Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

A NEW DATA FARMING PROCEDURE MODEL FOR A FARMING FOR MINING METHOD IN LOGISTICS NETWORKS

Joachim Hunker
Anne Antonia Scheidler
Markus Rabe

Department IT in Production and Logistics
TU Dortmund University
Leonhard-Euler-Straße 5
Dortmund, 44227, GERMANY

Hendrik van der Valk

Chair for Industrial Information Management
TU Dortmund University
Joseph-von-Fraunhofer-Straße 2-4
Dortmund, 44227, GERMANY

ABSTRACT

A key factor in maintaining a logistics network in a competitive state is gaining and visualizing knowledge. The process of gaining knowledge from a given data basis is called knowledge discovery in databases. Besides gathering observational data, simulation can be used to generate a suitable data basis for the knowledge discovery known as data farming, which is typically implemented as a study. Conducting such a study requires a suitable procedure model, describing and structuring the tasks of the process. However, existing procedure models focus on defense applications, while considerably less work was put into transferring the approaches to logistics networks. Therefore, the authors developed a procedure model for conducting a data farming study in logistics networks. In this work, we systematically introduce the necessary background and discuss existing approaches in the literature. Furthermore, we present a software framework that is used to support the process in a practical application context.

1 INTRODUCTION

A logistics network (LN) is a complex network containing a multitude of different processes and dependencies. To maintain such a LN in a good and competitive state, decision makers in supply chain management (SCM) are confronted with a high number of different logistics tasks. This results in multiple decision-making situations, which leads to the fact that an SCM must be supported accordingly (Teniwut and Hasyim 2020). One of the key factors to aid decision makers in finding answers to these tasks is gaining and visualizing knowledge. However, this can no longer be realized manually because of the complexity of the LN. The process of gaining and visualizing knowledge is known under the term knowledge discovery in databases (KDD), which is viewed as a procedure model ranging from the phase of collecting data to the phase of visualization of the results (Fayyad et al. 1996). A procedure model can be described as a standardized, organizational, and formal way to describe an ideal-typical flow of a process. In short, a common understanding is to divide the process into a sequential progression of phases. The phases contain steps, tools and personnel required to produce a result. To perform value-adding KDD requires a given data basis, which typically consists of observational data. However, observational data often lack data quality because of missing or out-of-range values (García et al. 2015). An alternative to the time-consuming and costly collecting and preprocessing of observational data is using simulation to generate synthetical data for the KDD-process, which is known by the term data farming (DF). DF is the process of using a simulation model combined with targeted experiment planning for the generation of large-scale synthetical data (Horne and Meyer 2005). In this context, the authors developed a farming for mining approach where we combine both DF and KDD in one method (Hunker et al. 2021). DF is typically understood as a

978-1-6654-7661-4/22/$31.00 ©2022 IEEE

procedure model, structuring the different phases ranging from model development to the analysis of the results. The field of application of these models is typically in a military context (Horne and Meyer 2016), although first approaches of a transfer to other domains such as production are presented (Feldkamp et al. 2015). Suitable procedure models are a necessary component to ensure the successful implementation of a DF study within the context of a decision situation.

Due to the focus on research and application in a military context, most of the existing procedure models lack relevance in the context of LNs. Specific knowledge is required for decision support in logistics networks. In order to discover knowledge in a targeted and value-adding manner, input data is required, which in our case is provided by means of data farming. Existing procedural models of data farming have essentially been developed in a military application context. However, these models cannot simply be transferred one-to-one to another application domain. For example, both the tasks and the data required differ greatly between the different domains. Therefore, this paper deals with the adaptation of existing models and presents an adapted process model in the context of logistics networks. In particular, the procedure model takes into account that DF is used in the mentioned context to generate a data basis that is suitable for the application of KDD.

The paper is structured as follows: Section 2 introduces the relevant background on KDD and DF as well as a structured literature review on related work regarding DF procedure models. On this basis, Section 3 presents our procedure model for the described context. Section 4 uses the developed procedure model to support a software implementation. The paper closes in Section 5 with a conclusion and an outlook on further research.

2 THEORETICAL BACKGROUND

The following sections will introduce the background of this paper. First, we introduce KDD in the context of our problem domain, LNs. Second, we discuss simulation and DF to generate a suitable synthecial data basis for the KDD process. The chapter closes with a structured literature review regarding procedure models for DF.

2.1 Knowledge Discovery in Logistics Networks

A large amount of data accumulates in large LNs. However, the potential knowledge hidden in the data must first be made accessible through suitable data analysis procedures, in order to contribute to the value creation of the companies (Freitag et al. 2015). Many of the procedures are listed in the literature under the structured procedure models of KDD. Among the best-known models are the Knowledge Discovery in Databases by Fayyad et al. (1996) and the CRISP data mining model (Cleve and Lämmel 2014). The models have in common that extensive data preprocessing is usually necessary to successfully apply the essential step, data mining. Unprocessed data are often unordered, erroneous, incomplete, partially redundant, or negligible in the context of the knowledge task (Runkler 2010). In addition, most data mining techniques have specific requirements for the data to be processed, which makes data processing method-dependent. Many data mining methods require a specific data type, which results in a large number of transformation options due to the large number of different data types (Aggarwal 2015). In an LN, different data types can be found, which can generally be assigned to three main categories (Cleve and Lämmel 2014). The categories include the nominal data type, with which the data can be classified into categories by means of designation (e.g., warehouse type, transhipment point location). This designation has no mathematical meaning and does not allow complex arithmetic operations within or between individual categories (Bramer 2016). Ordinal data can be sorted, i.e., they can be arranged in-order. An example from the LN is a sorting according to warehouse type. A major professional challenge is that there are often several sorting options for the elements of the LN. ten Hompel and Schmidt (2008), for example, distinguish between more than 20 so-called storage parameters, several of which allow for sorting (e.g., number of storage locations, number of levels and aisles, optimal height ratios). The third main category comprises metric data. These are numerical, discrete, or continuous data to which arithmetic operators can be applied. Data from these main categories can and often must be transformed during the application of the KDD procedure models.

The individualisation of data processing also becomes clear at the professional level, so that the relevance of the selected data depends on the application. Therefore, the choice of data must be adapted to the respective selected analysis method (Aggarwal 2015). Especially in the case of complex logistical issues such as the detection of bottlenecks in supply chains or the clarification of delays within transport routes, the selection of relevant data is not possible in advance. In addition, the risk of neglecting interesting data is disproportionately greater with very small samples.

From these preliminary considerations it becomes clear that pre-processed, relevant data are of great advantage for the KDD models. Since in practice such conditions cannot be generally assumed, the consideration turns to synthetical data. These can be generated by means of DF, a simulation-based technique.

2.2 Simulation and Data Farming

Simulation is a well established process for analyzing complex systems such as LNs and is defined as the “representation of a system with its dynamic processes in an experimentable model to reach findings, which are transferable to reality; in particular, the processes are developed over time” (Verein Deutscher Ingenieure 2014, p. 3). Simulation is an integral part of logistics assistance systems that are used to support decisions in SCM (Dross and Rabe 2014). Typically, simulation is applied as part of a project in the form of a simulation study. To conduct such a study in purposeful and targeted manner, procedure models are used. To address the mentioned flaws of observational data in the KDD process, simulation can be used to generate a suitable data basis. The process of generating synthetical data by means of simulation is called DF.

The term DF was coined by Brandstein and Horne (1998) and is used as a metaphor for the targeted cultivation of synthetical data by using a simulation model (Sanchez 2018). One of the key aspects to maximize data output based on the used model is design of experiments (DOE). Since simulation models of LNs are rather complex and contain a vast amount of factors, an appropriate design to systematically vary factors of interest is necessary. The commonly used ”trial-and-error” or ”one-factor-at-a-time” approaches are not suitable for designing experiments in an effective and farmable way. Due the curse of dimensionality, designs such as 2^k or m^k factorial are only suitable for a small number of factors. Space-filling-designs, for example, latin hypercubes, have proven to be a suitable design to use in a DOE (Sanchez 2006). With the use of a high performance computational environment, the experiments are run to generate synthetical data which can be analyzed, for example, by means of knowledge discovery. Approaches to analyze the data basis in combination with KDD are presented, for example, by Feldkamp et al. (2015) for manufacturing simulation and Hunker et al. (2021) in the context of supply chains. Similar to simulation, DF is applied in form of a study and for a structured, purposeful execution of such a study, procedure models are used.

2.3 Procedure Models for Simulation and Data Farming

Following Brinkkemper (1996), a method is a set of directions and rules to perform a project in a systematic way, containing activities that produce products. Activities in turn can be assigned to individual elements. Typically, procedure models are an element of methods or frameworks. They are used to structure and guide the process, e.g., of a simulation study. When analyzing procedure models for the research domain of simulation and related areas, one can identify different classes of procedure models. These can be distinguished on the basis of classifiers. For example, the purpose of use and the descriptive means are typical classifiers with regard to procedure models. A common class of such models are procedure models for conducting a simulation study. A wide spectrum of this class of procedure models can be identified in the established literature. The differences concern, for example, the complexity and the application domain of the models. Table 1 shows an overview of widely used, well established simulation procedure models.

It should be noted that some procedure models are established for a much longer time than the date of the exemplary source given.

Table 1: Overview of simulation procedure models.

Source	# Phases	Relevant phases
Balci (1998)	10	Problem formulation, model formulation and programming, experimentation, verification and validation (V&V).
Shannon (1998)	12	Problem definition, input data preparation, model formulation and translation, experimental design, V&V, documentation.
Chew and Sullivan (2000)	6	Requirements, conceptual mode, design, implementation, integration and test, use.
Rabe et al. (2008)	7	Task definition, model formalization and implementation, experiments and analysis, data collection and preparation.
Banks (2010)	12	Problem formulation, model conceptualization and translation, experimental design, runs and analysis, documentation and reporting.
Law (2015)	10	Problem formulation, study planning, data collection, model definition and programming, experiment design, production runs and output analysis.
Stoldt and Putz (2017)	12	Task definition, system analysis, model formalization and implementation, data collection and preparation, experiments and analysis.

Although different, many identified simulation procedure models contain relevant phases for conducting a simulation study, for example, initialization, modeling, and V&V (Barton 2020). With regards to V&V, it can be stated that there are other, specific procedure models for V&V that describe another class of procedure models. For example, the simulation procedure model by Rabe et al. (2008) describes the explicit integration of V&V across all phases in the procedure model and, furthermore, introduces a specific triangle-shaped model to guide the process of V&V. Since the procedure model for simulation by Rabe et al. (2008) is well established and part of a guideline (Verein Deutscher Ingenieure 2014), it is presented in Figure 1.

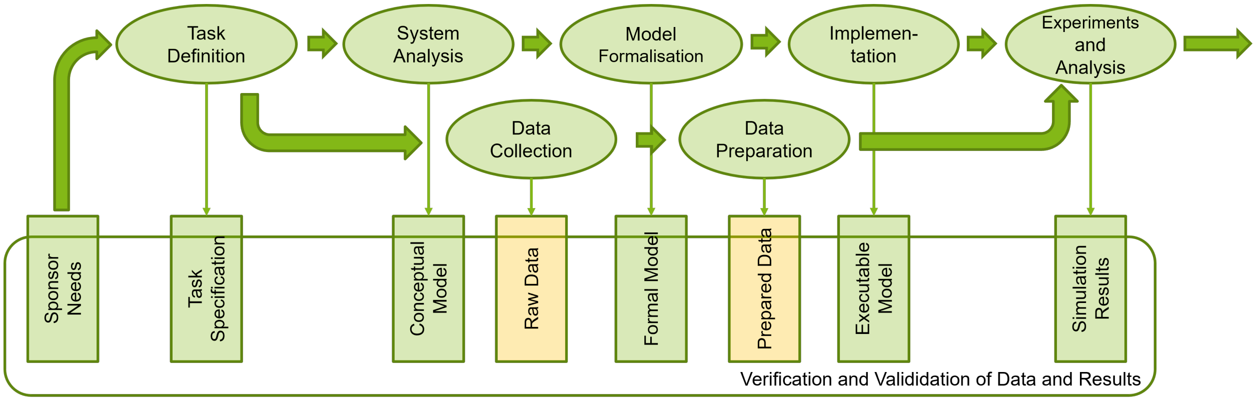

Figure 1: Procedure Model for Simulation according to Rabe et al. (2008).

Furthermore, procedure models for DF are another class of procedure models identified in the literature. Many of the models in the established literature are particularly characterized by the fact that they are applied in a military context. In Table 2, an overview of different DF procedure models is presented.

Table 2: Overview of DF procedure models.

Source	# Phases	Relevant phases
Lewis et al. (2001)	6	Edit scenario and DF parameters, execute DF space and scenario, explore results.
Horne and Meyer (2005)	5	Create scenarios, define and prepare data space, scenario execution, analysis and data mining, execute and examine scenarios.
Choo et al. (2008)	6	Scenario specification, design of experiments, simulation models, DF, regression and clustering.
Kallfass and Schlaak (2012)	5	Scenario definition, model and scenario, design of experiments, high performance computing, data analysis and visualization.
Horne and Seichter (2014)	5	Model development, rapid scenario prototyping, design of experiments, high performance computing, analysis and visualization.
Krol and Kitowski (2014)	4 (+2 additional)	Defining simulation, preparing input data, performing simulation, output data analysis.
Bruvoll et al. (2015)	7	Define objectives, design and develop simulation environment, execute simulation, analyze data and evaluate results.
Feldkamp et al. (2015)	4	Smart experiment design, data mining, visual analytics for input and output data.
Schubert et al. (2017)	5	Model development and rapid scenario prototyping, design of experiments, high performance computing, analysis and visualization with tool support.

Despite differences in structure, complexity, and granularity, overlaps can be found between the DF procedure models with regard to their main phases. For example, many models include phases for model development, DOE, experimentation, and analysis. As a reference and example, Figure 2 shows the established procedure model by Feldkamp et al. (2015), that was developed for the use in a manufacturing context.

As a first interim conclusion it can be stated that the literature overview in Table 1 and Table 2 as well as the following short discussion show similarities among a class of procedure models, but also between the classes of simulation and DF models. However, the simulation procedure models differ in important aspects, which is especially considered with respect to V&V. Moreover, it shows that most of the models are not specifically designed for the application in LN. Considering procedure models for DF, the literature shows that most of the models are designed for the application in military studies. Although DF procedure models have been introduced for DF in other domains, e.g., manufacturing, the authors of this paper are convinced that the models must consider the whole process of a DF study in a project context in LNs.

3 A DATA FARMING PROCEDURE MODEL

In the context of a farming for mining study for LNs, procedure models are an essential element of a method for a structured and targeted implementation. The related work presented and discussed in Section 2 shows a discrepancy between DF procedure models and established simulation procedure models. The goal of the new DF procedure model to be developed is to resolve the discrepancy between both classes of procedure models and to integrate and expand the DF procedure models into the established simulation standard. In

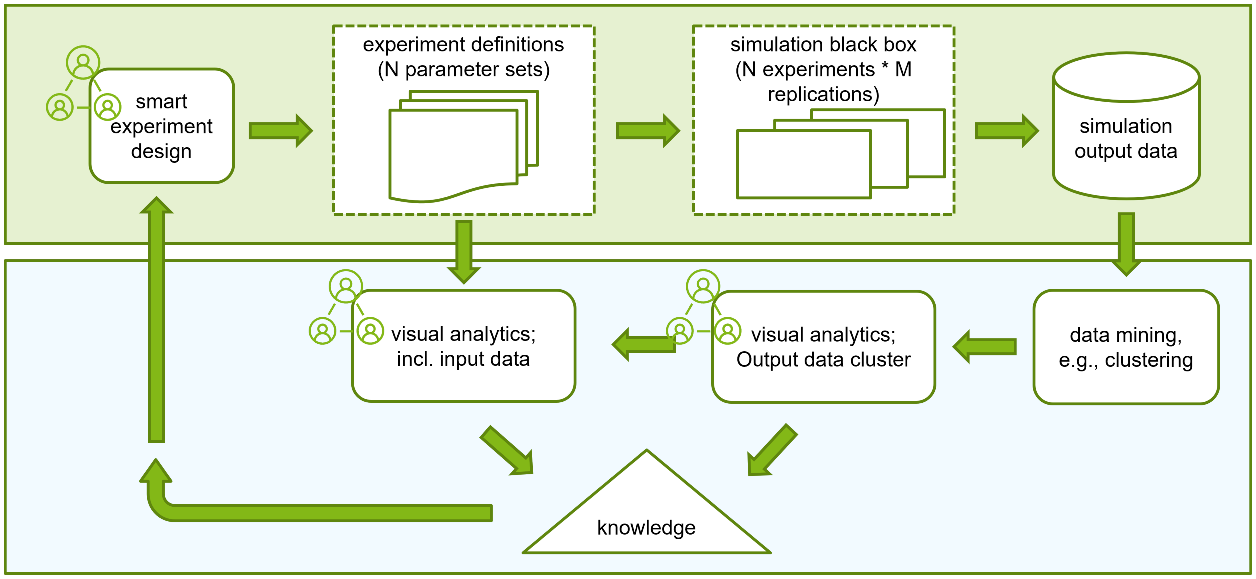
smart experiment design
experiment definitions (N parameter sets)
simulation black box (N experiments * M replications)
simulation output data
data mining, e.g., clustering
visual analytics; Output data cluster
visual analytics; incl. input data
knowledge

Figure 2: Process of knowledge discovery in simulation data according to Feldkamp et al. (2015).

order to determine the position of the procedure model within the farming for mining methodology, an overview of the method is shown in Figure 3.

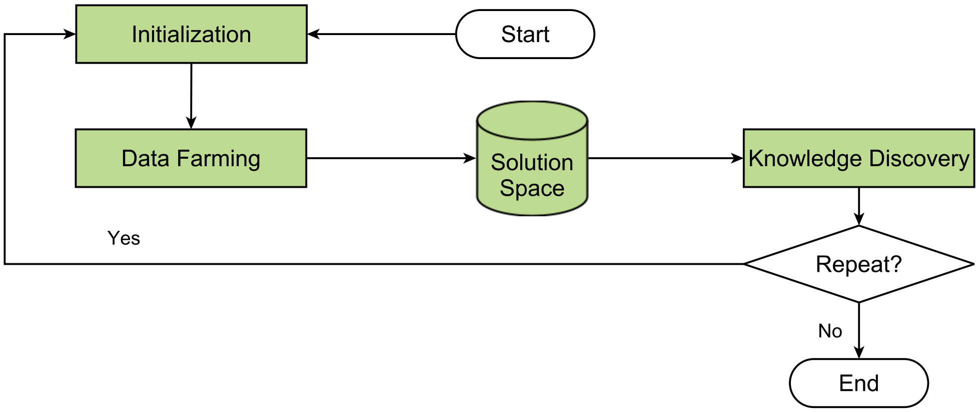

Figure 3: Overview of the Farming-for-Mining methodology.

The method consists of four main elements as shown in the figure. The first element is the initialization. We kindly refer the reader to Hunker et al. (2021) for a more in-depth discussion about the initialization and the connection between DF and KDD. DF is the second element of the method and describes the process of generating synthetical data by means of simulation. The third element of the method is the solution space, which includes a database system to manage the synthetical data that are generated by the DF. This concerns, for example, means to describe the structure of the synthetical data and storing and retrieving data in a purposeful way for the following KDD process. The process of knowledge discovery describes the fourth element of the methodology, which is concerned with data mining and the presentation and visualization of discovered correlations. The method provides the option of a cyclic approach to reuse insights gained during execution. We refer to this in our framework as the global cycle, because it takes place at the method level. As described above, procedure models are often part of a method. The new DF procedure model details the DF element as shown in Figure 3. To develop a new DF procedure model, we selected one model from each class of procedure models, simulation and DF, as a reference. First, we chose the simulation procedure model by Rabe et al. (2008) as reference model A (RMA), since it is well established and part of a guideline. The model is shown in Figure 1. Second, we chose the DF procedure model introduced by Feldkamp et al. (2015) as reference model B (RMB). The model is shown in Figure 2. This procedure model has, compared to other procedure models discussed before, with its focus a different application domain than military.

From both Section 2 and the introduction to Section 3, we can derive the following requirements and constraints for which our research shows that these are viable to derive a new DF procedure model. This concerns, for example, the following requirements:

- **Project in LNs:** DF is applied in form of a farming for mining project in LNs. A procedure model needs to include typical phases in form of initialization and evaluation.
- **Granularity:** DF as a process needs to be divided into an adequate number of phases. This is in particular relevant for the practical use of the model within LNs.
- **System formalization:** Integration of phases that have the goal of creating and deriving an appropriately formalized, executable simulation model or adapt an existing model.
- **DOE:** The experiment design is one of the main components of DF. A procedure model for DF must explicitly take this into account and integrate it accordingly as a phase, since a targeted DOE is indispensable for large-scale simulation experiments.
- **Data handling:** Since the procedure model is used to generate vast amounts of data based on a simulation model, relevant phases need to be included accordingly to manage the input and output data.
- **V&V:** A thorough and systematic V&V across all phase results is indispensable in the context of LNs.
- **Practical relevance**: This concerns, for example, the accompanying documentation of procedure models and the clear labeling of phases in terms of task and objective.

A first analysis shows that neither RMA nor RMB satisfy all requirements. However, the established procedure models provide an excellent basis for deriving synergies and we, therefore, combine and adapt them in order to benefit from their respective strengths. At this point, it should be noted that the following argumentation can be applied in wide parts to all of the above-mentioned procedure models. The requirements of a project in LNs can be fulfilled in the procedure models, for example, via suitable phases of initialization and evaluation. If both references are considered, this is partially fulfilled by RMA but not by RMB. RMA shows a phase for task definition with the result of a task specification based on sponsor needs, but is missing an evaluation phase in the later stages of the procedure model. RMB starts directly with the phase of smart experiment design and, therefore, does not consider any phases for a project initialization and, furthermore, no phase for an evaluation of results. Since DF requires an executable model, the procedure model needs to include phases for the development or adaptation of a simulation model. RMA differentiates three detailed phases from the conceptualization to the formalization to the implementation of a executable model. RMB does not include phases for a model development, since it is assumed that a simulation model is already available at the start of the procedure model. The requirement of including a DOE can be included by phases considering factor analysis, appropriate choice of an experiment design and creating a design matrix. This requirement is not met by RMA. In contrast, RMB includes specific phases for a DOE. The purpose of DF is to generate a vast amount of synthetical data, which can be analyzed accordingly. This has two implications. On the one hand, input data are necessary to create or adapt a simulation model. For a targeted DOE, they have to be gathered from different data sources of an LN. On the other hand, the structure and the amount of result data have to be stored appropriately in order to make them available to an analysis process of knowledge discovery. This must be ensured via suitable phases of data preprocessing and data postprocessing in the procedure models. This is partially the case in RMA, which includes phases to manage input data, whereas RMB includes phases (and respective tools) concerning the storage of synthetical data. V&V is of utmost importance in the domain of LNs, since a decision can include high risks that cannot be anticipated directly. It follows that wrong decisions based on wrong simulation results can lead to incalculable costs that exceed the cost of adequate V&V in a DF study. Therefore, a systematic approach is necessary, which is included across all phase results in RMA, whereas RMB does not include specific phases for a structured V&V. The final requirement concerning practical relevance is fulfilled by RMA and partially by RMB. Both procedure models are, although specific, well documented. With regards to labeling phases, the clear designation of phases is sometimes too unspecific in RMB, for example, concerning a phase called ”smart” experiment design. As a conclusion from the analysis of both reference models, it can be stated that both procedure models offer an excellent basis for deriving a new procedure model specifically geared to the DF in LNs and, thus, compensating for the weaknesses in the aforementioned context.

In the following, we systematically present the structural expansion from RMA and RMB. Figure 4 shows our proposed new DF procedure model in the context of LNs.

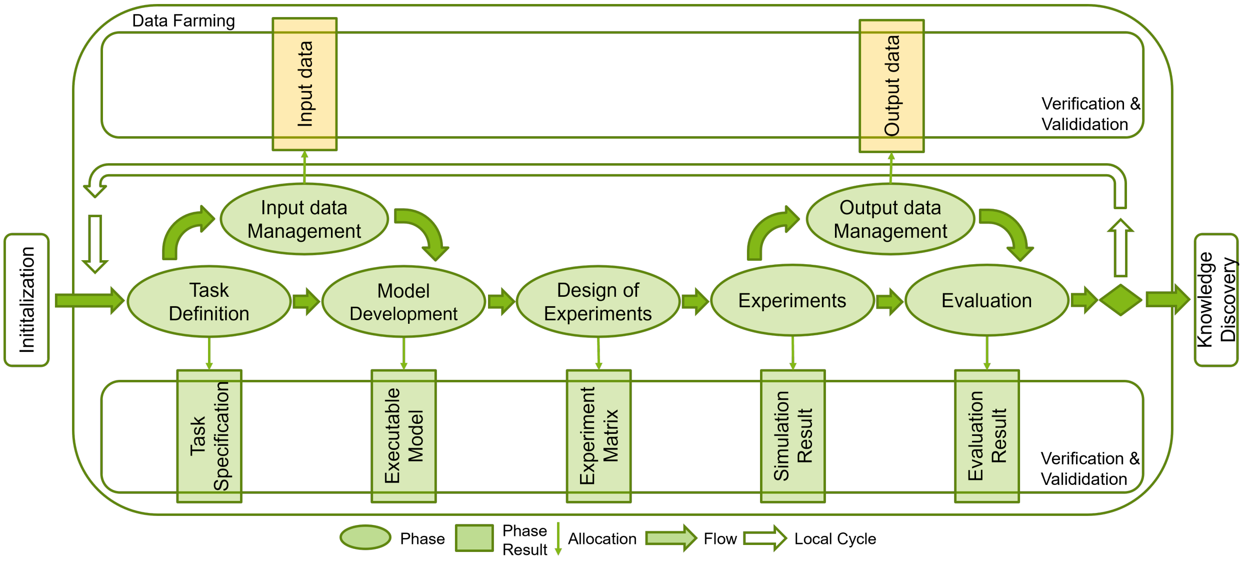

Figure 4: New DF procedure model.

Starting from the initialization method element, the DF starts with a task definition phase. The procedure model can be roughly divided into two parts, data management as well as modeling and experiments. These parts are interrelated and in part also take place in parallel, for example, with regard to experiments and output data management. In principle, the procedure model adopts the structural setup of RMA, i.e., phases and associated phase results are defined. All phase results are subjected to systematic V&V, meeting the requirement that comprehensive V&V should be carried out during all phases of the procedure model, which refers to the results of the phases due to the implicit test requirement. In view of a project-based implementation of a DF study, a task definition has to develop a common understanding of the DF task. This phase takes the results from the initialization to detail and process them in view of further phases, model development and input data management, resulting in a task specification that includes information on required system properties and boundary conditions. The phase of input data management contains activities to select and collect observational data from multiple sources of a LN and to process them for use as valid input data for the model development and DOE. The prerequisite for this is an initial investigation of the LN’s data landscape, in particular to uncover logical relationships. A further activity in this phase is an initial concretization of possible result data, for example with regard to structure, granularity, and storage. The model development summarizes the three phases of model development from RMA: system analysis, model formalization, and implementation. Due to the close proximity of the three reference phases in terms of content and expertise, they are combined in the new procedure model as individual activities in one phase. The result of the new phase is a farmable model implemented in a simulation software. The following phase deals with the DOE. This includes activities for identifying relevant factors, choosing an appropriate design, and necessary steps to generate a design matrix in a suitable format. This phase results in a design matrix, which in the following phase is used to conduct the experiments using the simulation model. Activities in this phase include planning the runs and replications, setting a seed, determination of the number of replications, setting up the transient phase (in non-terminating systems), and exection of the experiments. This results in a large amount of synthetical data as a phase result, which in turn have to be stored adequately. To guide the activities for handling the results of the simulation experiments, a phase for output data management has been integrated in the procedure model. Activities include plausibility checks, structuring, and processing of result data. Ultimately, the generated synthetical data will be stored persistently, e.g., in a database. This is the basis for a final phase of evaluation. The activities in this phase include planning and execution of statistical tests, e.g., a chi-squared test or using metamodeling to analyze input-output-behaviour. This provides a final assessment of the DF process and results before the synthetical data are used as input for the following knowledge discovery process. The DF procedure model does not provide a separate phase for further analysis. This is outsourced to a subsequent process of knowledge discovery. The procedure model also provides a local cycle, which allows for a repetition of the phase in case of a negative evaluation result.

4 SOFTWARE IMPLEMENTATION

Since the process of DF cannot be done manually and automation to run the experiments is a key element, we have developed and implemented a software concept to control the process of DF for knowledge discovery. The software helps decision makers in LNs by guiding the process of setting up a DF study with our proposed procedure model as part of the farming for mining method. The respective architecture for the software is shown in Figure 5.

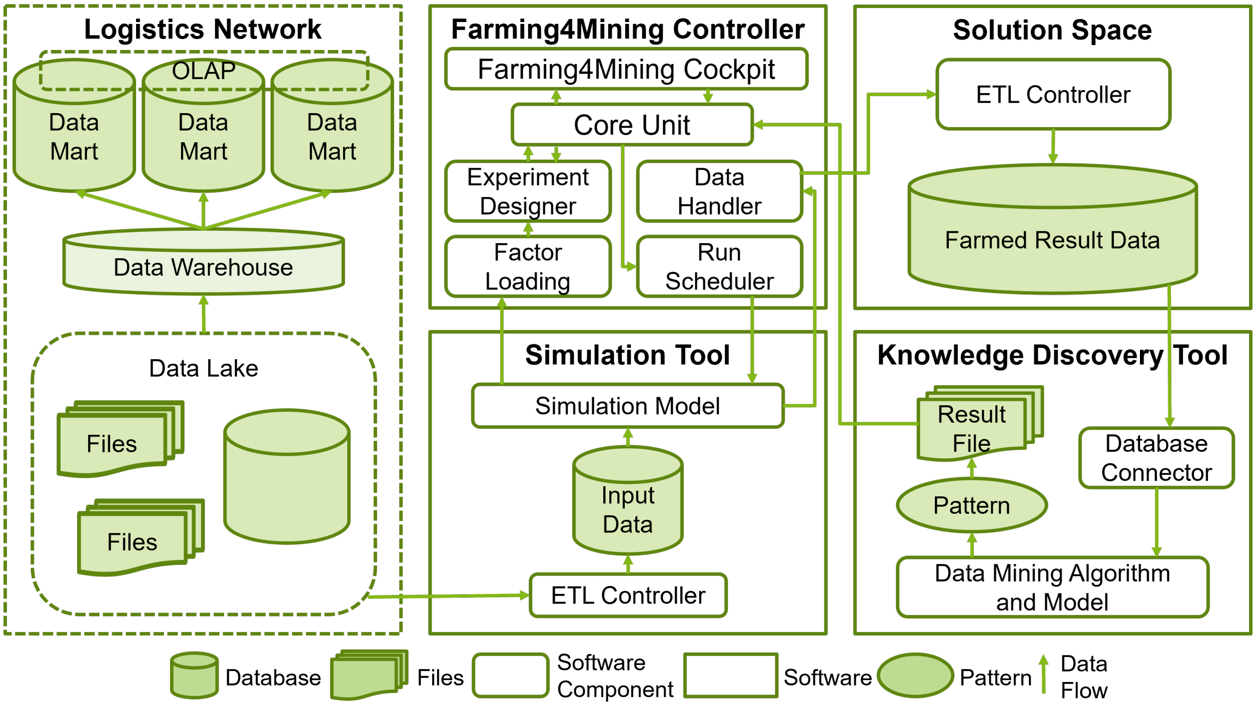

Figure 5: Architecture for the farming for mining software.

Different data sources of the data lake in the corresponding exemplary LN are shown on the left side. Furthermore, various elements for the existing data analysis processes in the LN are shown, with a data warehouse and specific data marts. The central part of the software architecture is the DF controller, which is a software written in Java. It contains a cockpit, which provides a surface for the decision maker and is, for example, responsible for the visualization of results. The core unit is positioned below the surface and centrally manages all control activities. It handles the design of experiments, offering different designs and a selection guide out of the box for the end user based on the different factors that are loaded from the simulation model. The simulation environment is another software element in the architecture. It consists of a database that is connected to a simulation model. The data stored in the database are observational data from the LN and used via extract, tranform, and load to instantiate the simulation model. In our case, the simulation tool we use for the experimentation with the software framework is Plant Simulation (Bangsow 2020). Furthermore, the core of the farming controller manages the experiments, which are run with the simulation model. Result data are stored in a central results database via a specific data handler. This is located in a software that is called solution space. This is the basis for the final element of the architecture, the knowledge discovery software. In our case, we use the established software RapidMiner (Kotu 2015). The synthetical data are loaded via a database connector into the software. After running data mining, the gained patterns are exported into files using the standard export features of RapidMiner. Result patterns (e.g., rules, trees, graphs, functional relations) are in the following process visualized in the cockpit, allowing for filtering of different patterns (e.g., based on their value of interestingness) and supporting intepretation and evaluation by the decision makers.

5 CONCLUSION AND OUTLOOK

After discussing the basics of knowledge discovery and data generation, the current state of the literature regarding procedure models from the areas of simulation and DF was examined. It has been found that the processes of simulation and DF are represented by two different classes of procedure models, which are, however, essentially based on congruent approaches. Thereupon, a well-structured procedure model was derived specifically for the domain of DF in LNs. This was adapted on the basis of two established, suitable procedure models from the respective classes. In this way, the separation of the two procedure model classes, which in our view does not lead to the desired results, is eliminated in order to combine the strengths of the respective procedure models as a synergy. This is, for example, shown in the emphasis and integration of two core elements from each class of procedure models for

a) V&V, since it is important to detect possible invalid phase results in a DF study as soon as possible to prohibit wrong inputs for the KDD process and subsequently invalid conclusions and decisions for a LN, and
b) DOE, as the complexity for LNs is reflected in the factors and is, thus, indispensable for a targeted DF study, especially with regard to the knowledge discovery process.

The integration and use of the procedure model in our framework proved to be very promising in theoretical and practical application. Furthermore, the refinement and balancing of the phases based on practical feedback from experts is an important part of further research, as especially the phases of data management are characterized by a multitude of complex steps. The software will be further developed to support decicison makers in LNs. This concerns for example the automation of V&V techniques, since manual V&V is not purposeful due to the complexity. In order to demonstrate the practical application possibilities, case studies using the developed procedure model as part of the overall methodology are a further research step. These will be worked on for a straightforward and targeted practical applicability in the context of a farming for mining approach in LNs.

REFERENCES

Aggarwal, C. C. 2015. *Data Mining*. Cham: Springer International Publishing.

Balci, O. 1998. "Verification, Validation, and Testing". In *Handbook of Simulation*, edited by J. Banks, 335–393. Hoboken: John Wiley & Sons.

Bangsow, S. 2020. *Tecnomatix Plant Simulation*. Cham: Springer International Publishing.

Banks, J. 2010. *Discrete-Event System Simulation*. 5th ed ed. Upper Saddle River, NJ: Pearson.

Barton, R. R. 2020. "Tutorial: Metamodeling for Simulation". In *Proceedings of the 2020 Winter Simulation Conference (WSC)*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, and T. R. T. Roeder, 1102–1116. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Bramer, M. 2016. *Principles of Data Mining*. London: Springer.

Brandstein, A. G., and G. E. Horne. 1998. *Data Farming: A Meta-technique for Research in the 21st Century*. Quantico, Virginia: Marine Corps Combat Development Command Publication.

Brinkkemper, S. 1996. "Method Engineering: Engineering of Information Systems Development Methods and Tools". *Information and Software Technology* 38(4):275–280.

Bruvoll, S., J. E. Hannay, G. K. Svendsen, M. L. Asprusten, K. M. Fauske, V. B. Kvernelv, R. A. Løvlid, and J. Hyndøy. 2015. "Simulation-Supported Wargaming for Analysis of Plans". In *NATO Modelling and Simulation Group Symposium on M&S Support to Operational Tasks Including War Gaming, Logistics, Cyber Defence (STO-MP-MSG-133)*, Paper No. 12: NATO Science and Technology Organization.

Chew, J., and C. Sullivan. 2000. "Verification, Validation, and Accreditation in the Life Cycle of Models and Simulation". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 813–818. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Choo, C. S., E. C. Ng, D. Ang, and C. L. Chua. 2008. "Data Farming in Singapore: A Brief History". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 1448–1455. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cleve, J., and U. Lämmel. 2014. *Data Mining*. München: De Gruyter.

Dross, F., and M. Rabe. 2014. "A SimHeuristic Framework as a Decision Support System for Large Logistics Networks With Complex KPIs". In *Proceedings of the 22nd Symposium Simulationstechnik 2014*, edited by J. Wittmann and C. Deatcu, 247–254. Wien, Österreich: Arbeitsgemeinschaft Simulation.

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. 1996. "From Data Mining to Knowledge Discovery in Databases". *AI Magazine* 17(3):37–54.

Feldkamp, N., S. Bergmann, and S. Straßburger. 2015. "Visual Analytics of Manufacturing Simulation Data". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, Chan, W. K. V., I. Moon, Roeder, T. M. K., C. Macal, and M. D. Rossett, 779–790. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Freitag, M., M. Kück, A. Ait Alla, and M. Lütjen. 2015. "Potenziale von Data Science in Produktion und Logistik: Teil 1: Eine Einführung in aktuelle Ansätze der Data Science". *Industrie 4.0 Management* 31(5):22–26.

García, S., J. Luengo, and F. Herrera. 2015. *Data Preprocessing in Data Mining*, Volume 72. Cham: Springer.

Horne, G., and T. Meyer. 2016. "Data Farming Process and Initial Network Analysis Capabilities". *Axioms* 5(1):1–17.

Horne, G., and S. Seichter. 2014. "Data Farming in Support of NATO Operations – Methodology and Proof-of-Concept". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and Miller J. A., 2355–2363. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Horne, G. E., and T. E. Meyer. 2005. "Data Farming: Discovering Surprise". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joins, 1082–1087. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hunker, J., A. Wuttke, A. A. Scheidler, and M. Rabe. 2021. "A Farming-for-Mining-Framework to Gain Knowledge in Supply Chains". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kallfass, D., and T. Schlaak. 2012. "NATO MSG-088 Case Study Results to Demonstrate the Benefit of Using Data Farming for Military Decision Support". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 2481–2492. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kotu, V. 2015. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Waltham, MA: Morgan Kaufmann.

Krol, D., and J. Kitowski. 2014. "Towards Adaptable Data Farming in Clouds". In *Proceedings of the 2014 IEEE Fourth International Conference on Big Data and Cloud Computing*, edited by J. Chen and L. T. Yang, 283–284: Institute of Electrical and Electronics Engineers, Inc.

Law, A. M. 2015. *Simulation Modeling and Analysis*. Fifth ed. New York: McGraw-Hill Education.

Lewis, A., J. Berlin, T. Meyer, S. Kruglikov, S. Miller, J. W. Lyver, and R. Gharavi. 2001. "An Information System for Distillation Data Farming". In *Proceedings of the Thirteenth International Conference on Scientific and Statistical Database Management*, edited by L. Kerschberg and M. Kafatos, 274–277: Institute of Electrical and Electronics Engineers, Inc.

Rabe, M., S. Spieckermann, and S. Wenzel. 2008. "A New Procedure Model for Verification and Validation in Production and Logistics Simulation". In *Proceedings of the 2008 Winter Simulation Conference (WSC)*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 1717–1726. Piscataway: Institute of Electrical and Electronics Engineers, Inc.

Runkler, T. A. 2010. *Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*. Wiesbaden: Vieweg+Teubner.

Sanchez, S. M. 2006. "Work Smarter, not Harder: Guidelines for Designing Simulation Experiments". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 47–57. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sanchez, S. M. 2018. "Data Farming: Better Data, Not Just Big Data". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, A. Mustafee, S. J. Skoogh, and B. Johansson, 425–439. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schubert, J., S. Seichter, A. Zimmermann, D. Huber, D. Kallfass, and G. Svendsen. 2017. "Data Farming Decision Support for Operation Planning". In *Proceedings of the 11th NATO Operations Research and Analysis (OR&A) Conference*, 7.2–1 – 7.2–20: NATO Science and Technology Organization.

Shannon, R. E. 1998. "Introduction to the Art and Science of Simulation". In *Proceedings of the 1998 Winter Simulation Conference*, edited by D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 7–14. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Stoldt, J., and M. Putz. 2017. "Procedure Model for Efficient Simulation Studies which Consider the Flows of Materials and Energy Simultaneously". *Procedia CIRP* 61:122–127.

ten Hompel, M., and T. Schmidt. 2008. *Warehouse-Management: Organisation und Steuerung von Lager- und Kommissioniersystemen*. 3 ed. Berlin: Springer.

Teniwut, W. A., and C. L. Hasyim. 2020. "Decision Support System in Supply Chain: A Systematic Literature Review". *Uncertain Supply Chain Management* 8(1):131–148.

Verein Deutscher Ingenieure 2014. *VDI 3633 – Simulation of Systems in Materials Handling, Logistics and Production: Fundamentals*. Berlin, Germany: Beuth.

AUTHOR BIOGRAPHIES

JOACHIM HUNKER is a researcher at the department IT in Production and Logistics at the TU Dortmund University. He holds a Master of Science in Logistics, Infrastructure, and Mobility with a focus on IT in Logistics from the Technical University of Hamburg. He graduated with a master thesis on a hybrid scheduling approach of assembly lines of car manufacturers. His research focuses on simulation-based data generation and data analytics in logistics. His email address is joachim.hunker@tu-dortmund.de.

ANNE ANTONIA SCHEIDLER is a researcher at the department IT in Production and Logistics at the TU Dortmund University. Until 2012, she worked as IT consultant for different large companies. Key areas of her activity were business processes modeling and data concepts. She graduated in 2017 with a PhD thesis on methods for knowledge discovery using data patterns. Currently, her research focus is on Data Mining, concepts for input data, data in simulation, and supply chain information. Her e-mail address is anne-antonia.scheidler@tu-dortmund.de.

MARKUS RABE is a full professor for IT in Production and Logistics at the TU Dortmund University. Until 2010 he had been with Fraunhofer IPK in Berlin as head of the corporate logistics and processes department, head of the central IT department, and a member of the institute direction circle. His research focus is on information systems for supply chains, production planning, and simulation. Markus Rabe is vice chair of the "Simulation in Production and Logistics" group of the simulation society ASIM, member of the editorial board of the Journal of Simulation, member of several conference program committees, has chaired the ASIM SPL conference in 1998, 2000, 2004, 2008, and 2015, Local Chair of the WSC'2012 in Berlin and Proceedings Chair of the WSC'2018 and WSC'2019. More than 200 publications and editions report from his work. His email address is markus.rabe@tu-dortmund.de.

HENDRIK VAN DER VALK is a researcher at the Chair for Industrial Information Management at the TU Dortmund University and at the Fraunhofer Institute for Software and Systems Engineering. He holds a Master of Science in Mechanical Engineering with focus on simulation in logistics from TU Dortmund University. His research focuses on digitalisation of logistics as well as designing a reference architecture for digital twins and their connection to IoT platforms. He conducts his research as member of the Excellence Center for Logistics and IT. His e-mail address is hendrik.van-der-valk@tu-dortmund.de.