

On Reusing Data Mining in Business Processes - A Pattern-Based Approach

Dennis Wegener and Stefan Rüping

Fraunhofer IAIS, Schloss Birlinghoven,
53754 Sankt Augustin, Germany

{dennis.wegener, stefan.rueping}@iais.fraunhofer.de

Abstract. Today's business applications demand high flexibility in processing information and extracting knowledge from data. Thus, data mining becomes more and more an integral part of operating a business. However, the integration of data mining into business processes still requires a lot of coordination and manual adjustment. This paper aims at reducing this effort by reusing successful data mining solutions. We describe a novel approach on facilitating the integration based on process patterns for data mining and demonstrate that these patterns allow for easy reuse and can significantly speed up the process of integration. We empirically evaluate our approach in a case study of fraud detection in the health care domain.

Keywords: Data Mining Patterns, Business Processes, Reuse and Integration, BPM, CRISP.

1 Introduction

Businesses need to be more and more flexible in order to be competitive in today's economy. Two big forces that drive flexibility are data mining, which helps enterprises to understand their customers, processes, and themselves better, and service oriented architectures (SOA), which help to be faster in implementing new business strategies and products. As data mining becomes more and more an integral part of executing a business, data mining functionality needs to be integrated into SOA in the context of existing applications [1].

In previous work [2] we presented an initial discussion on how to integrate data mining into business processes. Here, we focus on how to enable the reuse of existing solutions that have been proven to be successful. Currently, there are two approaches of achieving this: Passing through a new CRISP process [3] while being inspired by existing solutions (e.g. by personal experience or reading respective documentation and scientific papers), and reuse at implementation level (e.g. by copy-and-paste of existing code and workflows). In this paper we argue that there is a need for an approach in between.

The first approach for reuse by following the CRISP process, which describes the general procedure of performing data mining projects, does not suffice, as this approach is often too general. E.g. it has been identified that CRISP-DM lacks

in the deployment phase [4], in guidance towards implementing particular tasks of data mining methodologies [5], and in the definition of phases important for engineering projects [6]. In addition, we detected that many redundancies and inefficiencies exist when following the CRISP standard data mining process model in parallel to standard BPM approaches [2].

The second approach is to utilize reuse at implementation level. Modern SOA-based Business Process Management (BPM) environments, e.g. based on standards like BPEL [7] and BPMN [8], provide flexible and user friendly environments and tools for designing, deploying and managing business applications. Data mining solutions can be reused in such environments by making use of available data mining workflows and services. However, this approach is often too specific for being reusable efficiently. For example in an analysis of a large set of real-world data mining workflows [9] we found out that the changes of a workflow during the lifetime of a data mining project are made to the same extent at the preprocessing and at the modeling part, which implies that understanding and representing the semantics of the data is a very important step. The changes for the preprocessing part consisted to 50% and for the modelling part to 75% of manual parameter optimizations. The challenge is to reuse such kinds of manual fine-tuning. Also, there are requirements specific to data mining. For instance, the choice of data mining algorithm has impact on the data understanding and the business understanding phase. Thus, data mining is a complex process that requires a lot of manual optimizations and is not always transferable due to the dependency to the data. A copy-and-paste approach, by taking over the data mining part from another business process, will only work if the business process into which the data mining is integrated has exactly the same properties as the original one. What is needed is a way to specify the correct level of abstraction and generalization for enabling reuse.

In this work, we aim at a formal representation of data mining processes to facilitate their reuse in business processes. As visualized in Fig. 1, the approach should support the description of the process at different levels of abstraction between the CRISP model as most general representation and executable workflows and code as most concrete representation. This requires to:

1. support the modeling of the data mining process as a whole, including its unique requirements, according to the CRISP tasks and phases
2. allow for the modeling and specification of tasks at implementation level in order to make use of existing code and services
3. simplify the integration into business processes in the context of modern BPM systems, which requires compatibility to relevant standards and tools.

This work contributes a new process model for the reuse and integration of existing data mining solutions in business processes in the context of modern BPM environments. This includes, based on CRISP, the definition of data mining process patterns, which encode reusable data mining solutions at the appropriate level of generality. They will be defined specific to the data mining problems to

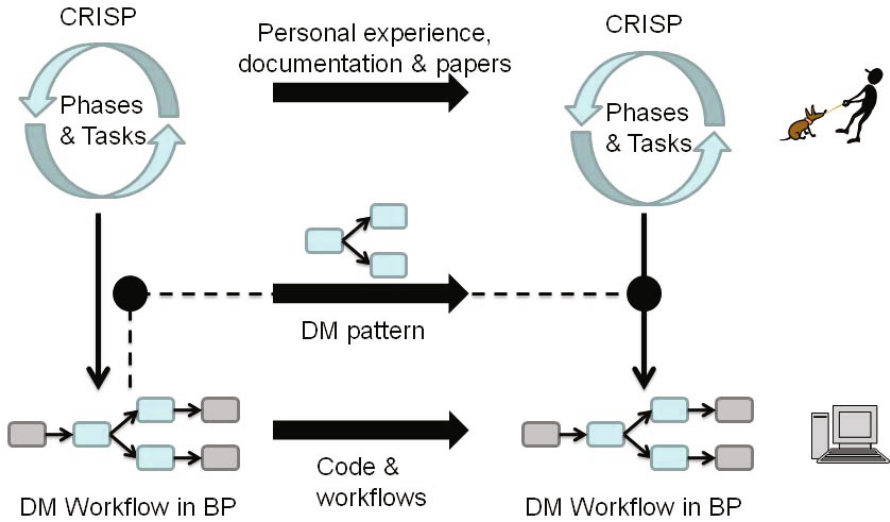


Fig. 1. Different strategies of reusing data mining

be addressed and will be based on a set of underlying data mining services. We will also define a process for applying these data mining patterns to business processes.

The remainder of the paper is as follows: Sec. 2 introduces the field of BPM and data mining and gives an overview over the state-of-the-art and related work. In Sec. 3 we introduce our approach for a new process model for the reuse and integration of data mining into business processes. In Sec. 4 we give a case study on the integration including an example of a data mining process pattern and its integration. Sec. 5 concludes.

2 State-of-the-Art and Related Work

A business process is a series of steps designed to produce a product or a service which includes all the activities that deliver particular results for a given customer [10]. Business process management (BPM) is a discipline combining software capabilities and business expertise to accelerate business process improvement and to facilitate business innovation. In the context of BPM there exists a variety of standards for process modeling, visualization and execution as well as tools supporting features like pre-deployment validation of processes and the automation of process deployment, hiding the complexity of the distribution middleware. BPEL [7] and BPMN [8] turned out to be one of the de-facto standards for the executable language for specifying interactions with web services resp. the graphical representation of processes. In the following, we will use the BPMN notation for the definition and visualization of process models.

Data mining, often also called knowledge discovery in databases (KDD), is the process of extracting (unknown) patterns from data [11]. There exists a variety

of different data mining methods and algorithms [12], which commonly involve the following classes of tasks: Inferring rudimentary rules, statistical modeling, constructing decision trees, constructing rules, mining association rules, linear models, instance-based learning, and clustering. CRISP-DM [3] is a standard process model for data mining that describes the life cycle of data mining projects as iterative process. The CRISP model consists of the following 6 phases:

- Business Understanding - understanding the project objectives and requirements from a business perspective and defining the data mining problem
- Data Understanding - getting to know the data and its quality
- Data Preparation - construct the final dataset from the initial raw data as input for the modeling
- Modeling - various modeling techniques are selected and applied, including the calibration of their specific settings
- Evaluation - assess how well the built model achieves the project objectives
- Deployment - the results of the data mining are delivered to the user.

Typically, the phases from Data Understanding to Evaluation are mainly performed only by the data miner, while the phases Business Understanding and Deployment involve both the data miner and the business user. The latter phases need a lot of knowledge transfer, coordination and manual adjustment.

In our work we refer to data mining as part of the application that is represented by a business process. Data Mining on event log data in order to construct processes (Process Mining) and data mining on business processes in order to improve the business processes as part of the field Business Intelligence are not in the scope of this paper. The question that we want to answer is how can we best reuse data mining as part of a bigger business processes.

In [13] a set of workflow patterns describing the control-flow perspective of workflow systems is defined. Such process patterns have plenty of advantages [14]: BPM processes serve as both the specification and the source code. The modeled processes become the solutions deployed and provide a simple communication tool between end-users, business analysts, developers and the management. Process patterns provide a proven and simple technique to shorten the learning curve and improve productivity and quality of the processes designed as they are simple to understand, learn and apply immediately. Thus, we propose to provide a similar approach for the integration and reuse of data mining - process patterns that represent templates for different data mining problems.

Java Data Mining (JDM) is a standard for developing data mining solutions [1]. JDM provides a web service interface which can be used to set up, integrate and manage data mining processes in the context of BPMN and BPEL based environments. BPEL processes are designed which include calls to the JDM web services. These services interface with the JDM API on top of a Data Mining Engine. However, they do not specify a general concept on how to redesign the business processes in detail and which requirements exist for the modeling of the business process when including data mining.

In [15] a dynamic data mining process system is introduced. The idea is to set up each data mining activity as web service, to model and execute the data

mining process in a BPEL environment and to get a PMML compliant model as result. The authors only present a single application as example and do not give a concept on how to model the data mining processes in general.

In [4] a methodology for the implementation of data mining into operational business processes is proposed, consisting of the phases Exploratory Data Mining, Deployment of IT into the Business Process, and Operational Data Mining. However, the authors do not specify detailed concepts for the modeling of the data mining process as part of the business process. In our work, we focus on how to perform the process (re-)design from the modeling point of view and aim at removing complexity for modeling and integration.

In [16] the authors propose a framework for the reuse of scientific workflows, which are also based on reusable patterns that include abstract tasks. However, their work focuses rather on grid-related tasks like copying, job execution and monitoring than on data mining specific tasks.

3 Process Patterns for Integration and Reuse

We aim at reusing existing data mining processes that have proven to be successful, and hence want to develop a formal and concrete definition of the steps that are involved in the data mining process and of the steps that are necessary to reuse it in new business processes. Thus, we focus on the reuse rather than on the data mining problem itself and consider a solution for the data mining problem to be available. In the area of data mining there exist a lot of situations and scenarios where existing solutions are reusable, especially when no research on new algorithms is necessary. These solutions can have different levels of generality. Lots of examples and ready-to-use algorithms are available in toolboxes, which only have to be integrated. However, due to a lack of formal support in practice a reuse and integration of existing solutions is not often or only informally done, which leads to a lot of unnecessary repetitive work. By our approach we want to facilitate the integration and reuse of these data mining solutions.

In the following, we will describe our approach in 3 steps. First, we will show how to modify the CRISP model to focus on the special case of reuse of existing solutions. Second, we will define the concept of data mining patterns. These patterns are created after the initial data mining process ended and include the definition, description and requirements of the data mining process, but are independent of the application scenario. The goal of these data mining patterns is to provide a flexible representation for different levels of generality. Third, we will describe how these patterns are applied for reuse by a meta process. This process will be started if a new application scenario is available and describes the tasks that can only be performed with knowledge on the application scenario.

3.1 Reuse and the CRISP Model

In the following, we will present how the CRISP phases and tasks from [3] differ in the case of executing CRISP from scratch and reusing existing solutions. Fig. 2 visualizes our approach.

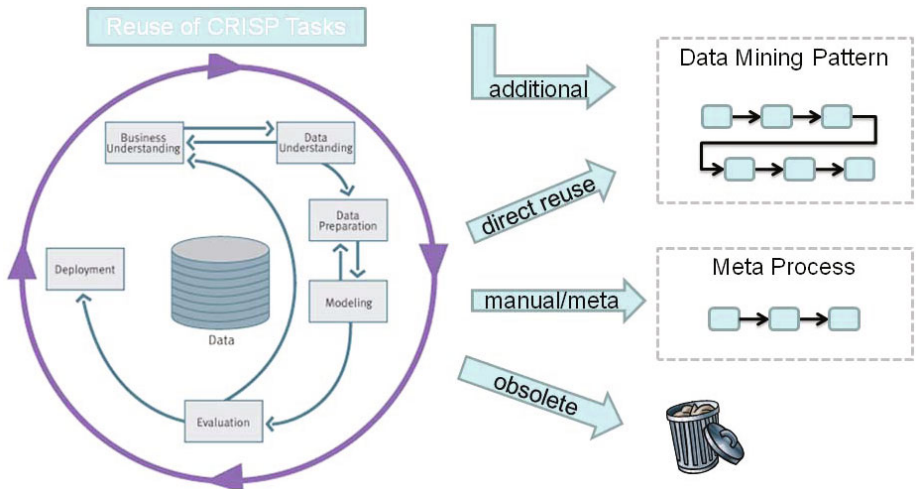


Fig. 2. Mapping CRISP tasks to data mining patterns and a meta-process for reuse

Business understanding phase. The task *Determine Business Objectives* is a general task that is independent of the data mining. We arrange this task at the start of the meta-process, as it provides the information needed for the choice of the pattern. The task *Assess Situation* involves the setup of an inventory of resources, a collection of requirements, assumptions etc. In our scenario, this task does not apply as the important information is already available through the business process. In addition, the data mining goal is already specified by the data mining pattern. Thus, we transform the *Determine Data Mining Goals* task into a task that checks if the data mining goal is still matching and arrange it at the beginning of the data mining pattern. In the deployed executable process, this task will be obsolete. From here, the data mining specific tasks start. In our scenario the *Produce Project Plan* task is outside of the scope, as the project plan consists of following the proposed approach for the integration.

Data understanding phase. The tasks *Collect Initial Data*, *Describe Data* and *Explore Data* are obsolete as we assume the data to be available through the modeled business process. The verification of the data quality is arranged as second task *Verify Data Quality* in the pattern. In the deployed executable process, this task will be obsolete as well.

Data Preparation phase. The *Select Data* task is a preprocessing task of the pattern. E.g., a *Collect Data* task and a *Collect Label* task (in case of supervised learning methods) or a *Data Available* gateway could be part of the BPMN pool that contains the tasks that have to be connected to the tasks of the business process. The further tasks *Clean Data*, *Construct Data*, *Integrate Data* and *Format Data* are also preprocessing task of the pattern.

Modeling phase. The tasks *Select Modeling Technique* and *Generate Test Design* do not apply, as this information is already contained in the pattern. The

Build Model task represents the training of the data mining model and specifies the modeling technique and is part of the pattern. The *Assess Model* task is split into the task *Testing* and the gateway *Model Finalized* inside the pattern. They specify the test design and the decision on when to deploy a model.

Evaluation phase. The task *Evaluate Results* involves a matching with the business objectives. Thus, we arrange this task at the meta-process. With this task, the data mining specific part ends. In contrast to CRISP, the evaluation takes place after the deployment of the process. The task *Review Process* is implicit contained in loops of the meta-process (changing the specification of tasks of a data mining pattern or choosing another pattern). The task *Determine Next Steps* does not apply as the next steps are defined by the meta-process.

Deployment phase. The planning of the deployment by the task *Plan Deployment* does not apply, as the way how to deploy the model and the process as a whole are defined by the pattern approach. The deployment itself is two-fold. On the one hand, there is the deployment of the data mining pattern into the business process. This task is arranged in the meta-process. On the other hand, there is the deployment of the data mining model, which is part of the data mining pattern. The task *Plan Monitoring and Maintenance* does not apply as well, as monitoring and maintenance is a general part of the BPM context. The tasks *Produce Final Report* and *Review Project* are outside of the scope, as we are not interested in such a kind of deployment and as a review of the steps of the overall process are considered to be already part of the BPM solution.

Additional Tasks. In addition to the tasks of the CRISP phases there are additional tasks necessary for the deployment, management and usage of the data mining model. Typically, business processes are not designed for ad-hoc but for long term usage. Thus, the data mining has to be able to adapt to changes in the data during time, e.g. by re-building and deploying the data mining model periodically. For re-building a model, a *Re-build model* gateway activity is arranged before the tasks for preprocessing and building the model. The task *DeployModel* is performed after the model evaluation task. In addition, a *Model apply* task for applying the model to new datasets and related additional preprocessing task are part of the data mining pattern.

3.2 Data Mining Patterns

In the following, we will present our approach for the specification of data mining patterns at different levels of generality. The CRISP-DM methodology includes a four-level breakdown, which describes the instantiation of the CRISP process model in order to get a CRISP process. The 6 CRISP **phases** consist of several **generic tasks** which cover all possible data mining applications. Out of these tasks **specialized tasks** are described which specify specific actions for a certain situation. Finally, the **process instances** represent a record of actions and decisions of an actual data mining engagement. This top-down approach is a

manual process which is not automated. In our approach, the CRISP breakdown is partially pre-defined by the process patterns, where some of the tasks are already defined on a detailed level.

The data mining tasks, including requirements and preconditions, can be modeled in a uniform way in the context of BPM by representing each task as a task in the business process. However, we need to take into account that reuse may in some cases only be possible at a general or conceptual level. Our approach needs to include the description of the general CRISP model, of executable workflows, and of abstractions in-between. Thus, we model the tasks different levels of granularity, resulting in the following hierarchy of tasks: **conceptual** (manual tasks, only a textual description is available), **configurable** (semi-automatic tasks, code is available but parameters need to be specified), and **executable** (automatic tasks, code and parameters are specified).

Process patterns are only valid for process execution if they are specified to the executable level. Some general CRISP tasks, e.g. *Determine Business Objectives*, cannot be specified as technical step. Thus, the executable specification for such tasks would be a check task or empty/obsolete task. In a meta-process, which describes how to apply a given pattern to a business process, these tasks are specified (with different choices) according to the levels of the hierarchy. E.g., a *Clean Data* task could be specified as human task, as code that deletes records with missing values or as separate DM process for the prediction of missing values. Details on the meta-process will be given later in Section 3.3.

Definition. We consider the phases and tasks of the modified CRISP process from 3.1, including additional tasks and information that are necessary for an execution of such a process, as the most general data mining pattern. Every concretion of this process for an application according to the presented hierarchy is also a data mining pattern.

The idea is to be able to describe all data mining processes, while the description is as detailed as adequate for the given scenario. A data mining pattern only contains tasks that can be specified without having to know the business process in order to keep the feature of reusability. Obviously, the practical relevance of a pattern depends on its generality. An example of a data mining pattern will be given later in Sec. 4.

3.3 The Meta-process

The process of applying a pattern for the integration of data mining into a given business process can be modeled as process in the BPM context. In our approach, this meta-process describes the steps needed to use a pattern for a given business process. Fig. 3 visualizes the meta-process and its steps. The first task is to define the business objectives of the application. After that, a data mining pattern is selected that matches with these business objectives. Then, the tasks of the selected pattern are specified to the executable level according to the hierarchy of our approach proposed or a given specification is chosen. If it is observed that the pattern cannot be specified as executable, the meta-process

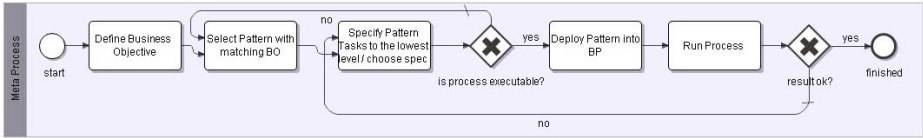


Fig. 3. The meta-process for applying a process pattern to a given business process

steps back to the task of choosing a new pattern. If the pattern is executable, it is deployed into the business process. After that, the integrated process is executed. If the result is satisfying, the meta-process is finished. If it is not satisfying, the meta-process steps back to the task of finding a new specification.

4 Case Study

In [17] a solution for fraud detection based on similarity learning [18] is presented. We choose this scenario as case study, as it already includes well specified business processes as well as a data mining solution that that allows for reuse. The approach for fraud detection from [18] is based on learning similarity measures for data records and is transferable for a generic class of fraud opportunities. The application scenario is based on detecting procurement fraud, e.g. an employee of a company placing an order to another company which is owned by himself. This is done by computing a similarity between employees and company owners based on several features such as name, address or bank accounts. We do not go into more details of the data mining method, as this is not important for understanding our pattern approach. Basically what is needed to apply this data mining solution to a problem is to first check if the problem is a procurement fraud problem, second to specify which features to be used for the similarity, and third to connect the inputs and outputs of the data mining process. For all other steps ready-to-use code is already available.

Fig. 4 shows an example of a pattern for this approach to procurement fraud detection. In the top pool *Requirements* the pre-requirements for applying this pattern are modeled. This includes checking the data mining goal (procurement fraud detection) and the data format as well as sending data, receiving the result and sending labeled data. The other pools *Data Mining Classification*, *Data Mining Model Building* and *Data Mining* contain the (partially already specified executable) tasks of the data mining process and the respective services. It can be seen that the process is a data mining pattern according to our definition from Sec. 3, which contains tasks of all 3 levels of the task hierarchy. E.g., the task *Check DM goal is supervised Fraud Detection* is a specialized task of *Check DM goal*. It is a conceptual task which describes in textual form that the goal of the integration of the data mining solution has to be supervised fraud detection in order to match with the pattern. During the integration, the user manually inspects this task and specializes it to the empty task. The task *Check if each attribute is based on a known data type* represents the requirement for a specific data format. The data mining solution is based on combining different similarity

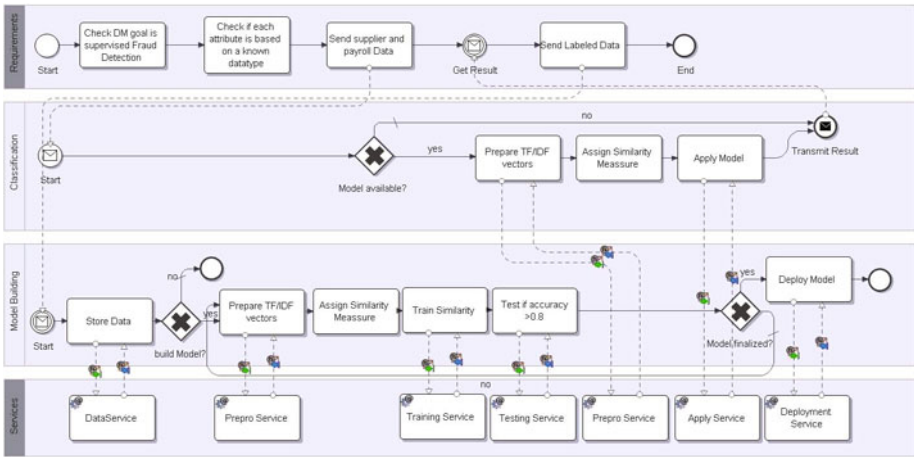


Fig. 4. An example of a data mining pattern

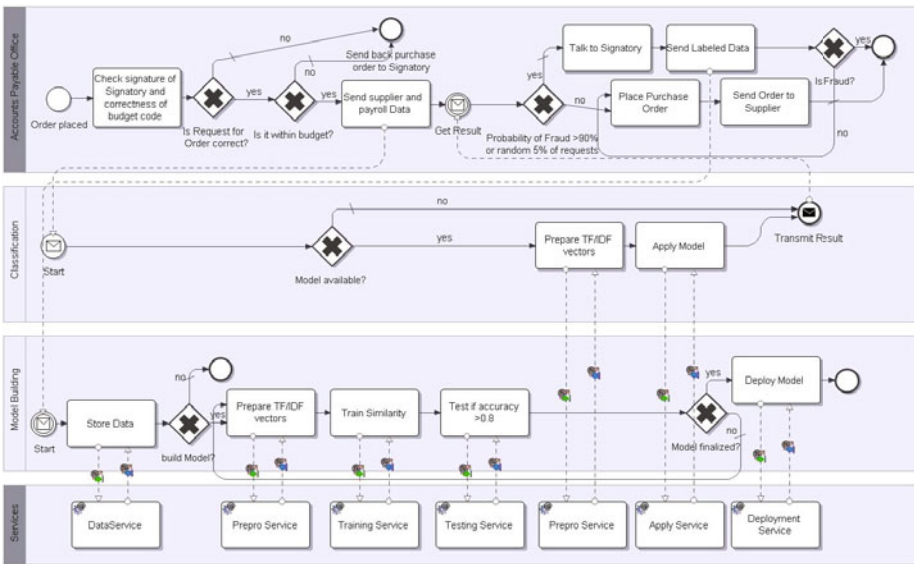


Fig. 5. An example of an integrated process

measures for attributes of different data types. For each of those types, a measure has to exist. The data is sent to the data mining part of the process via the task *Send supplier and payroll data* and the labels via *Send labeled Data*. The assignment of the similarity measures is performed at the task *Assign Similarity Measure*. It is a configurable specialization of the task *Train Model* that can be further specialized by the user, e.g. by a configuration file. All executable tasks, e.g. the task *Train Similarity*, are connected to an underlying service.

In [17] (Section 6) a set of business processes from the health care domain is presented which contains candidates for the integration of the fraud detection solution. We will focus on the business process *Purchase Order Inspection* from the RBH scenario, which consists of a random checking of several rules for a request for an order followed by the decision of placing or not placing this order. Fig. 5 shows how the presented data mining pattern is integrated into the business process. Compared to the pattern, the task *Check DM goal* disappeared because the user took the decision that the data mining goal indeed matched and hence specialized it to an empty task. The tasks for sending data and receiving the results are connected with the tasks of the business process. The tasks *Check if each attribute is based on a known data type* and *Assign Similarity Measure* disappeared as well, as the user manually specified the checking and the assignment of the similarities according to his knowledge of the data (e.g. by a parameter file). During the integration, the pool *Requirements* and its tasks disappeared at all due to the specification and deployment tasks of the meta-process.

By this somewhat simplified example we have shown how to apply a data mining pattern to a business process in order to get an integrated process and how manual steps can be specified to executable tasks.

5 Conclusion and Future Work

This paper is a first step towards support for an automation of a reuse of successful data mining processes. We have presented a new process model for easy reuse and integration of data mining in different business processes. Our approach is based on CRISP and includes the definition of data mining patterns, a definition of a hierarchy of tasks to guide the specialization of abstract patterns to concrete processes, and a meta-process for applying patterns to business processes. These data mining patterns allow for representing the reusable parts of a data mining process at different levels of generalization and provide a simple formal description for the reuse and integration of data mining. We evaluated our approach in a fraud detection case study in the health care domain.

Our approach is focused on reusable cases for data mining problems. Thus, in contrast to the more general CRISP model, the data mining patterns are not applicable in all cases, as several parts of the process are already pre-defined in the pattern. However, by the pre-definition a lot of unnecessary repetitive work is avoided. E.g., if data mining is to be integrated in the context of a very similar problem, the pattern can be just used again instead of applying the full CRISP-DM process twice. CRISP is iterative and does not describe the reuse of existing solutions at all. In our approach, we describe the reusable part by data mining patterns. The iterations are modeled in an outside meta-process.

Details on the definition, architecture and implementation of the underlying services remain to future work. Furthermore, we aim at a formalization that allows for tool support for the design and specification of high-level requirements as well as for the parameterization. In addition, the following questions need

to be addressed in the future: How to model the data within the process? Key challenges are to define the data-related interfaces between the business processes and the data mining patterns as well as to specify how to get the data out of the business process. Is the hierarchy adequate or is a more detailed level needed? E.g., more structure could be provided by the use of an ontology [19]. What is a good pattern, how can the quality of a pattern be determined and how to select a pattern from a pattern database?

References

1. Hornick, M.F., Marcadé, E., Venkayala, S.: *Java Data Mining: Strategy, Standard, and Practice*. Morgan Kaufmann, San Francisco (2006)
2. Wegener, D., Rüping, S.: On Integrating Data Mining into Business Processes. In: Abramowicz, W., Tolksdorf, R. (eds.) *BIS 2010. LNBP*, vol. 47, pp. 183–194. Springer, Heidelberg (2010)
3. Shearer, C.: The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing* 5(4), 13–22 (2000)
4. Rupnik, R., Jaklič, J.: The Deployment of Data Mining into Operational Business Processes. In: Ponce, J., Karahoca, A. (eds.) *Data Mining and Knowledge Discovery in Real Life Applications, I-Tech*, Vienna, Austria (2009)
5. Sharma, S., Osei-Bryson, K.: Framework for formal implementation of the business understanding phase of data mining projects. *Expert Systems with Applications* 36(2) (2009)
6. Marbán, O., Segovia, J., Menasalvas, E., Fernández-Baizán, C.: Toward data mining engineering: A software engineering approach. *Information Systems* 34(1) (2009)
7. Jordan, D., Evdemon, J.: *Web Services Business Process Execution Language Version 2.0*. Technical report, OASIS Standard (2007)
8. White, S.A., Miers, D.: *BPMN Modeling and Reference Guide Understanding and Using BPMN*. Future Strategies Inc., Lighthouse Pt (2008)
9. Bremer, P.: *Erstellung einer Datenbasis von Workflowreihen aus realen Anwendungen*, Diploma Thesis, University of Bonn (2010) (in german)
10. White, S.: *Process Modeling Notations and Workflow Patterns*. In: Fischer, L. (ed.) *The Workflow Handbook 2004*. Future Strategies Inc., Lighthouse Point (2004)
11. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17, 37–54 (1996)
12. Witten, I.H., Frank, E.: *Data Mining: Practical machine learning tools and techniques*, 2nd edn. Morgan Kaufmann, San Francisco (2005)
13. Russell, N., ter Hofstede, A.H.M., van der Aalst, A.H.M., Mulyar, N.: *Workflow Control-Flow Patterns: A Revised View*. BPM Center Report BPM-06-22, BPM-center.org (2006)
14. Atwood, D.: *BPM Process Patterns: Repeatable Design for BPM Process Models*. BPTrends (May 2006)
15. Tsai, C., Tsai, M.: A Dynamic Web Service based Data Mining Process System. In: *Proc. of the Fifth International Conference on Computer and Information Technology CIT*, pp. 1033–1039. IEEE Computer Society, Washington (2005)
16. Altintas, I., Birnbaum, A., Baldrige, K., Sudholt, W., Miller, M.A., Amoreira, C., Potier, Y., Ludscher, B.: A Framework for the Design and Reuse of Grid Workflows. In: Herrero, P., S. Pérez, M., Robles, V. (eds.) *SAG 2004. LNCS*, vol. 3458, pp. 120–133. Springer, Heidelberg (2005)

17. iWebCare Project Deliverable D01 – Business process model of e-gov fraud detection processes in the health care domain (2006), <http://iwebcare.iisa-innov.com/documents/D1-BusinessProcessModelingv4.3.zip>
18. Rüping, S., Punko, N., Günter, B., Grosskreutz, H.: Procurement Fraud Discovery using Similarity Measure Learning. *Transactions on Case-based Reasoning* 1(1), 37–46 (2008)
19. Hilario, M., Kalousis, A., Nguyen, P., Woznica, A.: A Data Mining Ontology for Algorithm Selection and Meta-Learning. In: *Proc. of the ECML/PKDD 2009 Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery (SoKD 2009)*, Bled, Slovenia, pp. 76–87 (2009)