Simulation in Production
and Logistics 2015
Markus Rabe & Uwe Clausen (eds.)
Fraunhofer IRB Verlag, Stuttgart 2015

A Procedure Model for the Credible Measurability of Data Warehouse Metrics on Discrete-event Simulation Models of Logistics Systems

Eine Vorgehensweise zur glaubwürdigen Messbarkeit von Data-Warehouse-Kennzahlen an ereignisdiskreten Simulationsmodellen von Logistiksystemen

Markus Rabe, Felix Dross, Andrea Vennemann, TU Dortmund, Dortmund (Germany), markus.rabe@tu-dortmund.de, felix.dross@tu-dortmund.de, andrea.vennemann@tu-dortmund.de

Abstract: This paper presents a new procedure model for the credible measurability of data warehouse key performance indicators on simulation models of discrete-event logistics systems. The basis for the new procedure model was the procedure model of information acquisition by Bernhard and Wenzel (2005). The new model consists of two parts, an input data management and an output data management. The complete procedure model has been integrated into the procedure model for simulation including V&V by Rabe et al. (2008b). Furthermore, a concept for a control software for the automated processing of the input and output data has been developed and is presented in this paper.

1 Introduction

Discrete-event logistics systems (DELS) have been described as large-scale complex socio-technical systems which operate in an environment of uncertainty. Examples for such systems can be found in warehousing, manufacturing, and supply chains. Discrete-event simulation (DES) is the most commonly used analysis tool for evaluating DELS performance (McGinnis 2005). Data warehouses (DWHs) provide data structures in order to support tools for analytical decision making. In contrast to common databases, DWHs store data in redundant and aggregated ways, speeding up interactive analysis and providing data at sufficient aggregation levels (Ehmke et al. 2011). Online analytical processing (OLAP) software provides a fast, flexible and interactive access to the data in a DWH and enables the organisation, aggregation and visualisation of information. Data are presented in terms of hypercubes depicting multidimensional structures. The cubes visualise system performance measures (the cube's cells) in the context of their dimensions (the cube's borders) and therefore enable the flexible and multidimensional analysis of the data (Jarke et al. 2003;

Ehmke et al. 2011; Bauer and Günzel 2013). OLAP technology is often used to realise performance measurement systems (PMS). A PMS unites different performance measures that relate to each other in a hierarchical form and typically culminate in one key performance indicator (KPI) (Bauer and Günzel 2013; Dross and Rabe 2014). KPIs are used by the management to measure, control and steer DELS. While the combination of DES models with DWH technology has been described before, previous papers mainly focus on harnessing the technological power of OLAP on the input or output data of DES models. They do not describe the steps which are necessary to obtain a credible measurability of existing DWH-KPIs for the management on the output data of a DES model of a DELS. In this paper, the authors present a procedure model as well as a software architecture, which can be used to achieve a credible measurability of data warehouse metrics on discrete-event simulation models of logistics systems.

The paper is structured as follows: Section 2 provides an overview of related work. The authors list relevant literature which focuses on the combination of DES models with DWH technology as well as literature with related procedure models. Some of the presented procedure models have been used as a basis or have inspired the newly developed procedure model. Section 3 presents the developed procedure model. Section 4 introduces a software architecture which can be used to achieve a credible measurability of DWH-KPIs on a DES model. The procedure model presented in Section 3 can be used to support the implementation of the software architecture presented in Section 4. Section 5 closes the paper with a conclusion and an outlook on future research.

2 Related Work

Koutsoukis et al. (1999) and El-Darzi et al. (2001) describe the benefits of using DWH and OLAP software in the preparation of data for the estimation of input parameters for various decision models. The streaming of simulation output data to a data warehouse was mentioned by Banks (1997). The use of data warehouse technology as an approach for the interactive and flexible analysis of detailed state-transition data collected from DES models of logistics systems has been further described by Vasilakis et al. (2004) and Ehmke et al. (2011). However, none of them emphasises the importance of a credible measurability of management KPIs on a DES model. A situation where a credible measurability of DWH-KPIs on a DES model is particularly important has been extensively described by Dross and Rabe (2014). The authors describe a situation where a simulation model is used as the basis for a decision support system (DSS) which regularly receives the output of complex PMS. The DSS uses the outputs of the PMS and is aimed to evaluate possible actions regarding the DWH-KPIs in the DES model.

In order to ensure quality and increase the rapidity in DES projects in general, there already exist well-structured methodologies to follow (Law and Kelton 2000, Rabe et al. 2008a, 2008b; Wenzel et al. 2008; Verein deutscher Ingenieure e.V. 2014). Although it is a very important part of a simulation study, the process of collecting all relevant information for a DES model has been a comparably small topic in many classical discrete-event simulation books (e.g., Vincent 1998; Law and Kelton 2000; Robinson 2004; Pidd 2004; Banks et al. 2014). Consequently, dedicated procedure models for the data collection for discrete-event simulation models have been

developed. Lehtonen and Seppala (1997) present a methodology for data gathering and analysis in logistics simulation projects. Baron et al. (2001) developed checklists for the identification of necessary data for a simulation model. Based on Baron et al. (2001), Csanady et al. (2008) developed a methodology for the systematic information retrieval for simulation studies. Skoogh and Johansson (2008) present an extensive methodology for the input data management in DES projects. Their methodology can be embedded into a procedure model for simulation projects, such as the one described by Law and Kelton (2000). Bengtsson et al. (2009) present an approach to use a methodology to identify and collect data, then use an input data management software to extract and process the data. Another methodology for simulation input data management is presented by Hill and Onggo (2012; 2014). The methodology was derived from observations made in an action research conducted at a management consultancy company. The procedure model of information acquisition (PMIA) by Bernhard and Wenzel (2005) is another procedure model for the input data management phase of a simulation study. In contrast to the other procedure models described above, the PMIA is based on the idea of separating data and information. Furthermore, it provides a consistent V&V of the phase results. It has already been integrated into the procedure model for simulation including V&V by Rabe et al. (Bernhard et al. 2007; Jodin et al. 2009). Since the authors of this paper aimed for a credible measurability of DWH-KPIs, a consistent V&V of the phase results is a very important aspect. Therefore, the PMIA has been chosen as the basis for the newly developed procedure model presented in this paper. Unlike for the processing of the input data, there exists no concrete procedure model for the processing of the simulation output data. The VDI guideline 3633 sheet 3 provides some recommendations for the processing of the output data (Verein deutscher Ingenieure e.V. 1997). Actually, in most of the literature the emphasis lies on the statistical processing of the simulation output data (Banks et al. 2014; Law and Kelton 2000). Therefore, the output data management of the newly developed procedure model has been newly constructed according to the PMIA. Finally, most of these aspects have been integrated into the procedure model for simulation including V&V by Rabe et al. (2008a; 2008b).

3 Procedure Model

The developed procedure model can be divided into two logical parts, the input data management and the output data management. Figure 1 shows the complete procedure model as an extension to the procedure model by Rabe et al. (2008b). In the following sections, the authors will explain the new elements of the procedure model. For a detailed description of the older elements, the reader is referred to the respective literature. The authors assume that the desired DWH-KPIs have been documented in the task description of the simulation study.

3.1 Input Data Management

In this part of the procedure model, the required output data of the simulation model for the measurement of the desired DWH-KPIs are identified. Furthermore, the deviation to the respectively required raw data for the measurement of the desired DWH-KPIs is documented. This information is then used to identify and collect the necessary input data using the PMIA.

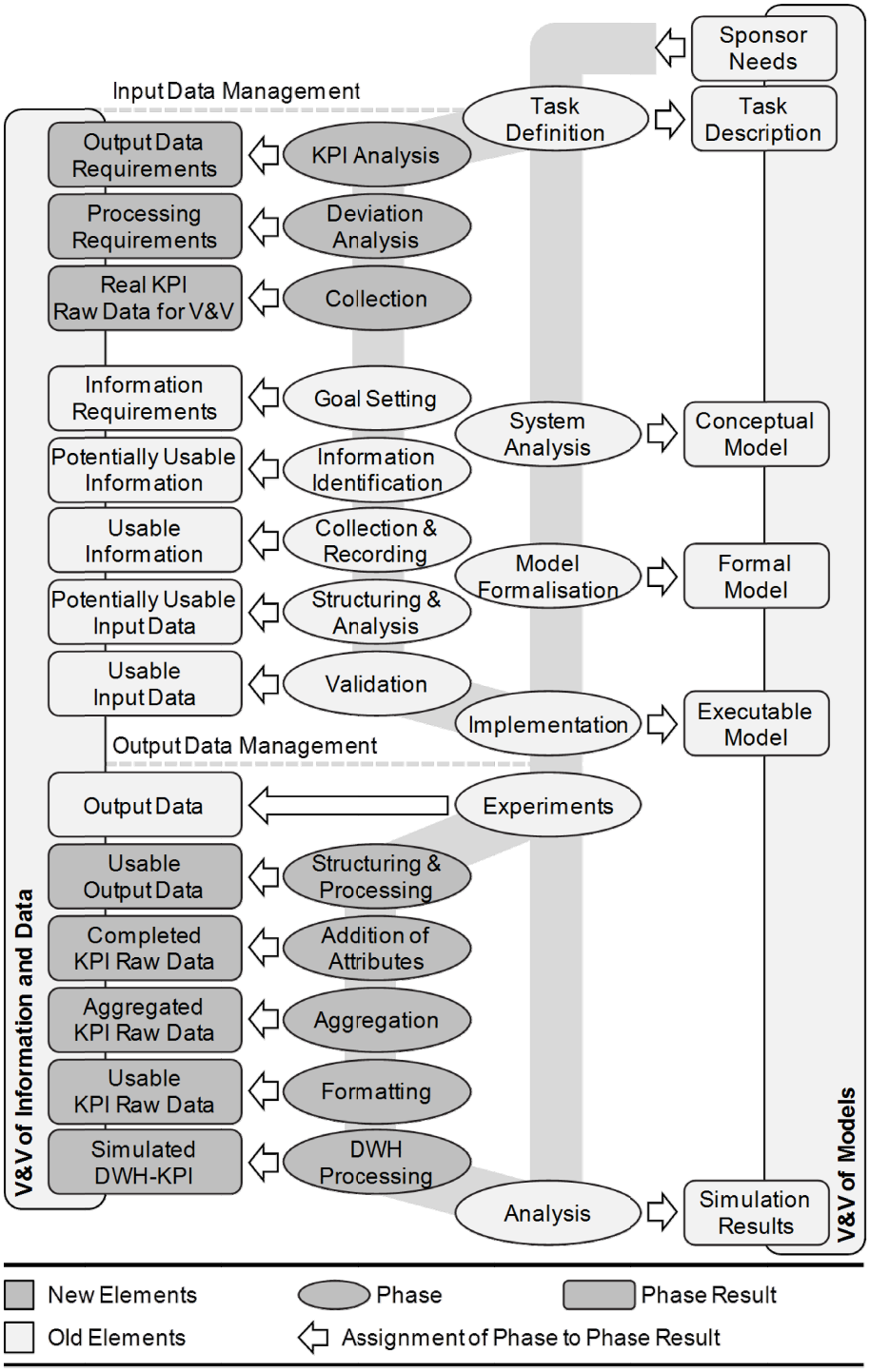

Figure 1: *Procedure model for a credible measurability of data warehouse metrics (Based on the procedure model for simulation including V&V by Rabe et al. 2008b)*

3.1.1 Identification of Necessary Simulation Output Data

The first new element in the procedure model is the KPI analysis. Within this phase, the first step is a general analysis of the DWH-KPIs, in which the focus lies on the logic of the KPIs. The technical aspects of the implementation in the DWH are initially ignored. The aim of the general analysis is to obtain an overview of the KPIs. The composition of the KPIs is determined by gradually dividing the calculation rules of the PMS in smaller parts in a top-down manner, which helps to resolve the complexity of the PMS. The analysis can be supported with KPI profiles, in which the structure of the PMS has been documented. If these are incomplete or non-existent, an interview with the PMS experts has to be conducted. In case of commonly known KPIs, the corresponding calculation rule can be taken from adequate literature, but it has to be ensured that the calculation rule has not to be adjusted in a company-specific way. If the analysis includes more than one KPI, the KPIs have to be analysed together, so that potential overlaps can be easily identified. To document the findings, the PMS can be represented graphically, e.g. in form of a tree structure. After the general analysis, the technical implementation of the DWH-KPIs in the DWH is examined. If an existing simulation tool is used for the simulation study, the simulation experts also have to identify which simulation output data can already be produced by the tool. The experts have to examine how the provided simulation output data fit to the raw data requirements of the DWH-KPIs. One important aspect to consider is that the lower layers of a PMS implementation in a DWH are mostly used to harmonise and aggregate the raw data which usually come from the operational systems. When using the output data of a simulation model as the data input source of a DWH, it might be possible to insert the data into much higher layers of the DWH. Hence, the simulation experts have to carefully identify which simulation output data will be used and in which specific input cube within the PMS implementations the simulation output data have to be written. Generally, choosing higher input cubes in the PMS implementation in the DWH leads to performance improvements in the calculation of the desired DWH-KPIs. With a suitable choice of the required simulation output data and the corresponding input cubes, the processing requirements can be minimised.

3.1.2 Identification of Processing Requirements

Ideally, the required output data of the simulation model and the corresponding input cubes in the DWH match in every aspect. In most cases though, the output data have to be processed to match the data requirements of the respective input cube. The goal of the deviation analysis is to identify those processing requirements. The processing requirements should be documented regarding attributes, aggregation and formatting. Regarding the attributes, the output data of the simulation model might not include certain attributes which are essential for the processing of the data in the DWH. In this case, the missing attributes have to be added to the output data. For example, the output data might include article numbers, but not corresponding article group numbers, which are required for further processing steps in the DWH. Regarding the aggregation, the output data may not meet the aggregation requirements of the DWH. For example, the output data may be recorded on a daily basis, but the input cube expects data on a monthly basis. Regarding the formatting, the output data might differ in terms of data types or in terms of the syntactical formatting.

3.1.3 Collecting Real KPI Raw Data for V&V

In order to achieve a sufficient match of simulated to real DWH-KPIs, it is mandatory to achieve a sufficient match of the simulation output data to the real raw data for the measurement of the DWH-KPIs. Hence, the real KPI raw data are useful to continuously validating the output data of the simulation model. In the collection phase, the real raw data for the DWH-KPIs are therefore collected and validated by calculating the desired DWH-KPIs manually, e.g. with a spreadsheet software.

3.1.4 PMIA

Based on the identified simulation output data requirements, the PMIA is now applied to determine the information needs for the simulation study and to obtain the relevant input data for the simulation model. These steps will not be discussed further here, because the steps have already been sufficiently described in the respective literature.

3.2 Output Data Management

Once the simulation model can produce the required output data, the output data have to be processed and loaded into the respective input cubes in the DWH. The output data management part of the procedure model focuses on the processing of the simulation output data in order to measure the desired DWH-KPIs. The authors assume that the simulation output data have already been validated e.g., using the real KPI raw data collected in the collection phase of the input data management.

3.2.1 Structuring and Processing of Simulation Output Data

First, the data needed to calculate the desired DWH-KPIs have to be selected from the overall simulation output data. If more than one input cube in the DWH is used, the selected output data have to be sorted respectively. Subsequently, a plausibility check with respect to the accuracy of the selection and sorting has to be performed. If more than one simulation run was performed, the output data of the different simulation runs now have to be combined into one data set using adequate statistical methods. Finally, a verification of the statistical processing steps and a validation of the processed output data have to be conducted.

3.2.2 Addition of Missing Attributes

If it is impossible to generate mandatory attributes of the simulation output data directly by an appropriate programming of the simulation model, these attributes have to be added to the output data. The missing attributes can usually be added using a relation to the simulation input data. For example, if the output data would only contain customer names, but customer numbers are required, the appropriate mapping of customer names to customer numbers could be obtained from the simulation input data. The completed data have to be checked for inconsistencies, e.g., using a desk test.

3.2.3 Aggregation

The aggregation of the data can be done along different dimensions. If the classification hierarchy of a dimension is generally known, such as the hierarchy of time, no further information for the aggregation process is required. If a company-

specific classification hierarchy is used, such as the classification into product groups, the classification levels have to be identified. As a prerequisite for the aggregation of the data, the completeness, disjointness and type compatibility of the data have to be ensured.

3.2.4 Formatting

As a final processing step, the data need to be formatted into a DWH-compatible form. For data types that are not strings, the data format is already defined by the respective data type. Therefore, formatting of the data is implicitly done with the conversion into the respective data type. For data types that contain a string, a syntactic formatting of the data might be necessary. This includes, for example, the adaptation of uppercase and lowercase letters within a string.

3.2.5 Measuring the DWH-KPI on the Simulation Output Data

Finally, the data are ready to be loaded into the input cubes in the DWH and the desired DHW-KPIs can be measured. The KPIs have to be validated with the real DWH-KPIs. If the simulated DWH-KPIs are sufficiently precise, the analysis of the DELS can be conducted now. Experimenting with the simulation model can lead to the desired insights regarding the connection of changes in the DELS to changes in the DWH-KPIs.

4 Software Architecture

Since the transformation of raw data into simulation input data and the processing of the output data into the required DWH data format is very time consuming if done manually, the authors have developed a concept for a control software which can perform the necessary steps automatically. The respective architecture is shown in Figure 2. The DWH system on the left is modelled after the reference architecture for DWH systems described by Bauer and Günzel (2013). Next to it is the simulation environment. The simulation of the DELS occurs in a data-driven simulation model. It consists of a generic simulation model with a connected database, from which the model is initialised. In the concrete implementation of the experimentation environment the software SimChain is used (SimPlan AG 2015). The architecture of SimChain has been described by Gutenschwager and Alicke (2004). The measurement of the DWH-KPIs from the output data of the DES model is realised with a logic copy of the original DWH, the shadowed DWH (Dross and Rabe 2014). An input data ETL module (Extract, Transform and Load) is used to automatically obtain and process the required input data. The raw input data might be obtained from the DWH or from additional data sources, which can be files or other databases. After each simulation experiment, the output data ETL module queries the simulation output data from the respective output data tables of the simulation tool, processes the data according to the steps described in the procedure model and loads the data into the shadowed DWH. The metadata repository is a database in which all additional information for the processing of the input or output data is stored, e.g. the mapping information for missing attributes and the attributes themselves. Some verification and validation checks (V&V) can be partly automated, so the simulation experts are supported by the control software. Possible V&V checks are the validation of the collected raw input data, the verification of the

input data processing steps, the validation of the data in the metadata repository and the verification of the processing steps in the output data ETL module. The software automatically performs the implemented V&V checks and shows an alert, if some irregularities have been found. In this case, the simulation experts have to examine the data and find the reason for the failing V&V checks.

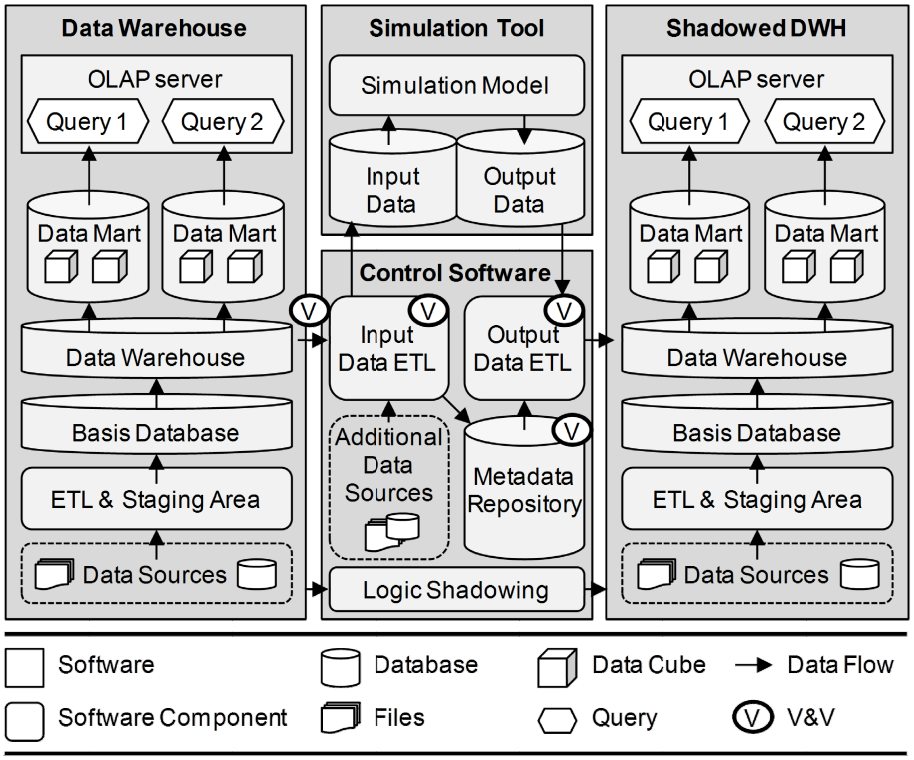

Figure 2: *Software architecture of the control software with V&V elements for the semi-automated updating of the simulation model*

5 Conclusion and Outlook

The procedure model described in this paper has already been successfully used by the authors in a simulation study. In this project it has already shown its usefulness and accurateness. The developed control software helped to automate some processing steps, but a manual V&V could not be completely substituted by the automated processes. In future projects, the authors will focus on advanced mechanisms for the automated V&V of the input and output data. This will be critical to achieve a sufficiently accurate measurement of the desired DWH-KPIs on the output data of the simulation model on a consistent basis. Only a credible measurability of DWH-KPIs will lead to a high acceptance of the simulation environment as a decision support instrument among the decision makers.

References

Banks, J.: The future of simulation software: A panel discussion. In: Andradottir, S.; Healy, K.J.; Withers, D.H.; Nelson, B.L. (eds.): Proceedings of the Winter Simulation Conference. Piscataway: IEEE 1997, pp. 1155–1162.

Banks, J.; Carson, J.S.; Nelson, B.L.; Nicol, D.M.: Discrete-event system simulation. Harlow (UK): Pearson 2014.

Baron, C.P.; Dietel, U.; Kreppenhofer, D.; Rabe, M.: Handlungsanleitung Simulation. In: Rabe, M.; Hellingrath, B. (eds.): Handlungsanleitung Simulation in Produktion und Logistik. San Diego (CA): SCS Int. 2001, pp. 117–190.

Bauer, A.; Günzel, H.: Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung. Heidelberg: dpunkt 2013.

Bengtsson, N.; Shao, G.; Johansson, B.; Lee, Y.T.; Leong, S.; Skoogh, A.; Mclean, C.: Input data management methodology for discrete event simulation. In: Rossetti, M.D.; Hill, R.R.; Johansson, B.; Dunkin, A.; Ingalls, R.G. (eds.): Proceedings of the Winter Simulation Conference. Piscataway: IEEE 2009, pp. 1335–1344.

Bernhard, J.; Jodin, D.; Hömberg, K.; Kuhnt, S.; Schürmann, C.; Wenzel, S.: Vorgehensmodell zur Informationsgewinnung: Prozessschritte und Methodennutzung. Technical Report 06008, TU Dortmund (Germany), 2007.

Bernhard, J.; Wenzel, S.: Information acquisition for model based analysis of large logistics networks. In: Merkuryev, Y.; Zobel, R.; Kerckhoffs, E. (eds.): Proceedings of the 19th European Conference on Modelling and Simulation. Riga (Latvia) 2005, pp. 37–42.

Csanady, K.; Bockel, B.; Wenzel, S.: Methodik zur systematischen Informationsgewinnung für Simulationsstudien. In: Rabe, M. (ed.): Advances in Simulation for Production and Logistics Applications. Stuttgart: Fraunhofer-IRB-Verlag 2008.

Dross, F.; Rabe, M.: A SimHeuristic framework as a decision support system for large logistics networks with complex KPIs. In: Wittmann, J.; Deatcu, C. (eds.): Proceedings of the 22nd Symposium Simulationstechnik, Berlin (Germany) 2014.

Ehmke, J.F.; Großhans, D.; Mattfeld, D.C.; Smith, L.D.: Interactive analysis of discrete-event logistics systems with support of a data warehouse. Computers in Industry 62 (2011) 6, pp. 578–586.

El-Darzi, E.; Vasilakis, C.; Chaussalet, T.; Millard, P.H.: A simulation model to evaluate the interaction between acute, rehabilitation, long-stay care and the community. In: Zanakis, S.H.; Doukidis, G.; Zopounidis, C. (eds.): Decision Making: Recent Developments and Worldwide Applications. Berlin, Heidelberg: Springer 2001, pp. 475–485.

Gutenschwager, K.; Alicke, K.: Supply Chain Simulation mit ICON-SimChain. In: Spengler, T.; Voß, S.; Kopfer, H. (eds.): Logistik Management. Heidelberg: Physica 2004, pp. 161–178.

Hill, J.; Onggo, S.: Data identification and collection methodology in a simulation project: An action research. In: Tjahjono, B.; Heavey, C.; Onggo, S.; van der Zee, D.-J. (eds.): Proceedings of the 6th Operational Research Society Simulation Workshop 2012, Birmingham (UK): The Operational Research Society 2012.

Hill, J.; Onggo, S.: Data identification and data collection methods in simulation: A case study at ORH Ltd. Journal of Simulation 8 (2014) 3, pp. 195–205.

Jarke, M.; Lenzerini, M.; Vassiliou, Y.; Vassiliadis, P.: Fundamentals of Data Warehousing. Berlin, Heidelberg: Springer 2003.

Jodin, D.; Kuhnt, S.; Wenzel, S.: Methodennutzungsmodell zur Informationsgewinnung in großen Netzen der Logistik. In: Buchholz, P.; Clausen, U. (eds.): Grosse Netze der Logistik. Berlin: Springer 2009.

Koutsoukis, N.-S.; Mitra, G.; Lucas, C.: Adapting on-line analytical processing for decision modelling: the interaction of information and decision technologies. Decision Support Systems 26 (1999) 1, pp. 1–30.

Law, A.M.; Kelton, W.D.: Simulation modeling and analysis. Boston (MA): McGraw-Hill 2000.

Lehtonen, J.; Seppala, U.: A methodology for data gathering and analysis in a logistics simulation project. Integrated Manufacturing Systems 8 (1997) 6, pp. 351–358.

McGinnis, L.F.: Technical and conceptual challenges in organizational simulation. In: Rouse, W.B.; Boff, K.R. (eds.): Organizational Simulation. Hoboken (NJ): Wiley 2005, pp. 273–298.

Pidd, M.: Computer simulation in management science. Hoboken (NJ): Wiley 2004.

Rabe, M.; Spieckermann, S.; Wenzel, S. (2008a) A new procedure model for verification and validation in production and logistics simulation. In: Mason, S.; Hill, R.; Mönch, L.; Rose, O.; Jefferson, T.; Fowler, J.W.; Ingalls, R.G. (eds.): Proceedings of the Winter Simulation Conference. Piscataway: IEEE 2008, pp. 1717–1726.

Rabe, M.; Spieckermann, S.; Wenzel, S. (2008b) Verifikation und Validierung für die Simulation in Produktion und Logistik: Vorgehensmodelle und Techniken. Berlin, Heidelberg: Springer 2008.

Robinson, S.: Simulation: The practice of model development and use. Hoboken (NJ): Wiley 2004.

SimPlan AG, 2015: SimChain. http://www.simchain.net/ accessed on 17.03.2015.

Skoogh, A.; Johansson, B.: A methodology for input data management in discrete event simulation projects. In: Mason, S.; Hill, R.; Mönch, L.; Rose, O.; Jefferson, T.; Fowler, J.W.; Ingalls, R.G. (eds.): Proceedings of the Winter Simulation Conference. Piscataway: IEEE 2008, pp. 1727–1735.

Vasilakis, C.; El-Darzi, E.; Chountas, P.: A data warehouse environment for storing and analyzing simulation output data. In: Ingalls, R.G.; Rossetti, M.D.; Smith, J.S.; Peters, B.A. (eds.): Proceedings of the Winter Simulation Conference. Piscataway: IEEE 2004, pp. 703–710.

Verein deutscher Ingenieure e.V.: Richtlinie 3633 Blatt 3 "Simulation von Logistik-, Materialfluß- und Produktionssystemen - Experimentplanung und -auswertung". Berlin: Beuth 1997.

Verein deutscher Ingenieure e.V.: Richtlinie 3633 Blatt 1 "Simulation von Logistik-, Materialfluss- und Produktionssystemen - Grundlagen". Berlin: Beuth 2014.

Vincent, S.: Input data analysis. In: Banks, J. (ed.): Handbook of Simulation. New York (NY): Wiley 1998, pp. 54–91.

Wenzel, S.; Collisi-Böhmer, S.; Pitsch, H.; Rose, O.; Weiß, M.: Qualitätskriterien für die Simulation in Produktion und Logistik: Planung und Durchführung von Simulationsstudien. Berlin, Heidelberg: Springer 2008.