Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

FEASIBILITY FOR AUTOMATIC DATA COLLECTION

Neil H. Robertson
Terrence Perera

Systems, Modelling and Integration Research Group
Sheffield Hallam University
Sheffield, S1 1WB, UK

ABSTRACT

It is argued that the data collection process is the most crucial and time consuming stage in the model building process. This is primarily due to the influence that data has in providing accurate simulation results. Data collection is an extremely time consuming process predominantly because the task is manually orientated. Hence, automating this process of data collection would be extremely advantageous. This paper presents how simulation tools could utilize the Corporate Business Systems as the potential source for simulation data. Subsequently a unique interface could be developed and implemented to provide this data directly to the simulation tool. Such an interface would prove to be an invaluable tool for users of simulation.

1 INTRODUCTION

The use of simulation has recently become even more widespread within the various manufacturing industries, it has also recently emerged in the service industry. The potential benefits of simulation are being realized in these industries, thus the demand for simulation is increasing. Since the 1970's, the use of these simulation tools has been led by those able to support and develop such technology, namely the automotive and aerospace industries. The potential benefits of such tools are both immense and diverse (Hitchcock et al 1994, Shukla et al 1996), although the primary benefits are said to be decision management and visualization. According to an influential body (IMTR 1999) 'no other technology offers more potential than Modelling and Simulation for improving products, perfecting processes, reducing design-to-manufacture cycle time, and reducing product realization costs' and then Davis (1999) further prognosticates that 'there will continue to be major advancements in simulation technologies'.

This paper will discuss the issues of data collection for simulation model building within the aerospace and automotive manufacturing industry. It is aimed at an audience with a specific interest in Discrete Event Simulation, although the points raised can equally be applied to most other types of simulation (Murphy et al 2000). This paper presents the way simulation models are built with a strong emphasis placed on the data collection issues. The results of a recent simulation survey regarding data collection and model building are then surmised. The paper proceeds to identify the Corporate Business System as a potential data source for simulation in order to accelerate the data collection process. Alternative methodologies and mechanisms to automatically interface the Corporate Business System to the simulation model are then described. The potential of these methodologies and mechanisms are discussed, together with some rare examples of these methodologies in practice. The authors conclude that the way forward for the simulation industry is to integrate the simulation tools to the Corporate Business Systems. A particular focus is then placed on the tool data requirements and the simulation data available in the Corporate Business Systems.

2 SIMULATION MODEL BUILDING

Generally, simulation models are built for one of three different applications, *System Design, System Management or Training* (Davis 1999). Modelling of new systems or changes to existing systems can both be classified as *System Design*; Discrete Event Simulation is usually used for this application. *System Management* has applications at an 'on-line' level, which are used for scheduling and real time system changes of the production network. Finally *Training* specifically refers to operation or task training, it can additionally be utilized for visual work instructions.

This paper is particularly concerned with the application of *System Design*. In this context, simulation tools are dependent upon providing decision support in order to model a new system or existing system. A strong emphasis is placed upon capturing the reality of a complex system, at the level of detail and accuracy required. As simulation models can only be as good as the data used and the assumptions made, both of which require great understanding. Hence, 'conducting a simulation study is a demanding

task, because multidisciplinary knowledge, experience and skills are needed' (Luker and Adelberger 1986).

It is strongly argued that data collection is the most crucial and time consuming stage in the model building process (Liyanage 1999, Oakshott 1997, Robinson and Bahatia 1995). This is predominately due to the influence that the data has over the level of accuracy attainable by the model.

The authors have recently conducted a survey on a broad range of simulation issues including those of data collection. This survey was conducted on 25 vetted simulation practitioners who completed a questionnaire designed by the authors. This survey was performed at the major annual gathering of the simulation community at the Winter Simulation Conference in Phoenix, Arizona in December 1999. Some of the findings are detailed in Table 1. The authors analyzed the results of the simulation survey and identified the problems and difficulties intrinsic to the data collection process. The key points identified are that data collection is predominately a manual process, unstructured, and there are many approaches to tackle common problem issues, such as data accuracy and duplication.

The survey has also highlighted that simulation is not considered to be part of the business process and with only one exception there was no interface established between the tools and the business systems. The results contained in Table 1 indicate that an automated data collection system would prove to be an invaluable tool for simulation users.

Table 1: Simulation Survey Issues and Responses

Issue	**Response**
Structured approach to data collection?	• 68% have a structured method for data collection, such as a data collection templates that are completed by the Project Team
Encountered data duplication	• 69% of simulation practitioners have encountered data duplication
Model use after project requirements are satisfied	• On average the respondents only discard 31% of models, although these models would probably have been used for System Design as opposed to System Management. • 27% of the respondent reuse the model for a different purpose, leaving 42% who keep the model updated • Some companies discard up to 70% of models, others keep them all updated
Is Simulation part of the business process?	• 62% of practitioners do not consider simulation to be embedded into their business process
How is the data supplied to the model?	• 60% of respondents indicated that they manually input the data to the model • The remainder principally have a direct link to an external system i.e. a spreadsheet • Some companies rely 100% on the manual method, whereas others rely 100% on a link to an external system
Methods used to ensure data accuracy, reliability and validity	The respondents indicated many different methods to tackle this issue, these are: • Model validation runs • Interviewing area experts • Model ('virtual') and plant ('real') comparisons • Basic 'sanity' checks • Personal experience • The 'customers' responsibility
Methods used to choose between duplicate data sources	The respondents indicated many different methods to tackle this issue, these are: • Most recent data • Most reliable data • Most local to the source/origin • Obtain team knowledge • Based on personal experience
Models develop and evolve how is data validity maintained?	• Most respondents indicated that it's the users responsibility to update the model, others stated that they used version or configuration control on the model • Although a few addition respondents indicated that they have developed external links to automatically modify the model
Methods used to source data, i.e. • Computer Based System • Paper Based System • People Based System	• Most used a variety of the three data systems to varying degrees, but this is dependant on the type of systems available • Some companies rely 95-100% on Computer based systems • Conversely some had a strong dependency on Paper based systems
Where is the majority of data held?	• The majority of respondents indicated that most of their data is held in local systems such as a spreadsheet • One case indicated that they have direct links to the corporate database for the bulk of their data

This simulation survey highlighted that the way in which data collection was approached was a key aspect. This essentially revolved around the method of supplying data to the model. Hence, the methodologies for data input were derived by the authors through observations made on various research exercises including follow-up interviews to the simulation survey and subsequent case studies (Murphy and Perera 2001).

3 DATA INPUT METHODOLOGIES

During this industrial analysis the authors identified four different methods to input the required data to the model. These are detailed below and also illustrated in Figure 1. The aim was to identify the current and emergent methods used to supply data to the model within industry. In Figure 1 methodologies '1' and '2' are currently used extensively within the industry, whereas methodologies '3' and '4' are only just emergent and as such, there is merely one isolated case evident for each methodology within industry. It is prognosticated that these two methods of integration offer great advancements with respect to data collection in the future, although it is dependent on their intended application type. Details of these cases are also provided below.

1. The model builder manually collects the required data via various mechanisms such as data templates completed by the project team, by using information spreadsheets or by interviewing individual domain experts. The data is then manually entered into the model as and when required. This is a simplistic method, especially in the larger manufacturing organisations who should be leading the way. The benefits are that it is a simple method to follow for model building, and all the data is verified by the modeler as it is entered into the coding. It has inherent drawbacks, principally due to its manual nature. These are namely the extended time, effort and errors. In addition the data is stored within the modelling tool and it is therefore a very inflexible system, as the coding will need to be changed if any of the data is modified.
2. As in the previous methodology, similar mechanisms are used to collect data, but this data is then amalgamated form the various sources onto a formatted 'master' spreadsheet. This spreadsheet contains the majority of the required data for the model. Once the model is built, the data is automatically imported from the spreadsheet to the model. Hence, the data is stored externally to the model, which enables flexibility of the data and the model. The im/exporting of the data required by the model is a significant advantage on the previous method, but it takes time and effort to amalgamate and format all the data. This methodology is becoming an increasingly popular method within industry.
3. The model utilizes an Intermediary Simulation Database that automatically retrieves and stores the required data from the sources within the Corporate Business Systems. The model 'reads' the required data from the 'integrated' database to run the model. Again the data is stored externally to the model, thereby introducing flexibility. In addition, as the data becomes available from the Corporate Business System, the intermediary database gathers and manages the model data. Thus time, effort and errors can be dramatically reduced if the intermediary database automatically refers to the data in the Corporate Business System. Except for one specific case detailed below, this method is not apparent within industry.

This methodology is currently being tackled and developed by a major US car manufacturer, who need this level of integration. Although the intermediary database shown in Figure 1 will in fact be a spreadsheet that is populated by the corporate business system. It is unclear how effective this case may prove as it is still under development.

4. In this methodology the model automatically collects data from the Corporate Business Systems via an interface as and when required in order to run the model. Essentially the model is referred to an external location to 'read' the data directly from the Corporate Data Systems. Again the data is stored externally to the model, thereby inducing flexibility. In addition, as the model is built it is referred to sources within the Corporate Business System. The automated referral system dramatically reduces time, effort and errors. There is a major drawback due to the complexity and size of this methodology, as there are so many different sources there will inevitably be an alternative source for the same data item. This data duplication may cause a problem in terms of data accuracy, reliability and validity. A further hindrance is that in some situations the data that is referred to in the model may not be available at the source when the model refers to it. This methodology is extremely complex to set up and it is apparent only in one isolated case within industry.

This methodology is being accomplished by a major US aerospace company, who has developed an interface that relates directly to the business system and the model. It is used for *System Management* as it used for routine modelling of the production line on a day to day basis. It rapidly modifies the production line when circumstances

change, such as amendments to the schedule provided by the corporate business system. It then recommends actions to be taken in order to meet the amended requirements. This set-up is for *System Management* rather than *System Design,* as the possible configurations of the system are already known, simplifying the modelling process for *System Management*.

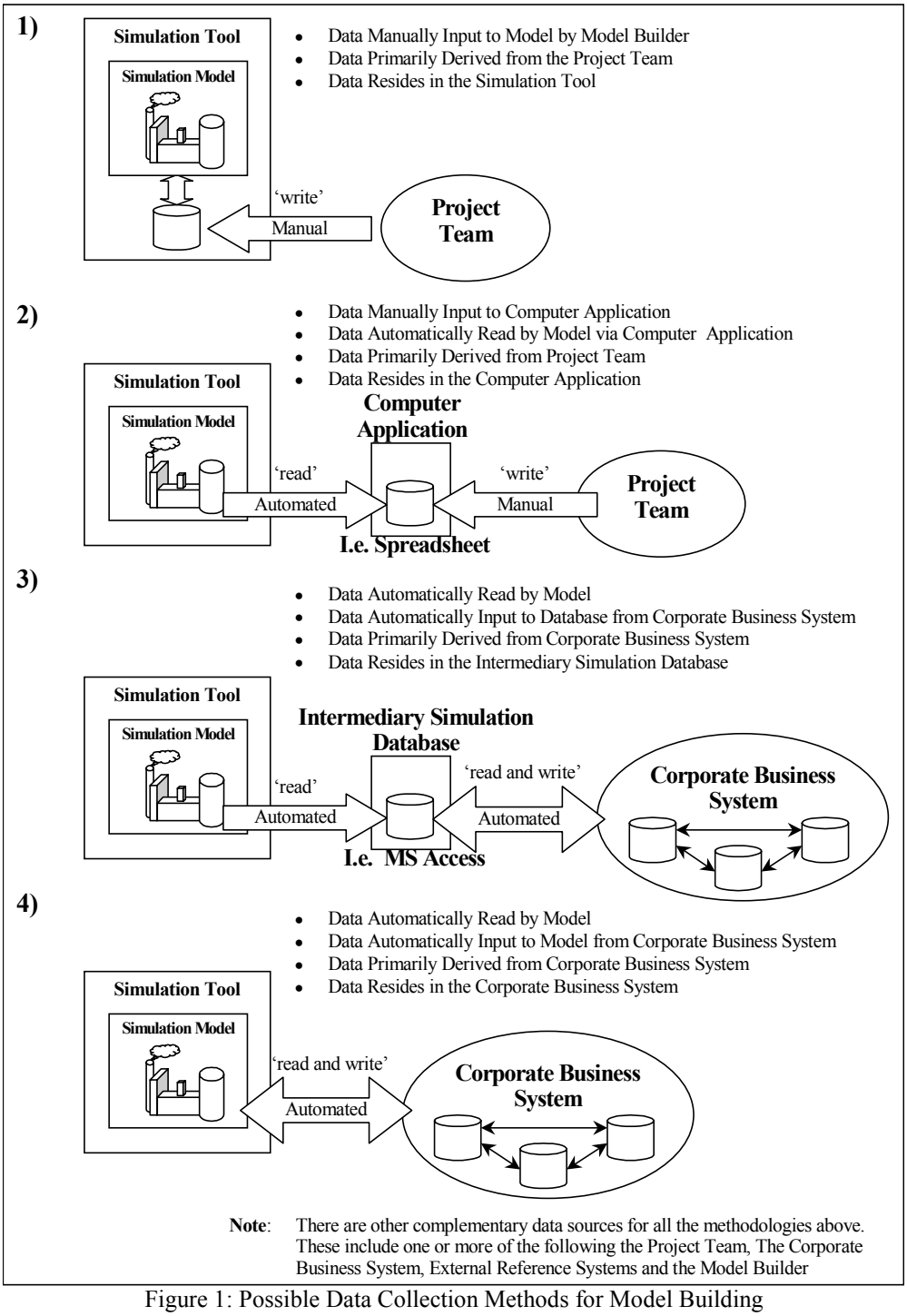

Figure 1: Possible Data Collection Methods for Model Building

4 INTEGRATION IS THE FUTURE FOR SIMULATION

It is apparent that the methodologies ('3' & '4') in Figure 1 to integrate the simulation model to the Corporate Business Systems could be the solution for the data collection process, as it would increase the data accuracy and reliability whilst also minimizing data collection efforts. This solution has been seen and acted upon by the case example provided for the two methodologies, and it is hypothesized that other organisations will follow this path. This integration is extremely difficult to establish, but the long term benefits such as easier data collection, increased data accuracy, reliability and validity, decreased data duplication, and being able to obtain the data as soon as it becomes available are all lucrative advantages. It is evident that these types of integration mentioned are particularly suited to *System Management* for applications in production planning, as the application of integration to *System Design* is more complex.

The Corporate Business Systems mentioned have over recent years evolved to be incorporated to a single extended information management system known as Enterprise Resource Planning (ERP). ERP systems are now being used to manage a huge amount of the information within an enterprise (Algeo 1998, Kishore and Seo 1999). They evolved out of Material Requirements Planning (MRP) and Manufacturing Requirements Planning (MRPII) (Ranky 1990, Chase et al 1998) and then crossed over into the fields of Finance, Sales and Human Resources. Hence, these interoperable systems can be divided into, Sales and Distribution, Human Resource, Finance and Manufacturing (Algeo 1998). ERP systems are particularly advanced in the larger industries, where such systems are very complex but highly beneficial, such as the aerospace and automotive industries.

Due to the complexity and level of detail required for model building, it is acknowledged that not all of the required information will be available through the ERP medium. Hence, other software tools such as CAD, CAM, PDM and some additional specialist tools are commonly used to support ERP systems, in order to complete the corporate business system. But still it must be acknowledged that the ERP system is a prime source for the model data as it contains information required by the tool.

5 SOLUTION

It is recognized that the manufacturing industry has recently adopted these extensive ERP systems in order to manage data across their enterprise. Hence, a logical solution to reduce data collection efforts during model building is to integrate the simulation data required by the simulation model with the data available in the ERP system. Such an automated solution, once established, would minimize the efforts for data collection and validation by the model builder, thus permitting the concentration of effort on other stages in the simulation process (Figure 1).

This solution is currently being addressed by the authors under a European research grant. This work programme involves three related aspects: the simulation tools used, the business system and the interface between the two. The research methodology used by the authors has so far resulted in the following:

- Identified the data requirements for each simulation tool through levels of decomposition,
- Identified the data stored within ERP systems,
- Proposed methodologies and mechanisms to enable integration (Figure 1).

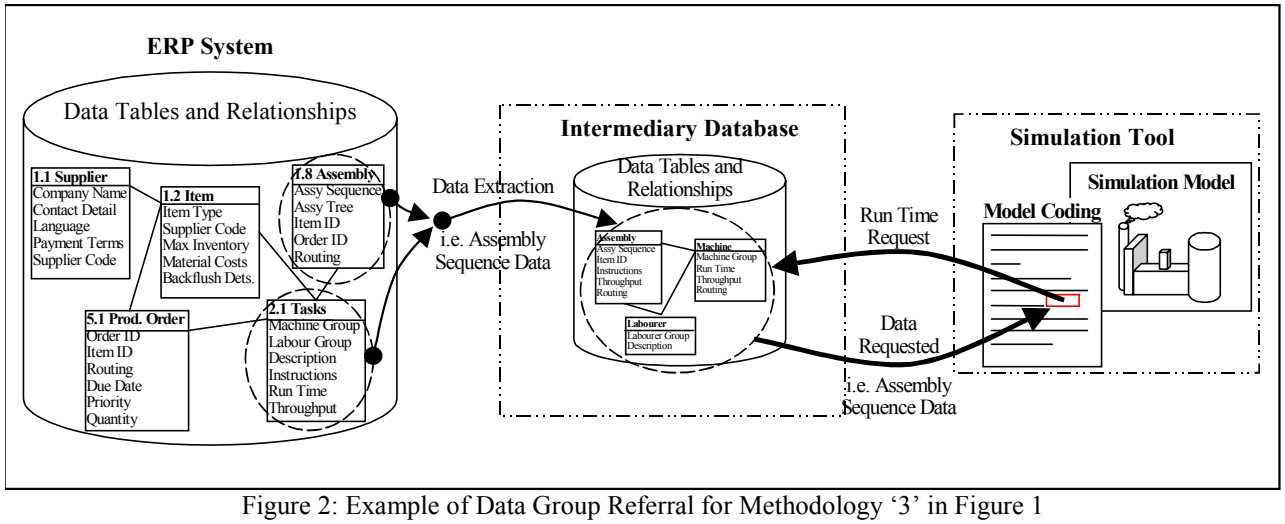

Figure 2: Example of Data Group Referral for Methodology '3' in Figure 1

This research is to be supplemented by the implementation of a physical 'test case' within industry. This will aim to provide the integration through the methodologies proposed, using a single simulation tool (DES) and a fully operational ERP system. The authors hypothesize that methodology '3' in Figure 1 will be tackled due to the inherent flexibility provided by the database structure. It is also envisaged that the database will form a point of reference for other data sources external to the ERP system, such as CAD, Product Data Management (PDM) and even the Project Team. The reason for the rejection of methodology '4' is due to the structural rigidity of the system and level of complexity imposed by the direct interface, especially when such a high level of data will need to be sourced. Hence, it is currently seen by the authors as unfeasible from a IT system perspective.

An example of methodology '3' is provided in Figure 2. It illustrates how the intermediary database would automatically extract and then store the simulation data contained within the ERP system. The intermediary database is updated through regular data extraction cycles, to 'mirror' the ERP system. The simulation model then refers to the intermediary database when it is run. Hence the coding that underpins the model requests the data from the database and subsequently receives the data. Enabling the model to run with 'real' data and provide results.

This 'test case' will be initiated by the development of a data map representing the actual data storage locations within a physical ERP system. Such an interface to an ERP system will inevitably raise some problems that will be addressed. These obstacles are envisaged to be those described below and they may explain why these two types of integration are not evident within industry.

- For *System Design* the data availability from an ERP system may be slow/intermittent, due to its iterative nature of model building
- There is no industry standard ERP structure. ERP systems have a variety of unique structures to address the specific needs of an organisation.
- Writing interfaces to compute and provide the data required by the models is a major effort (Jain 1999)
- Additional data that ERP systems do not contain may also be required, such as policies, procedures, etc.

The authors' research will aim to establish whether these obstacles can be overcome. The vendors of ERP systems have recognized the need for external applications to be able to access the data ERP systems contain, hence they have developed adapters or ports that enable the communication between the ERP system and different external or foreign applications. The adapter enables a multitude of different programming languages to 'talk' to the ERP system, thus the external application can read information through the 'portal' provided by the vendor. This level of communication has the potential to be taken a step further, as it is recognized that ERP vendors could in fact embed the functionality of simulation into its ERP application set. This may become apparent in the future if ERP vendors start collaborating with or acquiring simulation companies.

6 CONCLUSION

Two methodologies have been proposed to achieve such a solution and two leading industrial examples have been outlined. Both of these methodologies aim to reduce data collection efforts for model building by developing an automated interface between the simulation tools and the organisations corporate business systems, specifically the ERP system. The perceived benefits of such integration for data gathering are a vast reduction in terms of time and effort, unsurpassed levels of data accuracy, reliability and validity, a significant decline in data duplication and finally immediate availability of data as it is released into the business.

The concept of automatic data collection through an interface between the simulation model and Corporate Business Systems is still in the distant future. The pressure is on practitioners, researchers and vendors, to address these issues within the simulation industry. The initiative should be taken by the simulation vendors to develop interfaces with major ERP vendors; additionally this would also prove to be a 'natural way to move simulation into the business mainstream' (Banks 1999). The onus on the part of the researchers is that attempts and 'test cases' for this interface must be made in order for the potential to be realized and then disseminate it to industry. This action is currently being conducted and executed by the Modelling and Simulation Research Group at Sheffield Hallam University, UK, and a future paper will report these findings.

REFERENCES

Algeo, M.B. 1998. Enterprise Resource Planning Systems for Manufacturing. In Information Technology for Engineering and Manufacturing by National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, September 9-10th

Banks, J. 1999. What Does Industry Need from Simulation Vendors in Y2K and After? - A Panel Discussion. In Proceedings of the 1999 Winter Simulation Conference, Arizona, USA, December

Chase, R.B, Aquilano, N.J., Jacobs, F.R. 1998. Production and Operations Management: Manufacturing and Services, Irwin/McGraw-Hill, 8th ed.

Davis, W.J. 1999. Simulation: Technologies in the New Millennium. In Proceedings of the 1999 Winter Simulation Conference, Arizona, USA, December

Hitchcock, M., Baker, A., Brink, J. 1994. The Role of Hybrid Systems Theory in Virtual Manufacturing. In Proceedings of the IEEE Conference, p 345-350

IMTR. 1999. Modelling and Simulation. Integrated Manufacturing Technology Roadmapping Project (IMTR), Tennessee, USA, December

Jain, S. 1999. Simulation in the Next Millennium. In Proceedings of the 1999 Winter Simulation Conference, Arizona, USA, December

Kishore, R., Seo, Y. 1999. What is ERP?, Management Information Systems, http://wings.buffalo.edu/academic/department/som/students/clubs/KMBA/work/mgs/ywerp/ywerp.htm Accessed on 12/04/2001

Liyanage, K. 1999. Methodology for Input Data Modelling in the Simulation of Manufacturing Systems. Thesis (PhD), Sheffield Hallam University, UK

Luker, S.R.E., Adelberger, H.H. 1986. Intelligent Simulation Environments. In Intelligent Simulation Environment, Eds SCS, San Diego, p150-156.

Murphy, C., Murphy, S., Perera, T., Robertson, N. 2000. Simulation in Aerospace Industry: Applications, Software and Integration. In Conference on Achieving Excellence in Aircraft and Aviation, Atlanta, Georgia, October 16-17th

Murphy, S., Perera, T. 2001. Sucesses and Failures in Global Development of Simulation. (To be published), Sheffield Hallam University, UK

Oakshott, L.A. 1997. Business Modelling and Simulation. Pitman Publishers, London, UK

Ranky, P.G. 1990. Manufacturing Database Management and Knowledge Based Expert Systems, World Class CIM Series, Biddles Ltd, Surrey, UK, 1st ed.

Robinson, S., Bhatia, V. 1995. Secrets of Successful Simulation Projects. In Proceedings of the 1995 Winter Simulation Conference, Arlington, USA, p 61-67

Shukla C., Vazquez M., Chen F. 1996. Virtual Manufacturing: An Overview. In International Conference on Computers and Industrial Engineering, Vol. 31, No.1/2, p 79-82

AUTHOR BIOGRAPHIES

NEIL ROBERTSON is a Research Assistant within the Systems Modelling and Integration Research Group at Sheffield Hallam University. He is currently actively involved in a large EPSRC funded project investigating the use of simulation tools within industry. His specific interest is in the exchange of data between simulation tools and the business systems. He is also presently engaged in obtaining an MPhil award for this research. His email address is <n.robertson@shu.ac.uk>

TERRENCE PERERA is Head of Enterprise and Systems Engineering within the School of Engineering at Sheffield Hallam University. He also leads the Systems Modelling and Integration Research Group. His current research interests include the implementation, integration and practice of virtual modelling tools within all industrial sectors. His email address is <t.d.perera@shu.ac.uk>