Proceedings of the 2009 Winter Simulation Conference
M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, eds.

INPUT DATA MANAGEMENT METHODOLOGY FOR DISCRETE EVENT SIMULATION

Nils Bengtsson
Production Modeling Corporation
Gothenburg, 411 04, SWEDEN

Guodong Shao
Y. Tina Lee
Swee Leong
Charles McLean
Manufacturing Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, USA

Björn Johansson
Product and Production Development
Chalmers University of Technology
Gothenburg, 412 96, SWEDEN

ABSTRACT

Input Data Management (IDM) is a time consuming and costly process for Discrete Event Simulation (DES) projects. In this paper, a methodology for IDM in DES projects is presented. The approach is to use a methodology to identify and collect data, then use an IDM software to extract and process the data. The IDM software will structure and present the data in Core Manufacturing Simulation Data (CMSD) format which is a standard data format for any DES software. The methodology is implemented in a project at National Institute of Standards and Technology (NIST) in collaboration with an aerospace industry partner.

1 INTRODUCTION

Discrete Event Simulation (DES) has proved itself to be an effective tool for complex processes analysis (Ericsson 2005; Banks et al. 2000). A problem of using DES is the required efforts and costs spent on processing input data from various data sources in order to ensure valid simulation results. Large amounts of time in a DES project is needed for gathering and extracting data (Skoogh and Johansson 2007). Most of the time, the needed information can be found in various Information Technology systems (IT-systems) in the companies. However, data is usually not in the right format required for DES and IT-systems do not have a standardized way of communicating with each other. This makes it hard to integrate several IT-systems or DES software, customized interfaces for exchanging information are often needed to be developed. A reusable, neutral, standardized interface should help reduce the effort and cost related to Input Data Management (IDM) in DES projects (Johansson et al. 2007). Researchers at the National Institute of Standards and Technology (NIST) have developed Core Manufacturing Simulation Data (CMSD) (SISO 2009) to create a neutral format between common production software applications and DES tools. The concept has already been tested in pilot implementations in some case studies (Heilala et al. 2008, Johansson, and Zachrisson 2006, Johansson et al. 2007). For example, the CMSD was used to generate input data that can be reused for DES models developed using both Enterprise Dynamics (ED) and Plant Simulation (Johansson et al. 2007). The Generic Data Management Tool (GDM-Tool) has been developed at Chalmers University of Technology (Chalmers) to help structuring input data and reducing the time needed for extracting the data from various data sources. The GDM- Tool makes it possible to reuse connections to databases and the pre-configured data from the prior simulation run. It can also write data into the CMSD format if needed (Balderud and Olofsson 2008).

The industrial partner in the context of this paper often experiences disturbances in production and material handling processes where parts are missing or lost. They need to spend time searching for the missing parts, and sometimes need to re-order the lost parts. This problem has incurred big impacts and causes delays in the manufacturing processes and complicates even more by the global supply chain. In order to minimize the disturbances, the industrial partner is investigating potential investments in new tracking technology. To support the decision making for the investment, a DES model is developed and used to analyze the different scenarios. Input data for the simulation model is needed from many different production planning and control systems within the company.

In order to manage all the input data needed for the DES model, a methodology for the IDM has been proposed. The methodology involves how to handle the input data with a structured methodology, using GDM-Tool and CMSD, as well as the relationships in-between them. GDM-Tool processes the data, identifies distribution functions for the data set, and converts the data to CMSD format. The CMSD provides a data specification for efficient exchange of manufacturing life-cycle data in a simulation environment. Once the data is generated in CMSD format, it can be re-used as input for the different DES software.

Although each module has been implemented in various projects by researchers at NIST and Chalmers (Skoogh and Johansson 2008; Johansson et al. 2007, Balderud and Olofsson 2008), this is the first time GDM-Tool and CMSD have been used together as an IDM methodology.

The methodology is introduced in section 2. A prototype implementation with an industrial partner in aerospace industry is discussed in section 3. Discussion and a summary of future work are presented in section 4.

2 PROPOSED METHODOLOGY FOR IDM

The proposed methodology includes the input data collection, GDM-Tool, CMSD, and DES modules as showed in Figure 1. Data is being identified, located, and collected first. During the same time the connection between the different data sources and GDM-Tool is being set up. GDM-Tool has several plug-ins to process data including deleting and adding data elements, identifying distribution functions for the data set, and writing the data to CMSD format. By using GDM-Tool, it is easy to re-use the same data for other DES projects at the same company, for example, the connections to the databases are already in place from the last project; GDM-Tool will only need to execute the same sequence of plug-ins or add new data for the new project. Once the data has been generated in CMSD format, it can be re-used as input for any DES software within the company. More details for each module are discussed in the following subsections.

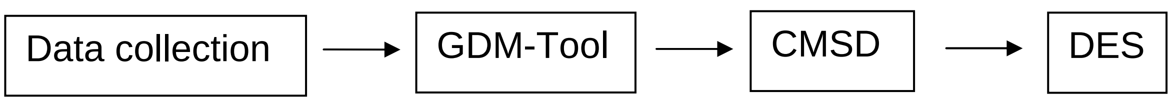

Figure 1: Overview of proposed methodology for IDM

2.1 Discrete event simulation

A DES model can be used to analyze what-if scenarios for a real world problem and provide valuable information for decision makers, for example, with a DES model of manufacturing system, the changing requirements from new production equipment, new layout for a factory, or a new transport system can be evaluated before the investment is made. Bottleneck analysis can also be performed. DES model can give the answer where the bottleneck of the system is located at specific times (Roser et al. 2001). The results from the simulation will be very similar for the real system, if the model is validated correctly (Banks et al. 2000).

To increase the chance of having a successful DES project, several activities need to be conducted. Banks et al. (2000) describes a methodology similar to the one shown in Figure 2. The main difference is higher detail in the startup and ending phases of the project. The framework has also been discussed in detail and applied in a DES Master Thesis in Sweden (Bengtsson and Palander 2007).

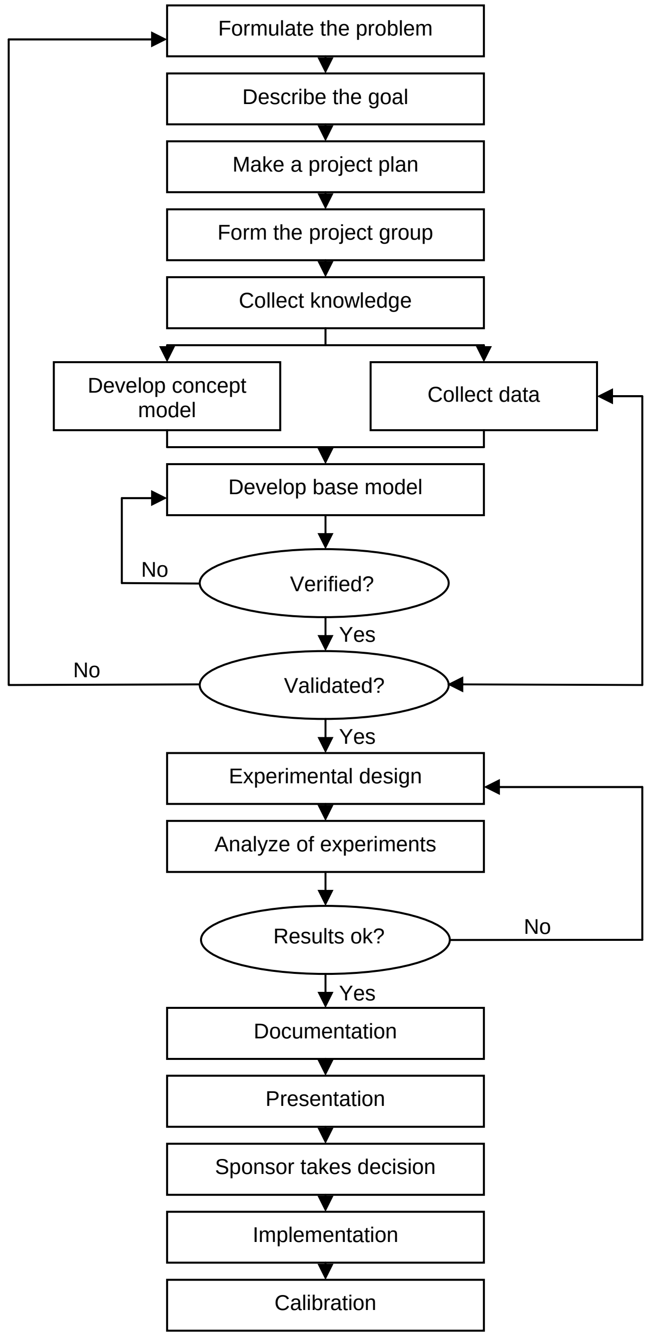

Figure 2: Overview framework for simulation project

In a DES project it is very important to fully understand the problem and document it together with how the modeled system works with complete logic. It is vital to have the sponsor to agree on the problem to solve and goal of the project before starting to develop the model,. The main tasks includes formulating the problem, describing the goal, and collecting knowledge and data.

The input data is critical for a DES model to be useful. If the input data is not correct, the results will not be useful, trash in will only have trash out no matter how fancy the model is. Data collection is a time consuming effort. In general, around 31 % of the simulation project time (Skoogh and Johansson 2007) is spent on data collection. If this percentage could be lower, it would be possible to decrease the cost for DES projects. A major part of the total simulation project time is spent on addressing interoperability problems due to the fact that data is not stored in the right format and it's time consuming to extract the data (Skoogh and Johansson 2007; Johansson et al. 2007).

2.2 Input Data collection

The framework presented in Figure 2 is not detailed enough in regards to the data collection effort in a DES project. Data collection is a critical step in every DES project, the data quality is an important factor for the model validity. A data input processing methodology has been developed for DES projects by Skoogh and Johansson (2008). The methodology consists of the following 13 activities:

- Identify and define relevant parameters
- Specify accuracy requirements
- Identify available data
- Choose methods for gathering of not available data
- Will all specified data be found?
- Create data sheet
- Compile available data
- Gather non available data
- Prepare statistical or empirical representation
- Sufficient representation
- Validate data representations
- Validated?
- Finish final documentation

The goal of the methodology is to reduce the time and effort in data collection and the subsequent processing effort required of the input data that contribute to the higher cost for a DES project. For detail about the methodology, see the referred paper by Skoogh and Johansson (2008).

2.3 GDM-Tool

In an effort to reduce the time needed to develop DES models by decreasing the time for input data management, Chalmers and NIST have started to develop an expandable software based architecture for Generic Data Management. The effort is still on-going with pilot implementations, and improvements of the first version of GDM-tool initially developed by Balderud and Olofsson (2008). GDM-Tool is aimed to link production data stored in different IT-systems at companies into the DES models and convert the data according to the CMSD data structures and specifications. The software works by associating a sequence of small plug-ins together for execution. A typical sequence is reading in the data, processing it, writing it in CMSD data structures, and then outputting the data; as shown in Figure 3.

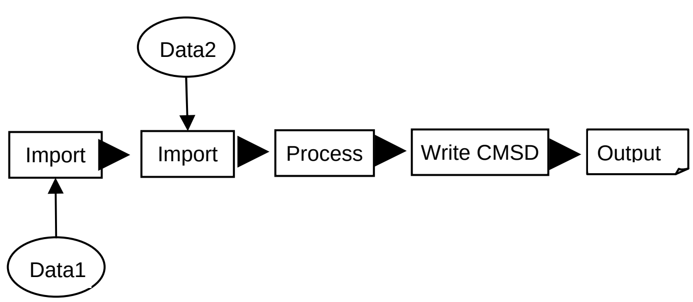

Figure 3: Overview of the plug-in structure with a typical sequence of Plug-ins

The plug-in based architecture in the software makes it flexible and easy to maintain. The goal is to have an extensive library with plug-ins enabling an easy to use interface that does not require the user to have programming skills in order to operate the software in the future. The user can operate the plug-ins by drag and drop features in GDM-Tool.

The GDM-Tool can view the processed data in various ways using graphical user interfaces. As of the time of writing this paper, There are 22 default plug-ins being developed with simple data processing, statistical functionality, and CMSD output. It currently can identify and fit data using the following distribution functions:

- Exponential
- Gamma
- Log-Normal
- Weibull

The distribution identification is based on Maximum Likelihood Estimation, Kolmogoroc-Smirnov test statistics, and Calculation of P-value which are all proven statistical methods. Manual choice based on P-value against P-value plots of the data can be also used to identify a distribution.

The CMSD output is implemented by using an internal function called Tags in GDM-Tool that is similar to a label. The Tags have all the information about each data element and where it should be placed in the CMSD structure. The generation of the XML files according to the CMSD structure is done by using the information in the Tags (Balderud and Olofsson 2008).

2.4 Core Manufacturing Simulation Data

Information management problems affect many aspects of manufacturing operations, but they are a particular hindrance to the creation and reuse of manufacturing simulations. NIST researchers, in collaboration with industrial partners, have been working on a standards development effort titled Core Manufacturing Simulation Data (CMSD) (SISO 2009) under the guidelines, policies, and procedures of the SISO (SISO 2008).

The CMSD specification describes a CMSD information model using the Unified Modeling Language (UML) (UML 2009) as the modeling language. The primary objective of this information model is to provide a data specification for efficient exchange of manufacturing life-cycle data in a simulation environment. The objective leads to:

- Foster the development and use of simulations in manufacturing operations
- Facilitate data exchange between simulation and other manufacturing software applications
- Enable and facilitate better testing and evaluation of manufacturing software
- Increase manufacturing application interoperability.

The CMSD information model addresses issues related to information management and manufacturing simulation development and provides a means to define information about many kinds of manufacturing objects. It facilitates the exchange of information between manufacturing-oriented simulations and other applications in manufacturing domains such as process planning, scheduling, inventory management, production management, and plant layout. The information model is not intended to be an all-inclusive definition of either the entire manufacturing domain or simulation domain. The model describes the essential or core entities in the manufacturing domain and the relationships between those entities needed to create manufacturing-oriented simulations.

The CMSD information model's UML representation has been organized using packages. UML packages, depicted as file folders, are UML constructs that can be used to organize model elements into groups. The CMSD information model consists of the following major UML packages:

- Layout
- Part Information
- Support
- Resource Information
- Production Operations
- Production Planning.

For more information and detailed definition about CMSD, se the referred article (SISO 2009).

3 PROTOTYPE IMPLEMENTATION IN AEROSPACE INDUSTRY

3.1 Project description

The industrial partner has experienced disturbances in production due to parts are getting misplaced or even lost upon delivery to their intended destinations. Enormous amounts of parts are handled on the factory floor on a daily basis, missing and lost parts can cause big problems to production schedules. The company is considering investing in new tracking equipment in order to have better tracking and control of parts within their processes. To assess the viability and business case of the new tracking equipment, a DES model is being built in order to aid the decision-making on how to tackle this problem with missing and lost parts. For input data for the DES model, the company has various manufacturing information, in varying format, stored in numerous databases within their organizations. To prepare the input data for the DES model, extraction and processing of the relevant manufacturing information from these databases will be required. This input data management will be both time-consuming and work extensive. However, the process is required every time data is updated in the database.

Another purpose of the project is to evaluate if CMSD can help bridge the IT-systems, databases, and the simulation together and ease the effort with extracting and handling large amounts of data. The evaluation will be conducted by using CMSD format as input data to a simulation model. The model is developed in Arena and represents the parts delivery and material handling processes at one section of the factory. The modeling result will be used for analyzing whether it is economically justifiable to invest in new identification and tracking technology for the processes.

The project is also a continuation on the work that has previous been done (Johansson et al. 2007; Heilala et al. 2008; Johansson and Zachrisson 2006; Riddick and Lee 2008) in evaluating and testing the CMSD information model with real world production scenarios, in order to further developing and validating the CMSD standard developing efforts.

Arena from Rockwell Automation was chosen as the software for the baseline simulation model. Arena's capability to handle Visual Basic for Applications (VBA) code simplifies the process of reading in the Extensible Markup Language (XML) document. The CMSD allows XML as the output structure. Use of the tools that are provided in Microsoft .Net framework (Microsoft, 2009) shortens the time needed to develop the CMSD interface in Arena. Other DES software's with XML reading capabilities could also be used in this project.

The GDM-Tool together with the CMSD information model is used for structuring the input data for the simulation model. The approach is that GDM-Tool is used for connecting to the databases, extracting the data, and reformatting the data according to the CMSD structures. An overview of the data flow diagram is presented in Figure 4.

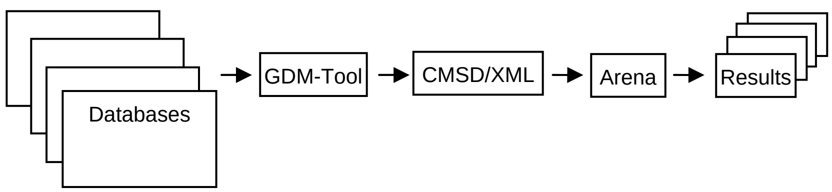

Figure 4: Prototype data flow diagram

3.2 Data

By using GDM-Tool to extract and process the data from the databases, it is easy to provide updated data with the latest production information for every simulation run. The current available production input data is in Microsoft Excel (Excel) format and needs to be processed before it can be used in the Arena model. The input data that need to be mapped into the CMSD structures for used in the Arena model includes:

- Arrival rates of different orders at the delivery locations
- Percentages for missing parts, lost parts, and parts found
- Times for searching missing parts
- Times for declaring missing parts as lost part after some time has been spent in searching for them
- Travel time for parts

The work with identifying and collecting the required data for the simulation model is being carried out by the industrial partner. This is done through interviewing key personnel working in the processes, accessing and assessing the data in databases, and determining the data format suitable for the GDM-Tool.

3.3 Simulation logic

Due to the fact that there are a lot of human factors played in the search process of lost products, it is very complicated to model it. For example, how can logic of where to start searching for products at the shop floor be formulated? The approach discussed in this paper was to focus on events after the parts had been delivered to their delivery location, when mechanics ready to use the parts but they could not locate the parts. the model is developed using the historical best estimated time to search for the missing parts by the plant floor personnel, without considering where the searches and the orders of searched-locations have been made in the facility. The positive side with this approach is that the search logic becomes easier to handle. The challenge is to find good distributions to model the search time. This approach satisfies the objective of the industrial partner.

There are many different scenarios to handle when a part is getting misplaced or lost in the process. A simplified version of simulation logic is showed in Figure 5. The logic gives three scenarios for a required part. In the first scenario, it can be delivered on time at the right place. This is the normal scenario that would not cause any delay for the production. The second scenario is that the part is misplaced, resources are needed to search for it, but the part can be found within a specified time. The last scenario is that the part is misplaced and could be not found in a declared time limit, the part is declared lost and needs to be reordered.

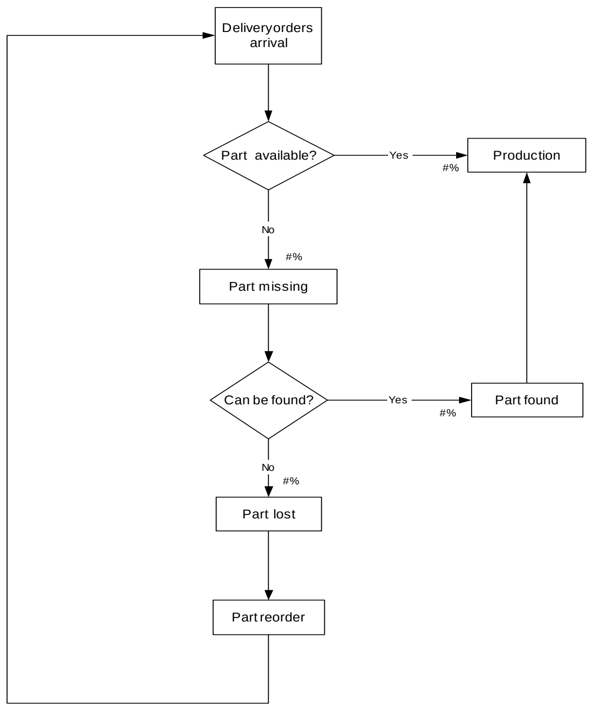

Figure 5: Schematic figure of search logic for different scenarios

3.4 Simulation output result

The result of the simulation is displayed in real time during the execution of Arena model; Figure 6 presents a sample simulation output. The focus is on presenting the results using the industrial partner's local vocabulary to describe the output rather than a general description. Though the simulation could run for years with the data and assumptions representative of the process model, the results show the total statistics for a period, in this case one month. It gives an overview on how many orders that have been handled in the simulation run and how successful the deliveries have been. It also gives the statistics on how many products have been missing and lost, and what affects that has had on delays within the factory.

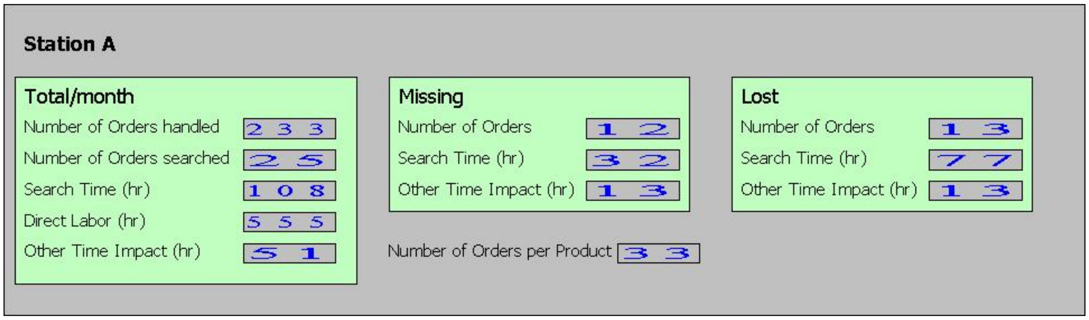

Figure 6: Sample output from the DES Arena model

Results on a monthly basis in the statistics are also provided in a Excel Spreadsheet that enables further processing of the output data. Excel also gives availability to customize the results further with graphs. An example from the result in Excel is shown in Figure 7.

Non RFID	Total		Missing/Found		Orders Lost/Re-order Parts
Delivery station A					
Total orders/mo	233		12		13
total orders/yr					
Search Time (hrs)	108		32		77
Direct Labor (hrs)	555				
Other Time Impact (hr)	51		13		13

Figure 7: Sample output from the DES Arena model in Excel

Both Arena and Excel report give the same results and data, but Arena only displays the current data for the simulation time while Excel stores results from specified time intervals. That is why Excel is used when analyzing behavior over time in the simulation model.

4 CONCLUSION AND FUTURE WORK

The methodology proposed and implemented in this paper addresses the problems related to input data for DES projects. The methodology is an integration of several novel concepts, tools, and specifications. The results so far showed that the methodology has a potential helping to decrease the time and cost related to IDM in DES projects.

The GDM-Tool is still under development for further enhancing its capabilities. The current efforts include supporting for the whole CMSD specification and adding more plug-ins, as well as the integration with the methodology for input data management.

The proposed methodology will be further tested in the ongoing project at NIST in collaboration with the industrial partner in the aerospace industry. Future projects are in the pipeline for planning and execution. These projects provide further opportunity to test, evaluate, improve, and develop the input data management methodology, the GDM-Tool, and the CMSD specification.

5 DISCLAIMER

The work described was funded by the United States Government and is not subject to copyright. No approval or endorsement of any commercial product by the National Institute of Standards and Technology is intended or implied.

REFERENCES

Balderud, J., and A., Olofsson. 2008. *A Plug-in Based Software Architecture for Generic Data Management*, Master's thesis, Department of Computer Science and Engineering. Chalmers University of Technology, Gothenburg.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2000. *Discrete-event system simulation*. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Bengtsson, N., and E., Palander. 2007. *Discrete Event Simulation of a line for Assembly of Office Chairs – Development of a simulation model at Kinnarps Production AB*, Master's thesis. Department of Product and Production Development. Chalmers University of Technology, Gothenburg.

Ericsson, U. 2005. *Diffusion of Discrete Event Simulation in Swedish Industry, One way to an Increased Understanding*, PhD Thesis, ISBN 91-7291-577-3, Department of Materials and Manufacturing Technology, Chalmers University of Technology, Gothenburg, Sweden

Johansson, M., and R. Zachrisson. 2006. *Modeling automotive manufacturing process.* Master's thesis. Department of Product and Production development. Chalmers University of Technology, Gothenburg.

Johansson, M., S. Leong, Y. T. Lee, F. Riddick, G. Shao, B. Johansson, A. Skoogh, and P. Klingstam. 2007. *A Test Implementation of the Core Manufacturing Simulation Data Specification*, In Proceedings of the 2007 Winter Simulation Conference, S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds., 1673-1681. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Heilala, J., S. Vatanen, H. Montonen, J. Tonteri, B. Johansson, J. Stahre and S Lind. 2008. Simulation-Based Sustainable Manufacturing System Design. In *Proceedings of the 2008 Winter Simulation Conference.* S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds., Miami, FL, USA

Microsoft. 2009, Overview of XML reading, available via <http://support.microsoft.com/kb/301225> [Access 5 April, 2009]

Riddick, F, and Y. T. Lee. 2008. Representing Layout Information in the CMSD Specification. In *Proceedings of the 2008 Winter Simulation Conference.* S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds., Miami, FL, USA.

Roser, C., M., Nakano, and M. Tanaka. 2001. A Practical Bottleneck Detection Method. In *Proceedings of the 2001 Winter Simulation Conference.* B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, Arlington, Virginia, USA:

Simulation Interoperability Standards Organization (SISO) Executive Committee. 2008: SISO Polices and Procedures, SISO-ADM-002-2008, 2008.

SISO 2009. Core Manufacturing Simulation Data information model part 1: UML model. Draft Product Development Group, Simulation Interoperability Standards Organization. Available via <http://discussions.sisostds.org/default.asp?action=9&boardid=2&read=39532&fid=24> [Accessed April 13, 2009]

Skoogh, A., and B. Johansson. 2008. A Methodology for Input Data Management in Discrete Event Simulation Projects. In *Proceedings of the 2008 Winter Simulation Conference.* S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds., Miami, FL, USA.

Skoogh, A., and B. Johansson. 2007. Time Consumption Analysis of Input Data Activities in Discrete Event Simulation Projects. In *proceedings of the 1st Swedish Production Symposium*. Gothenburg, Sweden.

UML Resource Page: Unified Modeling Language. Available via <http://www.uml.org/> [accessed January 3, 2009].

AUTHOR BIOGRAPHIES

NILS BENGTSSON is an applications engineer at Production Modeling Corporation Europe AB, he graduated with a Masters and a MSc. Degree from Chalmers University of Technology 2007. He is currently working as a guest researcher at National Institute of Standards and technology. His research interest includes discrete event simulation applicability and interoperability for industry. His email is <nbengtsson@pmcorp.com>.

GUODONG SHAO is a computer scientist in the Manufacturing Simulation and Modeling Group in The National Institute of Standards and Technology (NIST) Manufacturing Systems Integration Division. He has participated in research relating to FMS, CIMS, and manufacturing simulation integration for many years. He holds a Master's Degree from University of Maryland at College Park. He is a Ph.D. student at George Mason University. His email is <gshao@nist.gov>.

BJÖRN JOHANSSON is an Assistant Professor at Product and Production development, Chalmers University of Technology, currently also a guest researcher at National Institute of Standards and Technology in Gaithersburg, Maryland, USA. His research interest is in the area of discrete event simulation for manufacturing industries, modular modeling methodologies, environmental effects modeling, software development, user interfaces, and input data architectures are examples of interests. His email address is <bjorn.johansson@chalmers.se>.

Y. TINA LEE is a computer scientist in the Manufacturing Simulation and Modeling Group at NIST. She joined NIST in 1986. Her major responsibility in recent years has been to develop information models to support various manufacturing application areas. Previously she worked at the Contel Federal Systems and at the Sperry Corporation. She received her BS in

Mathematics from Providence College and MS in Applied Science from the College of William and Mary. Her e-mail address is <leet@nist.gov>.

SWEE LEONG is a senior manufacturing engineer in the Manufacturing Simulation and Modeling Group at the National Institute of Standards and Technology (NIST) Manufacturing System Integration Division since 1994. Prior to joining NIST, he worked at Ford Motor Company, John Deere, and IBM on different factory automation projects. His research interests are in modeling and simulation applications for the manufacturing industries and engineering tools integration program at NIST. Currently, Swee man-ages the Simulation Standards Consortium at NIST. He is chairman of the Core Manufacturing Simulation Data Product Development Group. He received a Bachelor and Master Degree in Industrial Engineering from Purdue University in West Lafayette, Indiana. He is a senior member of the Society of Manufacturing Engineers. His e-mail ad-dress is <leong@nist.gov>.

CHARLES MCLEAN is a computer scientist and Group Leader of the Manufacturing Simulation and Modeling Group at the National Institute of Standards and Technology. He has managed research programs in manufacturing simulation, engineering tool integration, product data standards, and manufacturing automation at NIST since 1982. He has authored more than 50 technical papers on topics in these areas. He is on the Executive Board of the Winter Simulation Conference and the Editorial Board of the International Journal of Production, Planning, and Control. He is formerly the Vice Chairman of the International Federation of Information Processing (IFIP) Working Group on Production Management Systems (WG 5.7). He holds an M.S. in Information Engineering from University of Illinois at Chicago and a B.A. from Cornell University. His e-mail address is charles.mclean@nist.gov.