INFORMATION ACQUISITION FOR MODEL BASED ANALYSIS OF LARGE LOGISTICS NETWORKS

Jochen Bernhard
Department of Simulation Concepts and Instruments
Fraunhofer-Institute for Material Flow and Logistics IML
44227 Dortmund, Germany
E-mail: bernhard@iml.fraunhofer.de
Sigrid Wenzel, Professor, Ph.D. Engineering
University of Kassel, Institute of Production Engineering and Logistics
Department of Mechanical Engineering
34125 Kassel, Germany
E-mail: s.wenzel@uni-kassel.de

KEYWORDS
Information Acquisition, Model Based Analysis, Simulation, Logistics Networks.

ABSTRACT

Bases of each efficient analysis of Large Logistics Networks are valid input data in the right quality, quantity and granularity. Considering these demands a goal and task oriented information acquisition by using several techniques of data acquisition, statistics and visualization is needed. This article gives an overview of current research results including theoretical definitions of terms like information, data, knowledge, a process oriented procedure model for information acquisition as well as different taxonomies of methods from data acquisition, statistics and visualization and their utilization. Finally, the article presents prospects of future research with the focus on the utilization of the information acquisition procedure model in applications.

INTRODUCTION

Large logistics networks LLN exist wherever a large number of different items are transported via several stations sometimes using alternating transport carriers. In accordance with delimitation and position, these logistics systems are called supplier networks, production networks, distribution networks, freight forwarding networks, container cycles, transportation networks etc. The components of these large networks (e.g. people, companies, resources, orders) are linked to each other by a multitude of various relationships.

The current developments in LLN lead to new structures and rules. Thus, in recent years, a multitude of new network-based enterprise structures such as alliances, joint ventures, supplier networks, fractal factories, segmented factories, virtual enterprises and process-orientated organization forms have been propagated and discussed.

These demands lead to added use of model based analysis and planning methods like simulation and optimization. In this context the input data in an appropriate quality, quantity and granularity are prerequisite for high-value results. Therefore, an interdisciplinary research team with partners from Data Acquisition, Statistics and Visualization decided to develop a methods utilization model for the gathering of qualitatively high-value input data.

In spite of or perhaps because of the current information overload the investment for a goal-oriented information acquisition in simulation projects is much too high with almost 50% of the total project time (Rabe and Hellingrath 2001). Therefore the logistics planner needs a toolbox for an efficient goal-oriented information acquisition, which offers a support of all planning activities.

Therefore, a primary research goal is the development and adaptation of procedure models for the acquisition of context based information for the determination of specific input data for the model based analysis (especially simulation and optimization) of LLN. The results lead to the development of a method toolbox for supporting the process of the information acquisition with respect to target group and task oriented method utilization.

DATA, INFORMATION, KNOWLEDGE

Any adequate approach to the acquisition of information for the modeling of logistics systems requires an exact differentiation between the terms data, information and knowledge and a distinct use of them. Mainly based on the information theory (Shannon and Weaver 1949; Nonaka and Takeuchi 1997) and an existing DIN standard (DIN 1995) following definitions were fixed:

Data consist of analog or digital signals or indications (syntax) and are used for the representation of information for the purpose of further processing.

Proceedings 19th European Conference on Modelling and Simulation
Yuri Merkuryev, Richard Zobel, Eugène Kerckhoffs © ECMS, 2005
ISBN 1-84233-112-4 (Set) / ISBN 1-84233-113-2 (CD)

Information covers statements and descriptions of characteristics and structures of certain objects or circumstances (semantic). Information is sent and received and due to this sender-receiver relation it is only valid for a specific purpose. Information is coded by data; it represents communicated and formalized knowledge and serves the increase of knowledge.

Knowledge ("mental models", skill, proficiency, know-how, experience, etc.) is bound to the consciousness and arises with the owner of the consciousness (individuals and/or collectives of individuals); knowledge is often difficult to formalize and hard to communicate.

These definitions for data and information built up the bases for all following sections.

INFORMATION ACQUISITION IN SIMULATION

The utilization of information acquisition in modeling depends basically on the purpose of modeling and the related modeling method. For example, if a purely descriptive model has to be developed, different information will be required than during an optimization task or a simulation. In the following, the information acquisition for simulation is focused.

The basic simulation study process is described in the Part 1 of the VDI-guideline 3633 (VDI 2000). The information acquisition represents a part step of this process and replaces the data acquisition. Input information for the procedure of information acquisition are the project task, the object of study, the determined investigation goal as well as the conceptual model (see Figure 1). The existence of a first conceptual model, which determines the model boundary and the planned abstraction level, is pre-requisite to carry out a purposeful and goal-oriented model information acquisition. The information acquisition itself delivers the concrete input data for the simulation model.

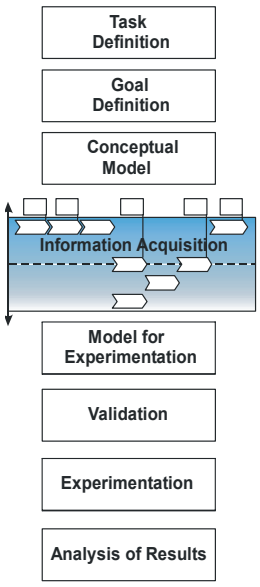

Figure 1: Information acquisition in the context of a simulation study.

PROCEDURE MODEL OF INFORMATION ACQUISITION

The procedure model of information acquisition links the information view to the data view in a process model. This includes the process steps of information collection and -estimation as well as the process steps of data management. Figure 2 gives an overview over the single process steps in their temporal sequence and in the respectively focused view of information and data. Where appropriate, iterations are necessary depending on the results of single process steps. For the processing of each process step several specific methods can be used.

The procedure model of information acquisition is based on the analysis of the given problem by considering the object of study, the aims of the study and the selected method of modeling. The *goal setting* process step is aided by taking characteristic information requirements of standard processes in logistics networks into account. After all this process step leads to principal information requirements for the given task and therefore for the model.

The following step *Identification of Information* is divided in three sub-steps: The *Analysis of Information Demand*, which derives the needed information from the principal information requirements, the *Analysis and Evaluation of the Existing Offer of Information*, including the detection and evaluation of all known and potentially useful sources of information, and the third sub-step *Assessment of the Available Information* which covers the comparison of the identified available information with the needed information by regarding aspects of expenses (e.g. availability, usability, costs, etc). This leads to the determination of potentially usable information.

The process step *Preparing the Collection of Information and Data* deals with the *Selection of Adequate Sources* and the *Identification and Selection of Appropriate Collection Methods* as well as the accomplishment of the corresponding preparation arrangements. With the aid of the evaluation of available information sources a decision is made on the sources to be used. Goal-oriented aspects of effort and benefit are the basis for the *Identification and Selection of Appropriate Collection Methods* to obtain an economical and goal-oriented utilization of the acquired information.

The *Collection of Information and Data* is composed of two sub-steps. By using the selected collection methods together with the chosen information sources the needed information is collected. Depending on the used collection method the collected information, for example as plain text, has to be transformed into corresponding data. In the next step the information and data has to be validated by the *Comparison of the Collected Data with the Needed Information*. This process step of information and data collection ensures that usable information with associated data is derived.

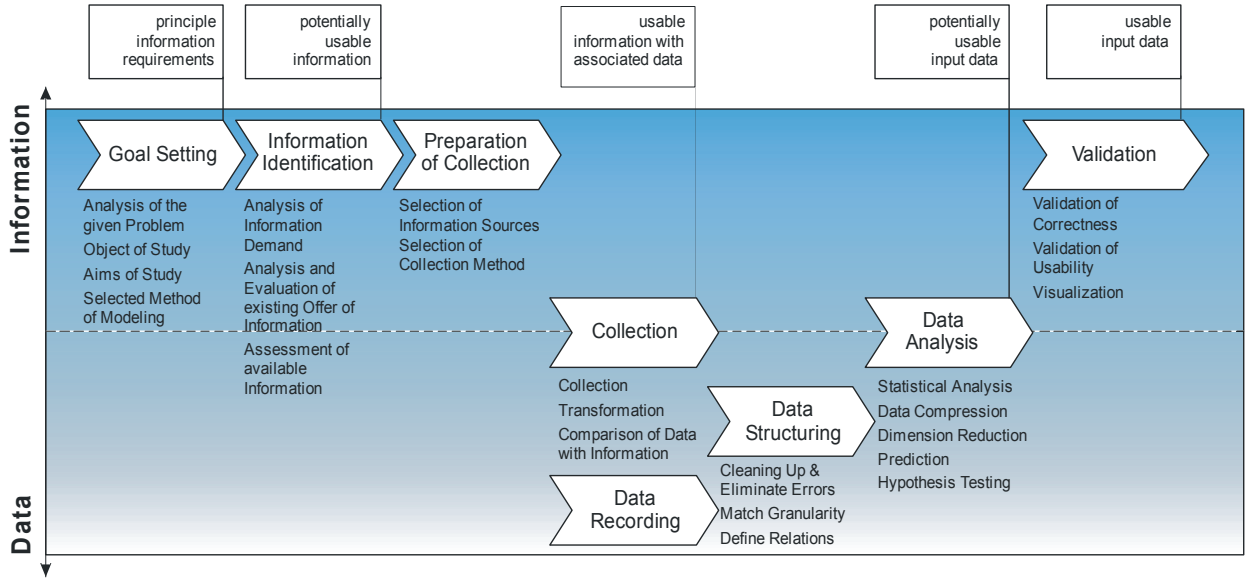

Figure 2: Procedure model of information acquisition.

Normally parallel to the process step described above the process of *Data Recording* includes all manually, semi-automatically or automatically operations to transfer data in a machine-readable structure and store them on a data medium. After this process step the data is available in digital form.

Data Structuring is a preprocessing step for the preparation of the recorded digital data for the following processes. This includes the clearing up, check on inconsistencies, as well as a scale-transformation taking the aims and the granularity level of the model based analysis into account. Also the definition of relations in data base systems by connecting indices and compatible variables leads to new logical rules.

The *Statistical Data Analysis* provides usable input data for the model based analysis of LLN. By using statistics methods for consistence checks (e.g. outlier identification, cross-validation), for reducing dimensions and complexity (e.g. variable selection, principal component analysis, classification methods and pattern recognition) or for a forecast the data are prepared, analyzed and interpreted with regard to quality and plausibility. On the one hand the data analysis works on the data, but on the other hand it depends directly from the previously identified information requirements.

The connected use of different statistical data analysis methods guarantees highly compressed and well prepared potentially usable input data only losing a minimum of information with respect to the collected data.

The *Validation* deals with the sufficient usability of the prepared data regarding the needed information and with its correctness and usage within the implemented simulation model. First, it has to be examined how much of the identified needed information is covered by the potentially usable input data. This validation process is supported by the usage of different methods of information visualization for the comparative and explorative analysis. In the second stage, the two steps of verification and validation of the potentially usable input data take place by regarding its correctness and usage within the implemented simulation model. These steps are equivalent to parts of the verification and validation process of the simulation model itself (VDI 2000). Finally the validation process yields usable input data for modeling, simulation and optimization of LLN.

TAXONOMIES FOR USING INFORMATION ACQUISITION METHODS

The utilization of the procedure model seen above necessitates enhancements for the purposive application of methods.

Therefore, parallel to the development of the procedure model different data acquisition, statistical and visualization methods were analyzed and classified regarding their use in the context of the information acquisition process. The worked out classification criteria represent the basis for the method taxonomies using in the different process steps to select a method for the concrete application.

In the area of the data acquisition, a comprehensive method screening was made for this both in internal areas of data acquisition discipline and in external areas, for example market research, development psychology and empirical social research. On basis of a direct cost-benefit-comparison the analyzed data acquisition methods were classified and rated with respect to their applicability for the data acquisition in LLN. Basis for the work within the range of the information and data acquisition is the distinction of the data acquisition methods in two method classes, the primary and the secondary collection. For primary collection

Criteria		Specifications							
Primary Function: Intention		Identification	Localization	Correlation	Association	Comparison	Structures and Patterns	Grouping	Classification
Type of Information to be Represented		qualitative	quantitative	qualitative and quantitative					
Relationship between Information-Objects		independent	relational	circular	hierarchical	network			
Information Encoding	Number of Viewable Information Components	1	2	3	4	5	6	7	n
	Dimension of Dependent Variables	nominal discrete	ordinal discrete	interval discrete	interval continuous	ratio discrete	ratio continuous		

Figure 3: Taxonomy for visualization methods from the information-related point of view

information and data are acquired directly for the investigation purpose, for example by means of questionnaires or interview as well as by self- and foreign-observation. A secondary collection is based on already existing data like external business data, internal business data collected for other purposes, official statistics etc. Known techniques are the inventory, the analysis of documents and the comparison of business companies. These primary and secondary methods have been analyzed by the authors especially in view of their usability to collect information on logistics systems (Hömberg and Jodin 2003). Goal-oriented aspects of effort and benefit are the basis for integrating the collection methods in category schemata. Only by choosing the most capable method, the economical and goal-oriented utilization of the acquired information can be achieved.

Based on multitude of different statistical analysis methods for basically different tasks and objectives, the applicable and useful groups of methods for the context of modeling of LLN were identified (Fender and Kuhnt 2003). The classification of a single method within a group by regarding the use in LLN could be carried out exemplary for the extensive method group Cluster Analysis. At this especially not only the complexity and quality of data, but also the data size becomes main criteria for the taxonomy.

An important step in the information acquisition process is the analysis of usable information and available data, whereby these often originate from different, heterogeneous data sources. The choice of suitable statistical analysis methods depends directly on the given goal and task, the kind of the available information, the structure of data records as well as the necessary granularity of the needed input data. For the multiplicity of different tasks and objectives as well as different data types in LLN usually a group of possible statistical methods exists for each area of application. Within these individual groups, different user-specific sources of error during selection and application of the suitable analysis method have to be considered, as for example the formulation of wrong assumptions, the neglecting and mistaking of relevant information as well as the deriving of wrong conclusions from an assumption. Due to these errors there also arises strengthened likewise error in input data for the simulation, so that the results of a simulation, based on this input data, have no force of evidence (Gather et al. 2000; Wenzel et al. 2001). Against this background groups of statistic methods could be identified, from which a suitable procedure can be selected, depending on the use within the information acquisition. Only within these selected groups, the taxonomy of the methods specific for the application in LLN can be made. This taxonomy and their criteria have to consider the complexity and quality of the data, the size of the available data sources as well as statistical quality.

For the development of the taxonomy of visualization methods, different visualization techniques used in the area of Scientific Visualization and Information Visualization (e. g. Chi 2000; Harris 1999; Robertson 1991) were identified and analyzed. Furthermore a catalog of criteria together with the criteria specifications was built up by using existing taxonomies. Also research results of Visual Data Mining have been taken into account. But the existing taxonomies only obtain graphical criteria or are built up from the viewpoint of the simulation in production and logistics (VDI 2003). Criteria from the information-related point of view, that offer the utilization of the methods with respect to the task and target group, previously were not considered. The developed taxonomy represents on the one hand the basis for the classification of application relevant visualization methods on the other hand the need inquiry from user view for the method selection (Wenzel et al. 2003).

In addition to the criteria from the graphical point of view the taxonomy for visualization methods is completed by a survey of the criteria from the information-related point of view. These are the primary

functions of the visualization method in terms of its potential use and the description of the information which can be represented and imparted by the chosen visualization method. The taxonomy is shown in figure 3, a detailed description of all criteria and their specifications can be found in (Wenzel et al. 2003).

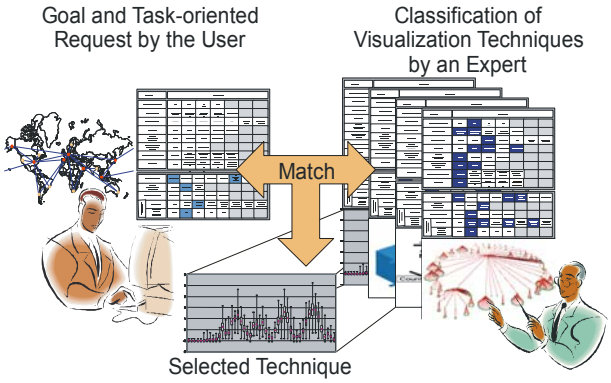

Figure 4: Taxonomy utilization for method selection.

The taxonomies were developed to match the requirements of a user with the specific characteristics of data acquisition, statistical or visualization methods. The precondition is the complete classification of all methods which are relevant for model-based planning of LLN. The user specifies his/her requirements with regard to task and purpose and other requirements, e. g. working field conventions, by means of the same taxonomy. However, this does not require a complete specification. By comparing his/her requirements with the entirety of available methods the user is led to a category of possible methods or, ideally, to just one method (see Figure 4). This comparison could be realized by an appropriate software program. The final selection has to be done manually by the user himself on the basis of extended criteria such as preparation effort and interpretation effort of the target group.

STANDARD LOGISTICS PROCESSES AND DATA FOR INFORMATION IDENTIFICATION

To assist the process step *Information Identification* of the procedure model a systematic and standardized classification, description and evaluation of the needed input data in the context of the logistics processes are meaningful and necessary. For this reason typical standard processes on different levels of abstraction were and still has to be identified and - on the assumption that the simulation is used as analysis method - characteristic input data for the analysis and planning were assigned to logistics processes. Thereupon the relevant data for the modeling with the identified standard processes were selected and the needed quality, quantity and granularity was defined (e.g. topicality, accuracy, view period etc.). The modeling of the LLN based on the process chain paradigm (Kuhn 1995 and VDI 2001) and the use of these standard processes with the assigned data leads to an aggregating of all necessary input data on the necessary level.

FUTURE PROSPECTS

The future research activities focus on a comprehensive definition and evaluation of the classification criteria for data acquisition, statistical and visualization methods. These classification criteria serve the development of method categories for simplifying the selection and utilization of specific methods. Also the derivation and definition of information categories, which are directly diverted from identified standard processes of logistics, as well as the generalization of the procedure model for the scope of LLN and not only for single applications are in the focus of the future work. The integrative methods utilization by using the procedure model has to be expanded with respect to a synergetic utilization of the methods. This requires the recognition of use and application specific dependencies of the methods of data acquisition, statistical analysis and visualization among each other. These dependencies have to be formulated as meta-information. In this context, especially several quality criteria are useful to get appropriate information as well as the corresponding input data.

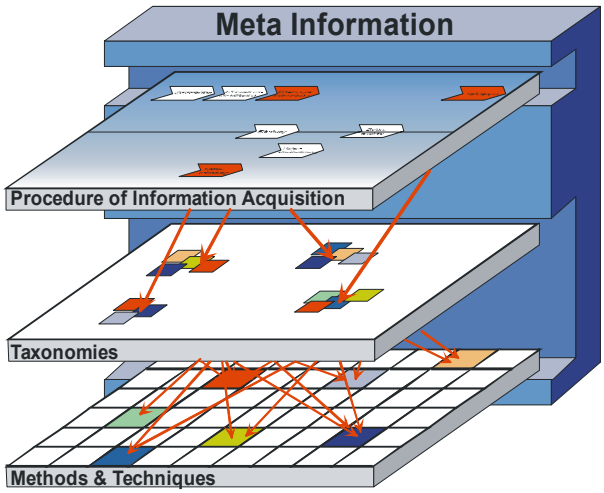

Figure 5: Methods Utilization Model.

The described results complement the methods toolbox and - together with the procedure model – lead to an integrated methods utilization model (see Figure 5), to carry out a purposeful information acquisition based on a given task definition and analysis method.

ADDITIONAL INFORMATION

Additional information can be found under <http://www.sfb559.uni-dortmund.de>.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center "Modelling of Large Logistics Networks" (559).

REFERENCES

Chi, E. H. 2000. A Taxonomy of Visualization Techniques using the Data State Reference Model. In: Proceedings of the IEEE Symposium on Information Visualization. IEEE Computer Society: Washington, DC, USA, 69-75.

DIN Deutsches Institut für Normung e.V. 1995. Informationsverarbeitung – Part 1. DIN44300-1. Beuth, Berlin.

Fender, T. and S. Kuhnt. 2003. Statistisches Datenmanagement zur Informationsgewinnung in GNL. Technical Report – Collaborative Research Center 559 „Modellierung großer Netze in der Logistik“ 03006. University of Dortmund, ISSN 1612-1376.

Gather, U.; C. Becker and S. Kuhnt. 2000. Robust Methods for Complex Data Structures. In: Kiers, H. A. L.; J.-P. Rasson; P. J. F. Groenen and M. Schader. (Eds.). Data Analysis, Classification and Related Methods. Springer, Berlin Heidelberg New York, 315-320.

Harris, R. L. 1999. *Information Graphics. A Comprehensive Illustrated Reference. Visual Tools for Analyzing, Managing, and Communicating.* Oxford University Press.

Hömberg, K. and D. Jodin. 2003. Methodische Datenerhebung für effiziente Planungs- und Simulationsprojekte in der Logistik. In: Material Management (2003) 5. Grütter, Coburg/Paderborn. 20-25.

Kuhn, A. 1995. Prozessketten in der Logistik – Entwicklungstrends und Umsetzungsstrategien. Verlag Praxiswissen, Dortmund.

Nonaka, I. and H. Takeuchi. 1997. Die Organisation des Wissens. Campus, Frankfurt, New York.

Rabe, M. and B. Hellingrath. 2001. Handlungsanleitung Simulation in Produktion und Logistik. SCS. San Diego.

Robertson, P. K. 1991. A methodology for choosing data representations. *IEEE Computer Graphics & Applications* 11 (1991) 3, 56-67.

Shannon, C. E. and W. Weaver. 1949/1976. Mathematische Grundlagen der Informationstheorie. Oldenbourg Wirtschaftsverlag GmbH, München.

VDI Verein Deutscher Ingenieure. 2000. Simulation von Logistik-, Materialfluss- und Produktionssystemen. Guideline 3633 Page 1. Beuth, Berlin.

VDI Verein Deutscher Ingenieure. 2001. Processes and process orientation in production logistics Example: Automotive Industry. Guideline 3600. Beuth, Berlin.

VDI Verein Deutscher Ingenieure. 2003. Simulation und Visualisierung. Guideline 3633 Page 11. Beuth, Berlin.

Wenzel, S.; C. Becker and D. Jodin. 2001. Methoden der Informationsgewinnung zur Bestimmung des Eingangsdatenraums für Simulationsmodelle in der Logistik. In: Panreck, K.; and F. Dörrscheidt (Eds.). Simulationstechnik. Proceedings of the 15th Symposium ASIM 2001. SCS-Europe BVBA, Ghent. 417-422.

Wenzel, S.; J. Bernhard and U. Jessen. 2003. A Taxonomy of Visualization Techniques for Simulation in Production and Logistics. In: Chick, S.; P.J. Sanchez; D. Ferrin and D.J. Morrice (Eds.). Proceedings of the 2003 Winter Simulation Conference. IEEE, New Orleans, Louisiana, USA. 729-736.

AUTHOR BIOGRAPHIES

JOCHEN BERNHARD was born in 1970 and studied mechanical engineering at the University of Kassel, Germany. After his graduation he worked at the Volkswagen AG and became an employee at the Fraunhofer Institute for Material Flow and Logistics in 2001. Since May 2004 he is leading the team Simulation Concepts and Tools. His main activities include the development and realization of concepts as well as the practical application of systems in simulation, planning and training. Emphases of his work are elaborating concepts for the coupling of simulation systems, the use of information visualization methods in simulation, collaborative engineering and internet-based learning. In addition to the development and management of national research and industry projects, he is an active member of the ASIM (Arbeitsgemeinschaft Simulation) group Simulation in Production and Logistics and the technical committee for Simulation and Visualization, a section of Committee A5 Modeling and Simulation of the Association of Engineers Society for Materials Handling, Materials Flow and Logistics Engineering (VDI-FML). You can reach him by e-mail at <bernhard@iml.fraunhofer.de>.

Prof. Dr. **SIGRID WENZEL** studied computer science at the University of Dortmund. She was employed at the Department for Material Handling and Warehousing at the University of Dortmund from 1986 to 1989. Since 1990, she has been employed at the Fraunhofer Institute for Material Flow and Logistics in the area of simulation concepts and tools as well as in the area of visualization techniques for production and logistics. She was head of the department from 1992 to 2004, acting department head of the division of enterprise modeling from 1995 to 2004 and additionally manager of the Collaborative Research Center 559 - Modeling of Large Logistics Networks at the University of Dortmund from 2001 to 2004. Since May 2004 she is professor and head of the Department of Mechanical Engineering, University of Kassel, Institute for Production Engineering and Logistics. In addition to this, she is a board member of the Arbeitsgemeinschaft Simulation (ASIM), spokesperson for the ASIM group Simulation in Production and Logistics, head of the technical committee for Simulation and Visualization and an active member of the technical committee for Modeling Processes and for Digital Factory, sections of Committee A5 Modeling and Simulation of the Association of Engineers Society for Materials Handling, Materials Flow and Logistics Engineering (VDI-FML). You can reach her by e-mail at <s.wenzel@uni-kassel.de>.