diskursbereich in einem Modell integriert, dessen Erstellung nachfolgend erläutert wird.*

## **1 Problemstellung und Motivation**

Die industrielle Wertschöpfung erfolgt in fortwährend komplexer werdenden Produktionsprozessen, welche oft durch sehr vielschichtige Ursache-Wirkungsbeziehungen charakterisiert sind und hinsichtlich ihrer Planung, Durchführung, Steuerung und Kontrolle von umfangreichem Expertenwissen abhängig sind (Wiedenmann, 2001, S. 30 f.) Durch die zunehmende Verbesserung und Verbreitung von Sensorik und Aktuatorik steigert sich die Leistungsfähigkeit von autonomen Produktionssystemen, welche zukünftig selbständig miteinander durch Datennetze kommunizieren, entscheiden und agieren sollen.

Neben der geplanten technologischen Verbesserung im Prozessablauf stellt die enorme Menge an erzeugten Prozessdaten und das Management dieser Daten (Erhebung, Analyse und Verarbeitung) bereits heute eine zentrale Herausforderung, aber auch ein hohes Potential dar. Prozessdaten enthalten historisierte, wettbewerbsrelevante Informationen, welche die Charakteristik von Prozessen abbilden und daher neben dem menschlichen Expertenwissen einen gleichbedeutenden Stellenwert besitzen. Erklärungsmodelle sollen

dabei helfen, solche sozio-technische Systeme zu verstehen und anschließend zu verbessern (Kagermann, Wahlster & Helbig, 2013, S. 46 f.) Dieser Beitrag beschäftigt sich mit der methodischen Vorbereitung – Untersuchung der Geschäftsziele und Datenvorverarbeitung - von Data-Mining-Projekten zum Aufbau derartiger Erklärungsmodelle für diskrete Produktionsprozesse, welche über einen mächtigen Prozessdatenbestand verfügen.

Die Vorbereitung und Durchführung von Data-Mining-Projekten innerhalb von diskreten Produktionsprozessen stellt einen hohen Anspruch an die interdisziplinäre Zusammenarbeit zwischen Domänen- und Data-Mining-Experten. Dabei kommt es nicht allein auf die Daten an, sondern auch die Beschreibung der Daten und die untersuchte Domäne sind von großer Bedeutung für den Erfolg künftiger Projekte (Lukasz, Musilek, 2006, S. 19; Marban et al., 2007, S. 97 ff.; Mariscal, 2013, S. 160 ff.; Sharma & Osei-Bryson, 2009 S. 4114 ff.)

Mit CRISP-DM (Cross Industry Standard Process for Data Mining) basiert dieser Beitrag auf einem der meist verwendeten Vorgehen zur Durchführung von Data-Mining-Projekten, welches besonders im Anwendungsbereich der Industrie zu finden ist [Mariscal, 2010, pp. 139) Gemessen an der Evolution von Data-Mining-Vorgehensmodellen bildet CRISP-DM zum einen die Vereinigung von bereits sehr etablierten Vorgehen wie dem KDD-Prozess und industriellen Ansätzen (z.B. SEMMA) und zum anderen dient es als Ausgangspunkt für neue Ansätze (z.B. Cios et al. 2005, CRSIP-DM 2.0) (Azevedo, Santos, S. 185; Mariscal, 2010, S. 142). Das Referenzvorgehen CRISP-DM definiert und beschreibt pro Phase einzelne generische Aufgaben unabhängig vom Anwendungsbereich sowie den verwendeten Technologien, um Data-Mining-Projekte systematisiert durchführen zu können. Das Benutzerhandbuch als inhaltliche Erweiterung gibt ausführliche Tipps und Hinweise zu den einzelnen Phasen und deren Aufgaben (IBM, 2010, S. 3 f.; Lukasz & Musilek, S. 5). Die Lösung der Aufgaben ist jedoch von der jeweiligen Situation abhängig. Situationen werden durch einen Kontext definiert, welcher durch die Anwendungsdomäne und weitere Faktoren charakterisiert wird. CRISP-DM liefert ausschließlich ein sehr abstraktes Vorgehen für eine Zuordnung des generischen Modells auf konkrete Anwendungsbereiche (IBM, 2010, S. 4; Mariscal, 2010, S. 139). Zusammenfassend können daher folgende Begründungen für eine Konkretisierung (B) konstatiert werden:

- B1: Eine situationsbezogene Konkretisierung ist methodisch nicht gewährleistet.
- B2: Die Ermittlung der relevanten Datenquellen für eine Analyse bleibt sehr vage und isoliert vom eigentlichen Analyseobjekt.
- B3: Die Integration von Rollen (z. B. Domänen- u. Data-Mining-Experte) in das Vorgehen wird nicht geregelt.



- B4: Eindeutige Zusammenhänge und Abhängigkeiten zwischen den jeweiligen Ergebnissen der einzelnen Phasen werden außer Acht gelassen.
- B5: Ergebnisse werden in ihrer Form nicht spezifiziert und sind damit nur schwer wiederverwendbar.
- B6: Die Wiederverwendbarkeit von vorbereitenden Teilergebnissen bei unterschiedlichen Analyseverfahren ist nicht gegeben.

Bezugnehmend auf die Problemstellung und Motivation wird das Standardvorgehen CRISP-DM in ausgewählten Punkten der Analysevorbereitung spezialisiert und wiederverwendbar für diskrete Produktionsprozesse angepasst werden, um die interdisziplinäre Zusammenarbeit und damit den Aufbau von prozessspezifischen Erklärungsmodellen zu unterstützen.

## 2 Forschungsdesign

Das momentan etablierte Standardvorgehen für Data Mining CRISP-DM weist Konkretisierungsbedarf bezüglich einer Anwendung auf die Analyse von Prozessdaten diskreter Produktionsprozesse auf. Gerade in den Vorbereitungsphasen werden zahlreiche generische Aufgaben und Ergebnisse beschrieben, ohne eine konkrete Form zu definieren und eine Wiederverwendung zu adressieren. Entsprechend Design Science Research soll für die Lösung dieses Problems ein Artefakt in Form einer modellgestützten Methode konstruiert werden.

Aus der Problembeschreibung leitet sich die Notwendigkeit ab, die Vorbereitung von Data-Mining-Projekten im Kontext diskreter Produktionsprozesse methodisch zu unterstützen. Methoden werden dabei allgemein als Vorschriften bzw. Handlungsempfehlungen für Problemlösungen verstanden (Weller, 2010, S.36). Da sich Modelle als ein wichtiges Instrument der Erkenntnisgewinnung, Kommunikation und Dokumentation etabliert haben (Wand & Weber, 2002, S. 363), soll die methodische Unterstützung durch die systematische Verwendung von Modellen erfolgen. Damit lässt sich das zu erstellende Artefakt den modellgestützten Methoden zuordnen. Der Artikel verfolgt damit das Gestaltungsziel, eine modellgestützte Methode zur Vorbereitung von Data-Mining-Analysen im Kontext diskreter Produktionsprozesse zu entwickeln. Resultierend aus den vorangegangenen Betrachtungen ergeben sich für den Beitrag die folgenden Forschungsfragen:

- Welche Anforderungen bestehen an die Vorbereitungsphasen von Data-Mining-Projekten in diskreten Produktionsprozessen?
- Wie muss eine modellgestützte Methode gestaltet werden, um die Anforderungen zu erfüllen?

Zunächst werden die Anforderungen an diskrete Produktionsprozesse und die Datenvorverarbeitung in Data-Mining-Projekten beschrieben, die literaturgestützt erhoben worden und als Grundlage für die Konstruktion der modellgestützten Methode dienen.

Anschließend werden die Konstruktionsergebnisse präsentiert. Eine modellgestützte Methode besteht dabei stets aus einer Prozessbeschreibung, welche neben der Erstellung eines Modells, dessen Transformation und Nutzung beschreibt. Das konstruierte Modell muss konform zur verwendeten Modellierungssprache sein, die gemäß einer Sprachbeschreibung definiert ist [Weller, 2010, S. 42]. Der Beitrag führt als Sprachbeschreibung ein Prozessmetamodell ein, das alle für die Prozessdatenanalyse relevanten Komponenten, deren Beziehungen und Regeln definiert. Mit Hilfe des Prozessmetamodells wird der Produktionsprozess im Ist-Zustand modelliert. Neben der Ablaufstruktur des Prozesses ist speziell die Strukturierung von Prozessdaten sowie die Abbildung existierender Abhängigkeiten und vermuteter Ursache-Wirkungsbeziehungen das Ziel der Modellierung, welche zunächst unabhängig von den Data-Mining-Zielen sowie den eingesetzten Verfahren ist. Der so modellierte Prozess kann mit Hilfe der Methode in ein prozessspezifisches Datenmodell transformiert werden, aus dem wiederum analysespezifische Falldatensätze abgeleitet werden können, welche die Grundlage für die Anwendung konkreter Data-Mining-Verfahren darstellen. Der Teil Modelltransformation und -nutzung ist in Kapitel 4.2 als Research in Progress gekennzeichnet.

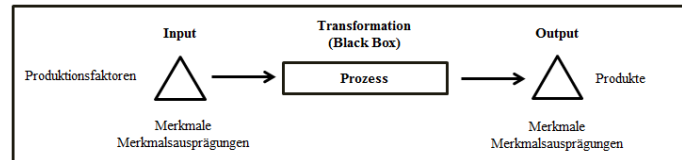
Abschließend erfolgt eine merkmalsbasierte Evaluation, die nachweist, wie die konstruierten Artefakte der modellgestützte Methode die Vorbereitungsphasen des Data Mining nach CRISP-DM unterstützt.

### **3 Anforderungen**

#### **3.1 Anforderungen diskreter Produktionsprozesse**

Innerhalb der betrieblichen Leistungserstellung bildet die Produktion einen Funktionsbereich, in dem Produktionsfaktoren miteinander kombiniert werden, so dass Endprodukte entstehen. Dabei wird die Produktion als Transformationsprozess angesehen. „Gegenstand der Produktion ist die Kombination und Transformation von Produktionsfaktoren (Input), so dass ein bestimmter Zweck (Output), das sogenannte Sachziel (z.B. die Herstellung von Automobilen), unter Beachtung des Formalziels (z.B. Gewinnmaximierung) bestmöglich erreicht wird. Die Ergebnisse des Transformationsprozesses sind die für den Absatzmarkt bzw. für weitere Transformationsprozesse bestimmten Güter oder Dienstleistungen, die nach ihrem Verwendungszweck als (End- oder Zwischen-) Produkte bezeichnet werden.“ (Kiener et al., 2012, S. 5). DANGELMAIER spricht davon, dass die

Inputs und Outputs eines Prozesses durch ihre relevanten Merkmalsausprägungen charakterisiert sind und sich darüber die Relation zwischen den einzelnen Inputs und Outputs herstellen lassen (Dangelmaier, 2009, S. 2ff.).



**Abbildung 1: Input-Output-System von Produktionsprozessen [Da09], S. 3**

Derartige Transformationsprozesse werden in der Regel als Input-Output-System beschrieben, welche durch eine Betrachtung der reinen Außensicht charakterisiert sind (Anforderung P1). Damit wird der Input-Output-Prozess wie in Abbildung 1 durch ein einziges Element mit dem nach außen wirksamen Objekten und deren Merkmalen beschrieben. Die Funktionalität des Transformationsschrittes bleibt dabei verborgen (Black-Box) und erfolgt entweder deterministisch oder stochastisch (Dangelmaier, 2009, S. 3). Produktionsprozesse können in einstufige und mehrstufige Prozesse unterschieden werden. Dabei bestehen die mehrstufigen Produktionsprozesse aus parallel oder sukzessiv ablaufenden einstufigen Produktionsprozessen. Somit stellen einstufige Produktionsprozesse stets eine Spezialisierung dar, welche innerhalb von komplexen Produktionsszenarien, zu mehrstufigen Produktionsprozessen kombiniert werden, was im Hinblick auf das Input-Output-System eine Komposition der einzelnen Elemente bedeutet. Innerhalb einer mehrstufigen Prozesskette wird auf die finalen Merkmale des Fertigungsobjektes hingearbeitet. Dazu müssen die Merkmalsänderungen an jedem Prozessschritt beherrschbar und transparent sein (Großmann & Wiemer, 2010, S. 856). Das Ziel eines jeden Produktionsprozesses ist ein reproduzierbarer Ablauf (Transformation), welcher stets zu einem konstanten Output führt (Weller, 2010, S. 70 f.; Wiedenmann, 2001, S. 27). Dafür sind gemäß GROSSMANN/WIEMER folgende Bedingungen zu erfüllen:

**Tabelle 1: Bedingungen einer reproduzierbaren Fertigung nach Großmann & Wiemer, S. 855f.**

Abk.	Bedingung	Beschreibung
P2	Definiertheit der finalen Produkteigenschaften	Festlegung des Fertigungsziels mit definierten Produkteigenschaften und dessen Toleranzgrenzen

P3	Durchgängigkeit der gesamten Prozesskette	Die Fertigung verläuft entlang einer durchgängigen Prozesskette und führt definiert und vollständig auf die finalen Eigenschaften des Produktes hin.
P4	Beherrschtheit der einzelnen Prozessschritte	Prozesse müssen bei einem gegebenen Input stets einen definierten Output liefern. (Definierte Merkmalsstruktur)
P5	Analysierbarkeit aller Prozessschritte	Zur Beherrschung eines Prozesses, muss jeder Prozessschritt beschrieben, analysiert und optimiert werden.
P6	Steuerbarkeit aller elementaren Zustandsänderungen	Jede elementare Zustandsänderung muss steuerbar sein.

In Bezug auf diese Bedingungen wird eine Einschränkung auf diskrete Produktionsprozesse eingeführt. Die Erweiterung „diskret“ erhalten Produktionsprozesse, in denen Produkte als abzählbare Einheiten hergestellt werden. Bei sogenannten Stückgutprozessen, können anders als bei kontinuierlichen Fertigungsprozessen, diskrete Schritte betrachtet, analysiert und gesteuert werden (Wiedenmann, 2001, S. 27f.).

### 3.2 Anforderungen an die Datenvorverarbeitung in Data-Mining-Projekten

Bezugnehmend auf die situative Anpassungsfähigkeit der zu konstruierenden modellgestützten Methode sind ausschließlich generelle Anforderungen zu ermitteln, welche die Phasen der Datenvorbereitung für eine Vielzahl an Data-Mining-Verfahren unterstützen. Die ermittelten Anforderungen sind mit den bereits in CRISP-DM implementierten Anforderungen für diese Phase abzugleichen, um die Nähe zum ausgewählten Standardprozess zu wahren. Die folgende Übersicht stellt die konsolidierten Anforderungen an die Datenvorverarbeitung dar und evaluiert diese durch weitere Quellen der Domäne:

**Tabelle 2: Anforderungen Datenvorverarbeitung**

Abk.	Anforderungen	Quellen
DM1	Umgang mit fehlenden Werten und deren Bedeutung (Missing Values) klären	(IBM, 2010), (Runkler, 2010), (Otte et al., 2004)
DM2	Skalenart ermitteln	(IBM, 2010) (Otte et al., 2004)
DM3	identische Formate/Schreibweisen pro Merkmal sicherstellen	(IBM, 2010), (Otte et al., 2004)

DM4	Wertebereiche von Merkmalen definieren	(IBM, 2010), (Otte et al., 2004)
DM5	Merkmalskorrelationen / Unabhängigkeiten entdecken	(IBM, 2010), (Otte et al., 2004)
DM6	statistische Lagewerte und Streuungsmaße zur Erkennung v. Datenrauschen/Ausreißern erheben	(IBM, 2010), (Runkler, 2010), (Otte et al., 2004)
DM7	Merkmalsbedeutungen / Relevanz ermitteln	(IBM, 2010), (Otte et al., 2004)
DM8	Balancierung der Daten untersuchen	(IBM, 2010), (Otte et al., 2004)
DM9	Schlüsselattribute/Schlüsselbeziehungen erkennen	(IBM, 2010), (Otte et al., 2004)
DM10	bekannte Ursachen und Wirkungen zwischen Merkmalen aufzeigen	(IBM, 2010)

Die Anforderungen der Datenvorbereitung fließen gemeinsam mit den Anforderungen diskreter Produktionsprozesse in die Konstruktion der modellgestützten Methode ein.

## 4 Konstruktionsergebnisse

### 4.1 Sprachbeschreibung

Ziel der modellgestützten Methode ist es, die Vorbereitung von Data-Mining-Analysen im Kontext diskreter Produktionsprozesse zu unterstützen. Dazu soll im ersten Schritt der zu analysierende Produktionsprozess im Ist-Zustand modelliert werden, um anschließend dieses Modell in analysespezifische Modelle zu transformieren, die für die Analyse der Prozess-Exemplardaten (Rohdaten des diskret gefertigten Produktes) genutzt werden können. Entsprechend dem Forschungsdesign erfordert eine modellgestützte Methode eine Sprachbeschreibung für eine Modellierungssprache, zu der die entstehenden Modelle konform sind. Jede Modellierungssprache verfügt dabei über eine festgelegte Syntax, die über eine Grammatik oder ein Metamodell beschrieben werden kann (Hesse & Mayr, 2008, S.389). Das vorgeschlagene Prozessmetamodell definiert dabei im Sinne einer Modellierungssprache alle möglichen Sprachkonzepte und Regeln zu deren Kombination (Wand & Weber, 2002, S. 364), um einen Produktionsprozess so zu modellieren, dass die Anforderungen diskreter Produktionsprozesse und an die Datenvorverarbeitung in Data-Mining-Projekten berücksichtigt werden. Als Ausgangspunkt für die Prozessmetamodellierung dient das Modell der Fertigungssteuerung (MFST) als etablierter Standard für diese Domäne (Dangelmaier und Felser, 1994, S. 35 f.; Großmann et al., S. 957). Durch

die Einschränkung des Anwendungsbereiches grenzt sich dieses Vorgehen von anderen Methoden der (Geschäfts-)prozessmodellierung ab (Dangelmaier & Felser, 1994, S. 36). Im MFST werden Produktionsprozesse als Input-Output-Systeme modelliert, in denen jeder Produktionsschritt als „Black-Box“ betrachtet wird, in den Güter hineinfließen (Inputs) und neue Güter hervorgebracht werden (Outputs) (Dangelmaier, 2009, S. 1 ff., 10 ff.). Der Produktionsprozess drückt sich dabei als Transformation von Input- und Output-Zuständen aus, die durch Merkmale und Merkmalsausprägungen charakterisiert sind. Mehrstufige Produktionsprozesse können in eine Kette einzelner Input-Output-Systeme zerlegt werden.

Abbildung 2 zeigt das entwickelte Prozessmetamodell, das in Anlehnung an MFST (Dangelmaier & Felser, 1994; Großmann et al., 2010) die Anforderungen diskreter Produktionsprozesse abdeckt und um die Anforderungen an die Datenvorverarbeitung in Data-Mining-Projekten erweitert wurde. Für die Darstellung des Prozessmetamodells wird ein Entity-Relationship-Modell gewählt und die Bedeutung der verwendeten Sprachkonzepte nachfolgend näher erläutert.

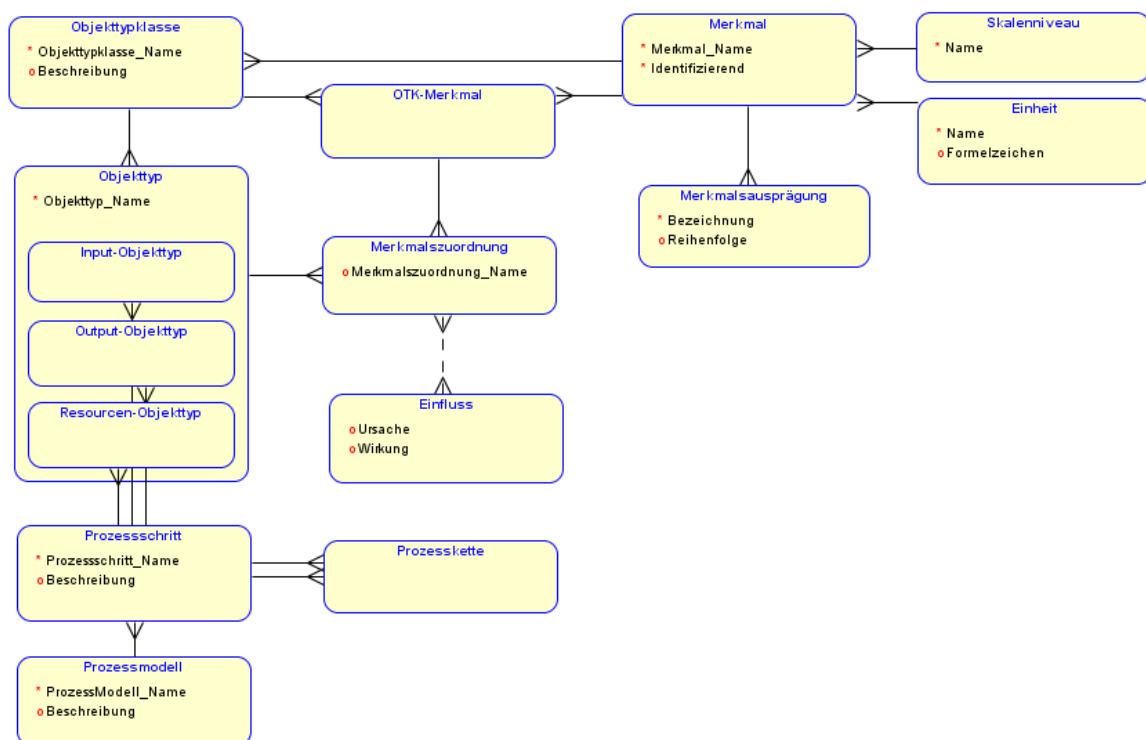


Abbildung 2: Prozessmetamodell – logische Sicht

Ein einstufiger Prozess bzw. ein *Prozessschritt* wird dabei grundsätzlich als eine Menge von *Objekttypen* repräsentiert, die unter Nutzung eines oder mehrerer *Ressourcen-Objekttypen* (z.B. Maschinen, Werkzeuge, u.a.) die Transformation eines oder mehrerer

**Input-Objekttypen** (z.B. Werkstücke, Material, u.a.) in ein oder mehrere **Output-Objekttypen** herbeiführen. Mehrstufige Prozesse werden über eine **Prozesskette** in eine chronologische Reihenfolge gebracht. Objekttypen abstrahieren individuelle Objekte (Exemplare) im Produktionsprozess und beschreiben immer genau einen Objektzustand. Sie können daher im Modell nur einmal verwendet werden. Selbst ein einziges zu bearbeitendes Produkt wird im Modell durch zwei Objekttypen abgebildet: einen für den Zustand vor der Bearbeitung (Input-Objekttyp) und einen für den Zustand nach der Bearbeitung (Output-Objekttyp) (Weller et al., 2010, S.76). **Objekttypklassen** (OTK) werden verwendet, um auf Modellebene dennoch den Zusammenhang zwischen gleichartigen Objekttypen hinsichtlich ihrer **Merkmale** abbilden zu können. Da die gleichen Merkmale von mehreren Objekttypklassen genutzt werden können, stellen die **OTK-Merkmale** diejenigen Merkmale dar, die aus der Gesamtheit aller Merkmale den Objekttypklassen zugeordnet sind. Für die Analyse mehrstufiger Prozesse ist zudem die Kenntnis von identifizierenden Schlüsselmerkmalen wesentlich, die direkt den Objekttypklassen zugewiesen werden. In der Regel wird dieses Merkmal ein Code oder eine Seriennummer sein mit der sich ein Exemplar eines Werkstückes im Produktionsprozess eindeutig bestimmen lässt.

Auf Modellebene können für Merkmale Sollwerte und Toleranzen für Merkmalsausprägungen vorgegeben werden, mit denen sich in der späteren Analyse der tatsächlichen Merkmalsausprägungen der Exemplardaten prozessuntypische Ausreißer erkennen lassen. Für Merkmale mit Nominal- oder Ordinalskala können im Prozessmetamodell gültige **Merkmalsausprägungen** definiert werden. Ein Vergleich dieses Wertebereichs mit dem tatsächlichen Vorkommens der Ausprägung in den Exemplardaten lässt auf die Balancierung dieses Merkmales schließen. Eine wichtige Anforderung aus dem Data Mining ist die Zuordnung des **Skalenniveaus** für jedes Merkmal. Daraus lässt sich später ableiten, welche Merkmale für welches Data-Mining-Verfahren in Frage kommen oder wie diese gegebenenfalls transformiert werden müssen. Für jedes Merkmal kann zusätzlich eine **Einheit** hinterlegt werden.

Jeder Objekttyp ist genau einer Objekttypklasse zugeordnet und besitzt daher alle Merkmale seiner Klasse, aber nicht alle Merkmale sind in jedem Prozessschritt von Interesse. Mit der **Merkmalszuordnung** können die für den entsprechenden Prozessschritt relevanten Zustände der verschiedenen Objekttypen einer Objekttypklasse beschrieben und für den jeweiligen Objekttyp gültige Sollwerte und Toleranzen lokal vorgegeben werden, welche die globalen Werte der Objekttypklasse überschreiben. Gleichzeitig lässt sich für den Objekttyp das Fehlen von Merkmalsausprägungen dokumentieren, was ein wesentliches Indiz für die Behandlung von Missing Values darstellt. Eine weitere wichtige im Prozessmetamodell abgebildete Anforderung, ist die Beschreibung von (vermuteten) Ursache-Wirkungsbeziehungen zwischen Merkmalen verschiedener Objekttypen. Ursache

und Wirkung lassen sich dabei als *Einfluss* in Form von Implikationen zwischen Aussagen, die mittels Variablen, Konstanten, Funktionen und (mathematischen u. logischen) Operatoren definiert werden, beschreiben. Gleichzeitig kann der Einfluss genutzt werden um Merkmalsabhängigkeiten innerhalb eines Objekttyps zu dokumentieren, was bei der anschließenden Auswahl von Eingangsmerkmalen für die Analyseverfahren unterstützt.

Aus dem Prozessmetamodell lässt sich ein relationales Datenmodell entwickeln, indem aus Entitäten Tabellen, aus Attributen Spalten und aus Relationen Schlüsselbeziehungen werden. Für die Implementierung des Prozessmetamodells und die Erfassung der analyse-spezifischen Prozessmetadaten und Objektzustände für die Modellerstellung eignet sich daher eine datenbankbasierte Anwendung.

## 4.2 Prozessbeschreibung

### Modellerstellung

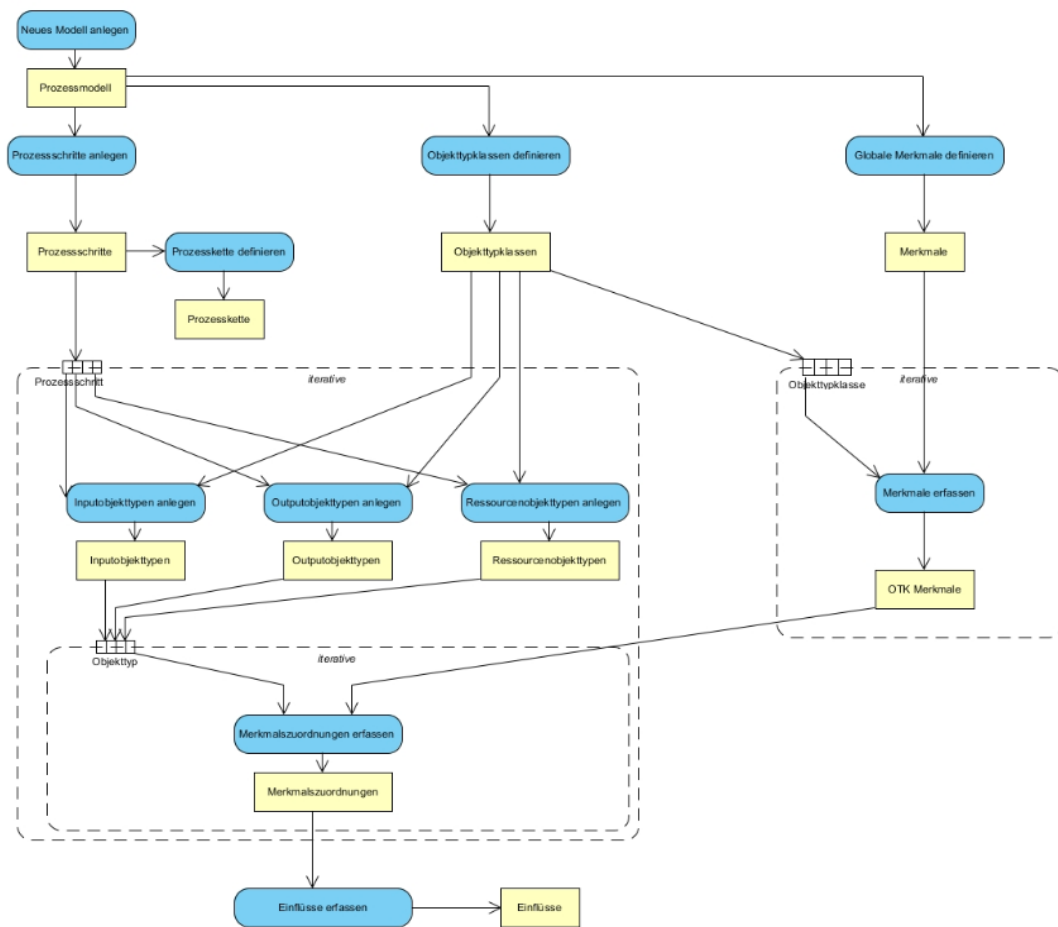
Gemäß Forschungsdesign benötigt die modellgestützte Methode neben der vorgestellten Sprachbeschreibung noch eine Prozessbeschreibung zur Erstellung eines Prozessmodells (Wiedenmann, 2001, S. 22; Weller, 2010, S. 42). Mit der Modellerstellung wird ein realer Produktionsprozess durch den Domänen-Experten, in ein Prozessmodell überführt, um anschließend für die Vorbereitung der spezifischen Prozessdaten eingesetzt zu werden.

In Anlehnung an die Vorgehensbeschreibung von CRISP-DM beginnt jedes Projekt mit der Formulierung der Geschäftsziele, welche den Anlass definieren und den Rahmen des Vorhabens bilden (Marban et al., 2007, S. 97 f.; Sharma & Osei-Bryson, 2009, S. 4116). Innerhalb der Analyse von Prozessen dienen die Geschäftsziele vordergründig der Beschreibung der geschäftlichen Erfolgsfaktoren und der Modellweite zur Fokussierung des zu untersuchenden Prozessausschnittes. Auf Grund der fehlenden Kenntnisse über die Input-Output-Relationen (Black-Box) zwischen den vorhandenen Prozessstrukturelementen werden für die Modellierung alle Objektzustände im ausgewählten Prozess betrachtet (Modellgranularität). Begrenzt durch die Modellweite erfolgt die Unterstützung der Situationsbeschreibung durch die Modellierung des konkreten Prozesses unter Verwendung der definierten Sprachbeschreibung.

Beginnend mit der Erfassung des Endproduktes und der beteiligten Fertigungsmittel als Objekttypklassen mit jeweils globalen Merkmalen und deren Ausprägungen, werden die einzelnen Prozessschritte mit ihrer Einordnung in der Prozesskette erfasst und einem Prozessmodell zugeordnet. Anschließend beginnt die Modellierung der einzelnen Prozessschritte in ihrer Außensicht, welche durch Input-, Output- und Ressourcen-Objekttypen sowie deren Merkmalszuordnungen definiert ist. Gemäß der globalen Merkmalszuord-



nung verfügt dabei jeder Objekttyp über die Merkmale seiner Objekttypklasse. Sollte innerhalb der Modellerstellung festgestellt werden, dass ein Merkmal fehlt, kann dies dem Merkmalskatalog hinzugefügt und über die Merkmalszuordnung verwendet werden. Im Anschluss an die Modellierung der Objekttypen können vermutete Ursache-Wirkungsbeziehungen innerhalb eines Objekttyps sowie zwischen Objekttypen in die Modellierung aufgenommen werden. Das beschriebene Vorgehen wird für jeden Prozessschritt innerhalb der Prozesskette durchgeführt. Das dazugehörige UML-Aktivitätsdiagramm des Ablaufes ist in Abbildung 3 dargestellt.



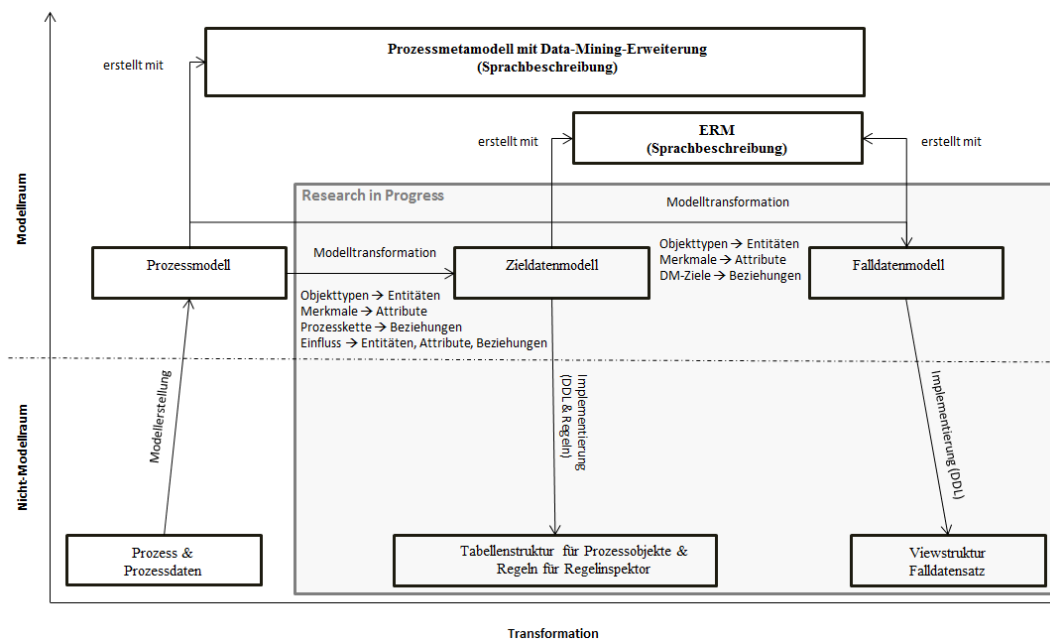
**Abbildung 3: Prozessbeschreibung Modellerstellung**

Reflektierend auf CRISP-DM wird innerhalb der Modellerstellung die Struktur des Prozesses erfasst und beschrieben. Anders als bei der darauffolgenden explorativen Datenuntersuchung und Überprüfung der Datenqualität, definiert die Modellerstellung Qualitätskriterien in Form von Metadaten und schafft Transparenz durch die Integration von Prozesswissen in Form von Ursache-Wirkungsbeziehungen. Durch die exakte Beschreibung

der Prozessstrukturen entsteht die Grundlage für eine semantisch korrekte Integration von relevanten Daten innerhalb der Datenaufbereitungsphase. Anhand dieser Aspekte wird sichtbar, dass durch das Prozessmodell eine Transformation des realen Prozesses in einen für Analysen aufbereiteten Untersuchungsbereich durchgeführt werden kann und somit die Überführung zwischen interdisziplinären Welten modellgestützt erfolgt.

### Modelltransformation und Modellnutzung (Research in Progress)

Neben der Beschreibung der Modellierungsmethode zur Erstellung eines Prozessmodells geben modellgestützte Methoden konkrete Hinweise zur Transformation und Nutzung der erstellten Modelle. Durch eine Modelltransformation kann eine entsprechende Lösung zunächst wiederum als Modell dargestellt werden. Dazu wird das erzeugte Prozessmodell als Repräsentant der Domäne derart modifiziert bis eine Lösung des beschriebenen Problems auf Modellebene gefunden ist (Weller, 2010, S. 40). Sind sich alle Beteiligten einig eine Problemlösung im Modellraum gefunden zu haben, muss diese Lösung auf das fachliche Problem der Realität (Nicht-Modellraum) übertragen werden. Dies erfolgt mit einer Beschreibung der Modellnutzung.



**Abbildung 4: Research in Progress - Zerlegung des Modellierungsproblems**

Bezogen auf die vorgestellte modellgestützte Methode soll das erzeugte Prozessmodell in ein Zieldatenmodell – in ERM-Sprachbeschreibung – überführt werden, um anschließend zur Generierung einer Tabellenstruktur für die konkreten Prozessdaten genutzt zu werden (siehe Abbildung 4). Diesbezüglich müssen Regelwerke definiert werden, welche den Übergang von einem Modell in das andere semiformal beschreiben. Die konkrete Nut-

zung des Zieldatenmodells erfolgt anhand eines generierten SQL-Skriptes, welches die relationale Tabellenstruktur anlegt und somit das Ziel für einen spezifischen Datenladeprozess aus den operativen Datenquellsystemen definiert. Weiterhin werden auf Basis der erfassten Prozessmetadaten (z.B. Sollwerte, Ursache-Wirkungsbeziehungen) Regeln generiert, welche zur Prüfung der Datenqualität auf die geladenen Prozessdaten angewendet werden.

Ausgehend von den modellierten Data-Mining-Zielen im Prozessmodell, werden alle vorgelagerten Objekttypen und deren Merkmale anhand eines identifizierenden Merkmals in einem analysegerechten Falldatenmodell – in ERM-Sprachbeschreibung – organisiert. Die konkrete Modellnutzung erfolgt über SQL-Skripte, welche je Data-Mining-Ziel eine View (Falldatensatz) auf die qualitätsgeprüfte Zieldatenstruktur definiert. Die erzeugten Falldatensätze werden anschließend als Ergebnis an die jeweiligen Data-Mining-Verfahren zur Analyse übergeben.

## 5 Evaluation und Ausblick

Die Systematisierung von Evaluationsmethoden innerhalb der Wirtschaftsinformatik (siehe Riege, Saat & Bucher, 2009, S. 75) bietet unterschiedliche Ansatzpunkte zur Evaluierung von entwickelten Artefakten. Im Beitrag wurde eine modellgestützte Methode als Lösungskandidat zur spezifischen Analysevorbereitung von Data-Mining-Projekten im Kontext von diskreten Produktionsprozessen vorgestellt. Anhand von abgeleiteten Konkretisierungsbedarfen und Anforderungen konnten die Sprachbeschreibung sowie die Prozessbeschreibung zur Modellerstellung fertig konstruiert und damit Teilaspekte der zweiten Forschungsfrage beantwortet werden. Die Prozessbeschreibung für die Modelltransformation und Modellnutzung wurden konzeptuell beschrieben und als Research in Progress gekennzeichnet. Für eine weiterführende Entwicklung ist es jedoch notwendig, die finalisierten Artefakte hinsichtlich ihrer korrekten Konstruktion auf Basis feststehender Anforderungen zu überprüfen (merkmalsbasierte Evaluation) (Riege, Saat & Bucher, 2009, S. 75). Die nachfolgende Übersicht macht die Ergebnisse der Evaluation ersichtlich und zeigt zugleich die konsolidierten Anforderungen zur Beantwortung der ersten Forschungsfrage auf.

**Tabelle 3: Merkmalsbasierte Evaluation**

Abk.	Anforderungen	Bewertung
<b>Begründungen für eine Konkretisierung von CRISP-DM</b>		
B1	Eine situationsbezogene Konkretisierung ist methodisch nicht gewährleistet.	Erfüllt – Input-Output-System ganzheitlich abgebildet

B2	Die Ermittlung der relevanten Datenquellen für eine Analyse bleibt sehr vage und isoliert vom eigentlichen Analyseobjekt.	Erfüllt – Objekttypen und deren Merkmalszuordnung beschreiben die exakte Zieldatenstruktur
B3	Die Integration von Rollen (z. B. Domänen- u. Data-Mining-Experte) in das Vorgehen wird nicht geregelt.	Teilweise – Domänen-Experte ist integriert, Data-Mining-Experte nach Modelltransformation
B4	Eindeutige Zusammenhänge und Abhängigkeiten zwischen den jeweiligen Ergebnissen der einzelnen Phasen werden außer Acht gelassen.	Teilweise – Prozessmodell als zentraler Ergebnispeicher
B5	Ergebnisse werden in ihrer Form nicht spezifiziert und sind damit nur schwer wiederverwendbar.	Teilweise – in Sprachbeschreibung definiert
B6	Die Wiederverwendbarkeit von vorbereitenden Teilergebnissen bei unterschiedlichen Analyseverfahren ist nicht gegeben.	Teilweise – wird nach Modelltransformation vollständig erfüllt
<b>Anforderungen diskreter Produktionsprozesse</b>		
P1	Beschreibung des Prozesses als Input-Output-System	Erfüllt – Sprach- und Prozessbeschreibung
P2	Definiertheit der finalen Produkteigenschaften	Erfüllt - Sprachbeschreibung
P3	Durchgängigkeit der gesamten Prozesskette	Erfüllt - Sprachbeschreibung
P4	Beherrschtheit der einzelnen Prozessschritte	Erfüllt - Sprachbeschreibung
P5	Analysierbarkeit aller Prozessschritte	Erfüllt - Sprachbeschreibung
P6	Steuerbarkeit aller elementaren Zustandsänderungen	Erfüllt - Sprachbeschreibung
<b>Anforderungen der Datenvorverarbeitung</b>		
DM1	Umgang mit fehlenden Werten und deren Bedeutung (Missing Values) klären	Erfüllt – Sprachbeschreibung
DM2	Skalenart ermitteln	Erfüllt - Sprachbeschreibung
DM3	identische Formate/Schreibweisen pro Merkmal sicherstellen	Erfüllt - Sprachbeschreibung
DM4	Wertebereiche von Merkmalen definieren	Erfüllt - Sprachbeschreibung
DM5	Merkmalskorrelationen / Unabhängigkeiten entdecken	Erfüllt - Sprachbeschreibung
DM6	statistische Lagewerte und Streuungsmaße zur Erkennung v. Datenrauschen/Ausreißern erheben	Erfüllt - Sprachbeschreibung
DM7	Merkmalsbedeutungen / Relevanz ermitteln	Erfüllt - Sprachbeschreibung
DM8	Balancierung der Daten untersuchen	Erfüllt - Sprachbeschreibung

DM9	Schlüsselattribute/Schlüsselbeziehungen erkennen	Erfüllt - Sprachbeschreibung
DM10	bekannte Ursachen und Wirkungen zwischen Merkmalen aufzeigen	Erfüllt - Sprachbeschreibung

Zusammenfassend kann die Konstruktion der existenten Ergebnisse positiv evaluiert werden. Die Anforderungspunkte B3 bis B6 sind aufgrund der noch umzusetzenden Modelltransformation und -nutzung nur teilweise erfüllt.

Die modellgestützte Methode ermöglicht bereits mit der Modellerstellung eine gezielte Zusammenführung der Fach- und Data-Mining-Domäne und schafft damit ein Kommunikationsmittel für eine aufgabenteilige, interdisziplinäre Bearbeitung von analytischen Fragestellungen in diskreten Produktionsprozessen. Durch die integrierte, modellgestützte Beschreibung des betrachteten Produktionsprozesses entsteht eine einheitliche Wissensbasis über semantische Ablaufbeziehungen sowie über die zugrundeliegende Datenstruktur und deren Besonderheiten. Mittels Zuführung von Metadaten erhält der Analyst notwendige Informationen über die Prozessdatenstruktur zur effizienten Konfiguration von Data-Mining-Verfahren.

Bezogen auf weitere Forschungstätigkeiten dient das Prozessmodell zur Generierung eines analysespezifischen Zieldatenmodells und spezifischer Falldatenmodelle für die vorhandenen Exemplardaten des jeweiligen Prozessschrittes sowie als Grundlage für eine strukturelle Überprüfung der Datenqualität und Verfahrensvoraussetzungen für die nachgelagerten Analysephasen. Die Konstruktion einer Prozessbeschreibung für die Modelltransformation und Modellnutzung sowie die Implementierung eines „Regelinspektors“ zur Realisierung der strukturellen Prüfungen auf Basis der integrierten Metadaten auf Merkmalsebene repräsentiert dazu den nächsten Forschungsschritt.

## 6 Literaturverzeichnis

- Azevedo, A., Santos, M. F. (eds.): KDD, Semma and CRISP-DM: A parallel overview (2008)
- Dangelmaier, W.: Theorie der Produktionsplanung und -steuerung Springer-Verlag, Berlin Heidelberg (2009)
- Dangelmaier, W., Felser, W.: Ganzheitliche Modellierung von Fertigungsprozessen. Ein erster Schritt bei der Konstruktion unternehmensspezifischer Fertigungssteuerungssysteme. The Electronic Library of Mathematics 1994, 34–48 (1994)
- Großmann, K., Wiemer, H., Weller, J., Großmann, K.K.: Reproduzierbare Fertigung in innovativen Prozessketten. Konzeption eines Beschreibungs- und Analysetools (Teil 2). ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb 105, 954–958 (2010)

- Großmann, K., Wiemer, H.: Reproduzierbare Fertigung in innovativen Prozessketten. Besonderheiten innovativer Prozessketten und methodische Ansätze für Ihre Beschreibung, Analyse und Führung (Teil 1). ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb 105, 855–859 (2010)
- Wolfgang Hesse, Heinrich C. Mayr: Modellierung in der Softwaretechnik: eine Bestandsaufnahme. Informatik-Spektrum, Vol. 31, No. 5., 377-393, (2008)
- IBM: CRISP-DM 1.0 - Step-by-step data mining guide (2010)
- Kagermann, H.; Wahlster, W.; Helbig, J.. (Hrsg.): Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0. Deutsche Akademie der Technikwissenschaften e.V., München, 2013.
- Kiener, S., Maier-Scheubeck, N., Obermaier, R., Weiß, M.: Produktions-Management. Grundlagen der Produktionsplanung und -steuerung Oldenbourg Verlag, München (2012)
- Knollmann, M., Meyer, M., Windt, K.: Data Mining-Methoden in der Produktionslogistik. Wissensgenerierung beim Umgang mit komplexen Daten und multikriteriellen Entscheidungen. Industrie Management, 51–55 (2012)
- Lukasz A., K., Musilek, P.: A survey of Knowledge Discovery and Data Mining process models. Cambridge University Press 2006, 1–24 (2006)
- Marban, O., et al.: From the Business Decision Modeling to the Use Case Modeling in Data Mining Projects (2007)
- Mariscal, G., Ó.M.C.F.: A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review 2010, 137–166 (2010)
- Runkler, T.A.: Data Mining. Methoden und Algorithmen intelligenter Datenanalyse, VIEWEG+TEUBNER, Wiesbaden (2010)
- Riege, C., Saat, J., Bucher, T.: Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. in: Becker, J., Krcmar, H., Niehaves, B.: Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik Physica-Verlag, Heidelberg (2009)
- Sharma, S., Osei-Bryson, K.-M.: Framework for formal implementation of the business understanding phase of data mining projects. Expert Systems with Applications, 4114–4124 (2009)
- Otte, R., Otte, V., Kaiser, V.: Data Mining für die industrielle Praxis Hanser Verlag, München, Wien (2004)
- Weller, J.: Modellgestützte Prozessverbesserung. Entwicklung einer wiederverwendungsorientierten Methode zur durchgängigen Unterstützung der Modellerstellung, -transformation und -nutzung im Rahmen der Prozessverbesserung Dresden (2010)

- 
- Weller, J., et al.: Modellierung in der Produktionstechnik: Ein Ansatz zur effektiven Generierung von Technologie-Know-how für die Absicherung einer reproduzierbaren Fertigung. In: Esswein, W., Jührisch, M., Turowski, K. (eds.): Modellierung betrieblicher Informationssysteme. Modellgestütztes Management, 69–86 (2010)
- Wiedenmann, H.: Modellierung von Produktionsprozessen als Beitrag zur Generierung von Termin- und Kapazitätsplanungs-Systemen bei variantenreicher Serienfertigung Jost-Jetter Verlag, Heimsheim (2001)
- Wand, Y.; Weber, R.: Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda. In: Information Systems Research, 13, 363–377 (2002)