A Methodology for Simulation Development on the Basis of Cause-and-Effect Modeling in E-Commerce

Axel Hummel, Heiko Kern
Business Information Systems
University of Leipzig
Johannisgasse 26, 04103 Leipzig, Germany
{hummel, kern}@informatik.uni-leipzig.de

René Keßler, Arndt Döhler
Intershop Communications AG
Intershop Tower, 07740 Jena, Germany
{r.kessler, a.doehler}@intershop.de

Abstract—The optimal configuration of an e-commerce system is a complex problem. The objective of the research project SimProgno is therefore the development of several e-commerce simulations and the integration of these simulations in order to support the shop managers in their decision making process. The structured development of these simulations is a challenging task for a development team. In this paper, we address this problem and present the SimProgno methodology which is used for the simulation development. This methodological framework is based on specific requirements of the research project and the e-commerce domain. Furthermore, we evaluate the methodology on the basis of its usage in the SimProgno project.

Keywords—methodology; simulation; e-commerce; cause-and-effect modeling; development process; System Dynamics; agent-based simulation

I. INTRODUCTION

The electronic commerce (e-commerce) is an important channel of distribution for many companies. An e-commerce channel is typically realized by using an online shop. The operation including installation, configuration, maintenance, etc. of an online shop is a complex problem which requires precise knowledge about the interdependencies between configuration parameters and the impact of these parameters within the e-commerce ecosystem. These interdependencies are often of multidimensional nature and include dimensions such as technical, economic, or social aspects. At the moment, shop managers decide on basis of their expert knowledge, current trends, and business objectives about an optimal shop configuration. This decision making process is insufficient because their decisions have a subjective character, are non-transparent, and the effects of the decisions are difficult to predict. An approach to support and improve the decision making process is the use of simulations covering the expert knowledge.

The simulation of a shop configuration can be separated in sub-simulations covering a certain e-commerce service such as payment, fulfillment, or marketing. Hence the aggregated simulation of a shop requires the development and integration of several sub-simulations. The structured development of these sub-simulations is a challenging task for a development team. A solution for this problem is the definition of a methodology including, for instance, a defined process, roles, responsibilities, or interfaces between all participants or expected deliverables. The advantages are the avoidance of failure during the development, and the raise of quality and efficiency of the simulation development process.

In this paper we present a methodology for simulation development. This methodology considers certain requirements such as the involvement of domain experts, the usage of different simulation techniques, and the usage of established methods and tools. Our proposed methodology is a result of the research project SimProgno [1] which aims to explore techniques to combine and integrate simulations in the e-commerce domain.

The paper is structured as follows. In the subsequent section we describe the requirements for our methodology. In Section 3 we discuss already existing methodologies relating to the SimProgno approach. The main part of the paper is presented in Section 4 and describes our process model. In Section 5 we evaluate our methodology with respect to practical use. Finally, we conclude in Section 6 with a discussion of our approach and describe future work.

II. REQUIREMENTS

The methodology has to meet several requirements to satisfy the development support of simulations. Based on the specific characteristics of the e-commerce domain and the objectives of SimProgno, we derived the following requirements.

(R1) Involvement of domain experts. Domain experts play a crucial role for the specification of the simulation
model. They have in-depth knowledge of the considered application domain but normally domain experts are unfamiliar with the development of simulation models. Therefore, a concept is required to involve domain experts.

(R2) Abstraction of certain simulation techniques. The model about the simulated system has to be described in a simulation technique-independent way. This allows a better understanding and description of the domain knowledge and enables the reuse in a changed context. This can be, for instance, changed simulation objectives or other available data.

(R3) Usage of established methods and tools. Already existing and well-functioning methods and tools should be used in the process steps. This enhances the quality of the process and the developed artifacts, eases the usage, and increase the acceptance.

(R4) Integration of simulations. The integration of simulations has to be possible in order to create an aggregated simulation. The integration process is outside the scope of the proposed methodology, but well-defined interfaces for the input and output parameters have to be specified and implemented during the development process.

III. RELATED METHODOLOGIES

Nowadays, there are many methodologies available, which consider the development process of simulation models. All these methodological frameworks focus on different aspects of the development process or refer to miscellaneous simulation techniques. Because of the applied simulation techniques in the SimProgno project, there are three different groups of methodologies that contribute to the proposed procedure of this paper.

First, there are general methods which are independent of a special simulation technique and can be used therefore for different approaches. Banks et al. [2] describe such a universal methodology consisting of twelve process steps. These steps are clustered in four phases: (1) discovery or orientation, (2) model building and data collection, (3) running of the model, and (4) implementation in the real world. The model conceptualization has a very general character. The authors give universal guidelines that can only be applied in a limited way for System Dynamics and agent-based simulations used in the SimProgno project.

The second group contains the methodological frameworks that consider System Dynamics modeling. Examples of such methods can be found in [3], [4], [5]. We discuss the work of Forrester [3] and Cavana & Maani [5]. Forrester proposes a methodology that includes the following six phases: describe the system, convert description to level and rate equations, simulate the model, design alternative policies and structures, educate and debate, and finally implement changes in policies and structure. The methodology described by Cavana and Maani consists of the five major phases problem structuring, causal loop modeling, dynamic modeling, scenario planning and modeling, as well as implementation and organizational learning. In conformity with the method by Banks et al., both methodologies cover the whole development process of simulation models. Although both methods consider the development of System Dynamics models, there is a substantial difference between the two approaches. Cavana and Maani create a causal loop diagram in the second step which is then refined to a complete System Dynamics model. Since causal loop diagrams describe cause-and-effect relationships in a way that is independent of a specific simulation technique, the authors abstract from the System Dynamics approach. Forrester develops the simulation model without such an abstraction and rejects the use of causal loop diagrams.

Finally, agent-oriented methodologies are another relevant group for the methodological framework of SimProgno. Today, there is no widely accepted approach for the development of agent-based simulation models available. Therefore, a huge number of methods can be found in the literature. Elamy and Far found over thirty different methods [6]. After the genealogy by Henderson-Sellers and Giorgini [7], most agent-oriented methodologies like Prometheus [8], AOR [9], MaSE [10] and Gaia [11] are refinements of object-oriented methods. Furthermore, there are methodological frameworks that have their origin in requirements engineering or in the field of knowledge engineering and artificial intelligence. An example of the former group is Tropos [12]. A well-known representative of the latter sort is the MAS-CommonKADS approach [13]. Comparisons and evaluations of the several methods show that there are large differences regarding the steps of the development procedure or the perspective of the approaches [6], [14], [15]. Because of their origin, these methodologies are not designed for the simulation context. Therefore, essential process steps like data collection and data mining or model validation are outside the scope of these methods.

In conclusion, the investigated methodologies are not entirely suitable for the e-commerce domain. Table I shows a comparison of the analyzed frameworks regarding the requirements of Section 2.

None of the methodologies provide guidelines for the involvement of domain experts. Nevertheless, all methodological frameworks include process steps such as problem formulation or requirements analysis which need specific domain knowledge. Therefore, an implicit involvement of domain experts can be assumed. Hence, we conclude that the methodologies fulfill the requirement R1 partially. The approach by Banks et al. [2] is the only method that is completely independent of a specific simulation technique, but
some of the process steps are very general and have to be substantiated. The agent-based methods and the methodology by Forrester focus on a certain simulation approach and reflect therefore only a special perspective of the development process. The method by Cavana & Maani lies in somewhere between the two groups. This method is primarily developed for System Dynamics, but the used causal loop diagrams represent knowledge of a special domain in a way that is independent of a certain simulation technique. As a consequence, only the methodology by Banks et al. fulfills the requirement R2. The method by Cavana & Maani partially satisfies R2, whereas the other two approaches are not independent of a specific simulation technique. Almost all methodologies used established methods and tools in the individual steps, such as stock-and-flow diagrams for System Dynamics modeling or UML diagrams for agent-based modeling. Therefore, the agent-based methods and also the approaches by Cavana & Maani as well as Forrester partially fulfill the requirement R3. Since the process steps of the method by Banks et al. are very general, no statements can be made about this requirement. Finally, none of the investigated methodologies consider integration issues during the several development steps. The definition and implementation of interfaces for input and output parameters is therefore not included in the analyzed methods.

IV. THE SIMPROGNO METHODOLOGY

The analysis in the previous section shows, that there is a lack of methodological frameworks satisfying our requirements. Therefore, we have developed an own methodology in SimProgno based on the approaches of Section 3. Before describing the certain process steps of the proposed methodological framework, we firstly introduce the involved stakeholders and their roles.

A. Roles and Stakeholders

The stakeholders can be abstracted to the three different roles domain expert, simulation modeler, and simulation developer. For the description of these roles and their interactions, we refer to the problem space and solution space characterizations by Czarnecki and Eisenecker [16]. Fig. 1 shows the roles and their assignment to the problem and solution space.

First, there is the domain expert who represents the so-called problem space. The domain expert has an in-depth knowledge of the considered application domain, especially about the fundamental cause-and-effect relationships. Furthermore, the domain expert is able to identify relevant scenarios, because of his practical knowledge about the customer requirements. Typical stakeholders that appear as domain experts are employees of specialist departments of the service provider like key account managers, sales, or business consultants. In SimProgno, there are further domain experts such as technology partners of the service provider.

The opposite part of a domain expert is the simulation developer, who has expert knowledge about the solution space, that is, he is familiar with simulation techniques, tools, and related methods. The simulation developer is responsible for the development of the simulation. This role can be filled by software developers, programmers, and research partners specialized in simulation.

The third role in our methodology is the simulation modeler who mediates between the domain experts and the simulation developer. He has general knowledge of the problem and solution space and the complete development process. This role is realized by the service provider in the SimProgno project.

B. Process Steps

In SimProgno, we developed a methodological framework which is structured in ten phases. Fig. 2 gives an overview of the development process which will be described in detail in the following subsections.

1) Scenario selection: The objective of this step is the selection of a relevant scenario. Therefore several methods of market analysis can be used in order to identify typical e-commerce problems which require additional decision support. We use three different approaches for finding potential problems. Firstly, the simulation modeler carried out an empirical online
Fig. 2. Overview of the process steps of the SimProgno methodology and the involved roles. The roles are shown at the right side (DE: domain expert, SM: simulation modeler, SD: simulation developer).

survey [17]. Secondly, we analyzed already existing studies published by leading market research companies such as Forrester, Gartner, and Nielsen. Thirdly, domain experts and customers of the service provider are interviewed by the simulation modeler in order to validate the results of the previous steps.

The outcome is a collection of relevant problems from which the simulation modeler choose one scenario. Mostly, the market analysis shows that there are different scenarios possible. The results of this step can therefore be used for additional simulation studies.

2) Problem structuring and simulation requirements definition: The result of this step is a precise formulation of the problem and the definition of corresponding requirements. The initial specification of the simulation’s objectives is done by the simulation modeler. Based on a workshop with domain experts, the specification will be refined. In order to obtain unambiguously objectives that are specific, measurable, assignable, realistic and time related, we apply the S.M.A.R.T. method [18]. Typical questions that should be discussed in the expert workshops are: Which are the most important problems of the experts in the selected scenario? Is simulation a suitable tool to address these problems? Which cause-and-effect relationships should be represented in the model and which are outside the scope?

3) Data collection and data mining: Reliable data play a crucial role in the development of simulation models. The phase data collection and data mining is therefore an essential part of almost all methodologies for simulation development. Since the collection and evaluation of data is very time consuming, this phase should be started as soon as possible. In the SimProgno methodology this phase begins during the second phase in an expert workshop. Based on the precise problem formulation related domain expert projects and available data sources of the domain experts are identified. The data collection and data mining phase runs parallel to the remaining steps of the development process.

On the one hand the simulation modeler uses different data mining methods such as cluster analysis, association analysis and regression analysis in order to identify relevant system parameters and cause-and-effect relationships. On the other hand, the data are necessary for the specification of the mathematical model, the calibration, and validation of the simulation. The results of the data collection and data mining phase are therefore important inputs of many other phases. These dependencies are indicated in Figure 2 by arrows, which point to the corresponding process phases.

For the development of the simulation model and its validation, we primarily use real transaction data that are historical data of customer behavior in online shops. Furthermore data based on surveys and studies are used, if the transaction data are insufficient.

4) Identification of system variables: The objective of this phase is the identification of relevant system variables. This takes place in a second expert workshop, which is structured in three parts. Firstly, it is necessary to collect the explicit and tacit knowledge of the involved domain experts. We use several creativity techniques [19] such as brainwriting and brainstorming for the elicitation of the expert knowledge. This step results in plenty of index cards, each of them labeled with a system variable. In the next step, the domain experts discuss the system variables in order to define a common semantic of these variables. It is essential to include the unit of measure in the term definition, since there are different units in most cases, especially if the variable can be measured over a certain period of time. During the discussion, the simulation modeler visualizes the index cards in form of a word cloud on a pin board. System variables which are referred by multiple index cards are considered to be particularly important for the simulation. Clustering of the word cloud is the final step of the expert workshop. The clusters are derived from the problem formulation and the corresponding requirements of the previous process phase.

Additionally, the system variables are classified
in three groups: (1) input parameters, (2) output parameters, and (3) auxiliary parameters. In order to be able to integrate different simulation models to a complex simulation, we define a fixed set of parameters. These global parameters represent typical key performance indicators of an online shop and are influenced by several simulation models. Therefore, the input and output parameters are divided into local and global input and output variables. Local variables are only used for one specific simulation. Auxiliary parameters represent internal variables of the simulation model which cannot be configured by the end user. This classification provides the basis for the specification of the interfaces.

5) Modeling of cause-and-effect relationships:

The result of this process phase is a so called causal loop diagram of the underlying cause-and-effect relationships characterizing the selected scenario. Causal loop diagrams are directed graphs which describe dependencies between factors of a system in a qualitative way [20].

This phase is realized in the third expert workshop. First, the determined system variables are related together, in order to identify dependencies between them. The relationships are symbolized by lines between the respective system variables. System variables that have no connections to the other ones can be moved in a special archive indicating that these parameters are analyzed later. Obviously, these system variables do not influence the other parameters at the present time. Next, the domain experts have to identify the cause and the effect of the several relationships, which are denoted by arrows pointing from the cause to the effect. Finally, the cause-and-effect relationships are specified in more detail. Especially the polarity (positive or negative) and the temporal effect (short-term or long-term) of the cause-and-effect relationships are relevant attributes.

Causal loop diagrams describe the tacit knowledge of the domain experts in a clear and explicit way. Therefore, the causal loop diagram is an important milestone in the development process. Since causal loop diagrams represent the cause-and-effect relationships in an abstract way, this formalism is the basis for all simulation models, whichever simulation technique is used.

Finally, the initial causal loop diagram of the expert workshop is refined by the simulation modeler on the basis of data analysis and literature research, since the domain experts may have a limited point of view. During the specification of the cause-and-effect relationships, it is possible that additional system variables are identified or that the problem formulation has to be revised. This requires the repetition of the process phases two respectively four.

6) Conceptual model development:

The result of the sixth phase is a complete conceptual simulation model which refines the identified cause-and-effect relationships. First, the simulation developer decides which simulation technique is the most suitable approach to extend the model of the cause-and-effect relationships corresponding to the specified requirements. At this time, we apply two different simulation techniques: System Dynamics and agent-based simulation. The decision, which approach is the most suitable one, is based on several criteria such as the homogeneity and number of the actors which have to be simulated, the structure of the real system, the type of the available data (individual data versus aggregated data of customer groups or countries), and the performance of the IT infrastructure [21], [22].

For the development of a specific simulation model, the methodologies of Section 3 can be used. In SimProgn, the simulation developer creates a stock-and-flow diagram in the case of System Dynamics. Developing stock-and-flow diagrams from causal loop diagrams is widely used in System Dynamics methodologies [20], [23]. For agent-based simulation, the simulation developer uses various types of UML diagrams to specify the static structure of the simulation model and its dynamic behavior.

The definition of the input and output interfaces completes the conceptual simulation model and is the last step of this phase. The interfaces specify the local input and output variables and additional which global parameters are used. Local parameters are specified by its name, range, unit, and an optional default value.

7) Validation of model conceptualization:

The validation of the model conceptualization is an important quality assurance measure. The domain experts check the simulation model regarding plausibility and the specified requirements. This process phase is realized by another workshop. Graphical model representations are used to communicate the simulation model. In case of System Dynamics, we use stock-and-flow diagrams. In case of agent-based simulations, we apply UML class diagrams, UML activity diagrams and UML sequence diagrams to describe the structure and behavior of the model. If inconsistencies are identified, a repetition of phase six is necessary in order to improve the simulation model.

8) Defining the mathematical model:

The result of this phase is a complete quantitative simulation model allowing precise statements about the behavior of the system over time. The proceeding we choose depends on the specific simulation technique. In System Dynamics, the dependencies between the system variables are specified by mathematical formulae. Moreover all stocks and auxiliary parameters are initialized. In case of agent-based simulations, the behavior of the agents is quantified by decision values. Since domain experts have mostly qualitative knowledge about the interrelations of the system variables, the mathematical
equations are primary based on the results of phase three (data collection and data mining).

9) Implementation: After specified a complete quantitative simulation model, the simulation developer implements this model. Today, there are specialized tools available for the implementation of simulation models. In SimProgno, we use the Sphinx SD Tools [24] for implementing System Dynamics models and Repast Simphony [25], [26] for agent-based models.

10) Verification and validation: The SimProgno methodology finishes with the verification and validation of the implemented simulation model. The first step is the verification phase in which the implementation of the simulation model is checked. The Sphinx SD Tools provide syntax checking so that syntactically incorrect System Dynamics models can be detected during the implementation phase. Furthermore, we check whether the implemented model is conform to the specified mathematical model of phase 8. The agent-based models are implemented using the Java API of Repast Simphony. In this case, we use widely known procedures from the software development domain to verify the implemented models. Banks et al. [2] suggest among others the following methods and tools: code review by external persons, logic flow diagrams, standard debuggers, and output of the input parameters to check if they are unchanged. If errors are found by the simulation developer, the implementation has to be corrected. This requires a repetition of the implementation phase.

Second step is the validation of the simulation. Here, the simulation model’s correctness will be checked, which means, the simulation model replicates the behavior of the real system. The validation is realized by means of simulation experiments. Firstly, we specify the values of all input parameters and the duration of the experiment. The values of the input parameters are mainly derived from the data base which is used for defining the mathematical model. Based on the specification of the input parameters and the data set, the expected outputs are defined. Furthermore, the domain experts specify a lower bound and an upper bound for every output parameter which determine the interval in that a value has to lie in order to be accepted as correct. If a simulation model uses random numbers, we have to specify how often a simulation experiment has to be repeated in order to reduce the influence of the random numbers. Additionally, we perform a sensitivity analysis to determine the impact of a single input parameter to the output of the simulation model. For this purpose, the simulation experiment is repeated several times each with a slight modification of the specific parameter. The effects of the changes are checked for plausibility by the domain experts. The identification of errors in the simulation model requires a new iteration of the development process starting with phase two, phase six, or phase eight.

V. Evaluation

In this section we evaluate the SimProgno methodology regarding the requirements of Section 2. We used the methodology to develop simulations for different scenarios such as payment behavior [27], online marketing, product search, and product recommendation. Generally, the proposed methodology works fine and fulfills the requirements.

The involvement of domain experts was an important requirement (R1) in our methodology. A challenge was the different level of abstraction between the domain experts and the simulation modeler. The knowledge of the domain experts needs to be filtered, structured and abstracted in order to develop universal and valid statements of the domain. A further challenge was the time restriction. Often the planned number of workshops and the duration of the workshops were too much. Hence, we must performed different steps of our methodology in one workshop. A solution for the abstraction and time problem is a good preparation or the familiarization with the simulation topic by the simulation modeler.

The support of different simulation techniques was a further important requirement (R2). We developed simulations with the System Dynamics approach as well as the agent-based approach. Both simulation techniques were applied successfully but the development of an agent-based simulation is more complex than the development of a System Dynamics simulation. The reason is the specification of the simulation with the help of causal loop diagrams. In the case of agent-based simulations these causal loop diagrams have to be transformed to the individual behavior of the agents.

The decision for a certain simulation technique occurs in step five. In some scenarios we made the decision before step five. Through the usage of simulation-specific concepts the development of a certain simulation was more efficient. But the reuse in other simulations was restricted which results sometimes in a completely new development.

Regarding the third requirement (R3), we used established methods during the development process. The actors are familiar with these methods. Thus, the methodology was accepted by the actors and the development of simulation models was successful.

Following the principles of the FAMAS Simulation Backbone Architecture [28], we apply also a lightweight approach to integrate the several simulation models. All sub-simulations of an aggregated simulation are coupled together by its data flows. We define a fixed set of global parameters that is used by the several sub-simulations for information exchange. The precise specification of the input and output parameters is
therefore a crucial step in order to enable the integration of simulations (R4).

VI. CONCLUSION

A proven methodology is an important basis in order to develop valid simulation models. Specific requirements that we derived from the e-commerce domain required the development of an own project-specific methodology. This methodology is based on already existing methodologies and considers some specific requirements. The core of our methodology is the process model consisting of ten steps. We apply successfully the methodology in real world scenarios and develop several simulations.

As future work we want to use our methodology to develop further simulations. Moreover, we want to integrate additional simulation techniques. The involvement of domain experts is also an important aspect in other domains, hence, our methodology seems to be suited for these domains too. The application to these domains can be proven in the future.

ACKNOWLEDGMENT

The work presented in this paper is undertaken as a part of the SimProgno project (01IS10042C). SimProgno is funded by the German Federal Ministry of Education and Research (BMBF).

REFERENCES