Cross-System Process Mining using RFID Technology

Christian Glaschke, Norbert Gronau and Benedict Bender
Chair of Business Information Systems, especially Systems and Processes, University of Potsdam, August-Bebel-Straße 89, Potsdam, Germany
{cglaschke, ngronau, bbender}@lswi.de

Keywords: Process Mining, RFID, Production, Cross-System.

Abstract: In times of digitalization, the collection and modeling of business processes is still a challenge for companies. The demand for trustworthy process models that reflect the actual execution steps therefore increases. The respective kinds of processes significantly determine both, business process analysis and the conception of future target processes and they are the starting point for any kind of change initiatives. Existing approaches to model as-is processes, like process mining, are exclusively focused on reconstruction. Therefore, transactional protocols and limited data from a single application system are used. Heterogeneous application landscapes and business processes that are executed across multiple application systems, on the contrary, are one of the main challenges in process mining research. Using RFID technology is hence one approach to close the existing gap between different application systems. This paper focuses on methods for data collection from real world objects via RFID technology and possible combinations with application data (process mining) in order to realize a cross system mining approach.

1 INTRODUCTION

Since more than 20 years, business process management is the leading paradigm for organizing and restructuring of corporations and public entities. Although nearly all kinds of companies use business process management in certain areas, there are still some challenges that need further analysis. To name only a few, we can draw the reader's attention to the following issues (Gronau, 2016):

- Improving learning while performing a business process
- Improved usage of person-bound knowledge that is generated in or used during the business process.
- Establishment of PDCA cycles (plan-do-check-act) in process management to be able to detect deviations and to correct them without interrupting the business process.
- Typical business processes nowadays are supported by enterprise systems like ERP, CRM or SCM systems. Nevertheless, there are normally deviations between the intended process covered by the ERP and the process actually performed by the company (Gronau, 2015)
- Irrespective of the high degree of automation, in most business processes, the human-machine interface is more important than ever because automated business processes inevitably have to be interrupted to allow the human being to decide or do something in order to propel the entire process. Today, however, the description of human interfaces is not intuitive at all.
- Last but not least, an increased real world awareness for nowadays available business process objects (like information, persons or customers' material) is needed. Also approaches to integrate these kinds of information into the respective processes are more necessary than ever.

Valid models of business processes are the basis for every optimization approach. Process mining has already shown some merits in some realms and is able to measure the actual performance of a business process - partially with the help of existing process models. Nevertheless, existing process mining approaches are limited in their scope to a single information system. Overcoming this barrier would not only bring the monitoring closer to the actual business process, it could also mean a big step ahead in the direction of process model discovery.

A main obstacle on the road to model improvement is the recognition of two entities with

Glaschke C., Gronau N. and Bender B.
Cross-System Process Mining using RFID Technology.
DOI: 10.5220/0006223501790186
In *Proceedings of the Sixth International Symposium on Business Modeling and Software Design (BMSD 2016)*, pages 179-186
ISBN: 978-989-758-190-8

different IDs in different information systems being one and the same (role, object, instance) in the real business process. For this purpose, i.e. to close this gap, this paper suggests to use RFID (Radio-frequency identification) technology.

2 PROCESS MINING WITH RFID

The focus of the present paper lies on the recognition of processes in heterogeneous application landscapes. In the course of digitalization, a lot of data on a company's business processes have become available. For instance, position vectors of real objects by means of RFID technology (Zhang et al., 2012), cyber-physical systems are able to realize process components autonomously and application systems have to be integrated over several levels in order to meet the requirements. To develop a system in this environment that allows for the reproduction and understanding of the business process is therefore the problem to be solved by the present work. This problem has already been defined and treated in other studies (Van der Aalst et al., 2012, Glaschke, 2015) and it aims at overcoming the system barriers. To achieve this aim, especially process mining is used, as this approach allows for an identification of process patterns through the analysis of a system's log files. The key idea of the present article is therefore to combine process mining with the data from RFID systems to thereby gain a comprehensive view on business processes.

In order to solve this problem, this paper will follow the Design Science Research Methodology (DSRM). According to Peffers et al. (2007) the DSRM involves six steps. Section one and two relate to the first two steps of the DSRM approach, namely the problem identification and definition of objectives. Section three discusses related techniques and technologies that could assist the cross-system based process mining. Section four corresponds to the design and development phase since it specifies the solution for cross-system based process mining. Furthermore, the second subsection demonstrates the solution with a theoretical example which relates to the demonstration phase. The evaluation of the approach needs to be conducted in an experimental setting for which the approach is described in the fifth section of this paper. With regard to the communication step the approach is presented in this paper, while the evaluation results will be published in further papers as well as necessary modifications of the approach.

Finally, the results will be critically discussed and an outlook will be given. The DSRM (Design Science research methodology) seems appropriate since the approach is motivated by theoretical as well as practical issues and involves the design of an appropriate solution for the given problem and it´s evaluation.

3 RELATED WORK

For this paper, two research topics are of utmost importance. First, we will focus on process mining to identify the existing cross application approaches of this method.

The second relevant research topic is the technology of RFID. With respect to this issue, we will particularly focus on works that link these two approaches. Concerning the technology of RFID, we will especially explore the question of how the data is recorded in the application landscape and how it is provided for further processing.

At the end of this section, a brief summary of the existing research gap and the most important corresponding findings will be given.

3.1 Process Mining

A well-known approach for process discovery is the concept of process mining, which was developed by Van der Aalst and his research group at the Technical University of Eindhoven (The Netherlands). This approach uses log files from application systems (for instance ERP systems) to reconstruct process models. To be successful in that effort, the application system has to provide the information needed in a specific manner (i.e. as shown in Table 1).

Table 1: Example of a log file.

PID	Activity	Worker	Timestamp
452	registration	55	2011-12-24, 11:10:21
452	investigation	56	2011-12-24, 11:15:21
452	consulting	33	2011-12-24, 12:17:10
452	dismissal	55	2011-12-24, 12:47:11
453	registration	55	2011-12-24, 11:16:35
453	investigation	56	2011-12-24, 11:27:12
453	consulting	12	2011-12-24, 11:52:37
453	dismissal	55	2011-12-24, 11:59:54
454	registration	55	2011-12-24, 11:11:21
454	investigation	55	2011-12-24, 11:15:21

An important component needed for this listing of process instances is the process identification number (PID). This number is used to create a process diagram based on more than just one process instance (Van der Aalst, 2012). In the background, petri networks are used here (i) to allow for the generation of process diagrams, (ii) to describe the different conditions of the process and (iii) to create a graph for visualization purposes and analysis (Van der Aalst, 2011a; Accorsi, 2012).

In an article from Thiede and Fuerstenau (2016), an extensive literature review on this topic is presented. Therefore, the top 20 AIS Journals between 2004 and 2015 are analysed. Additionally, literature from production and organisational journals have been screened for process mining and RFID implementation publications. It could be concluded that the usage of RFID to close the gap has not been researched so far.

As can be deduced from Thiede and Fuerstenau (2016), many research projects concentrate on the field of "digital services" and among these especially on the subtype of single systems. Thiede and Fuerstenau (2016) define the concept of "service" broadly as "the application of specialized knowledge skills through deeds, processes, and performances for the benefit of customers (Vargo and Lusch, 2004, p. 2). For instance, a customer buying an article in a web shop (digital service) receives it by a dispatching service (non-digital)."

Out of this studies the work on Cross-System Process Mining is analyzed. There are different approaches with wide scattered areas of application. Now we describe the most important ones. The first approach deals with data from smart environments. The main result is an algorithm to mine frequent patterns (Wen, 2015).

Another work deals with the theme of end to end Process Management. The authors describe implications for theory and practice based on the end to end process management approach. Especially the technological challenge is highlighted in the investigation of end to end processes (Madder et al., 2014).

Cross-organizational observations of processes, however, have so far only been presented by very few studies and also these work primarily focused the aspect of workflow synchronization. An example of this specific application of process mining is given by Zeng et al., (2013).

Zeng et al., (2013) define in their work several approaches for the synchronization of different log files:

1. Coordinated with synchronized activities.
2. Coordinated with messages exchanged.
3. Coordinated with shared resources.
4. Coordinated with abstract procedures.

These four coordination patterns can generally be used to link the different log files from various companies. Nevertheless, this solution proposal is based on the assumption that the respective processes remain closed systems only linked to each other at some specific points. As examples hereof, we can cite the cases of goods delivery (activity) or document exchange (messages). Based on these patterns, rules can be defined that determine when exactly a synchronization takes place in two log files or models.

3.2 RFID

Today and due to digitalization, data on numerous business objects is available. So, for instance, the position vectors of an object can nowadays be calculated in real-time due to RFID technologies (Yingfeng et al., 2012). These data can then be used to analyze the goods flow like demonstrated, for example, in Jakkhupan (2012). For this purpose, the objects involved in the process are first tracked and then the respective data is saved.

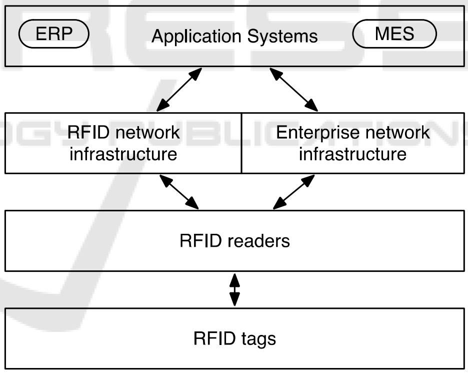

Figure 1: RFID infrastructure (Krishna, 2007).

In a typical RFID infrastructure as shown in Figure 1, the readers send unprocessed data to an application system. In other architectures, middleware has to be used for this purpose (Abad et al., 2012). Furthermore, cross system communication is made possible by writing collected information on the RFID tag to be read out and further processed at other points in the process (Finkenzeller, 2015).

A combination of both technologies, i.e. of RFID and process mining, has already been used by Zhou (2010) for the reconstruction of health care processes. For this approach, the techniques of process mining were applied to both, log files and the RFID infrastructure. In doing so, it could be evidenced that the log files of RFID infrastructures are also suitable for being used in the context of process mining applications and techniques (Gonzalez, 2006).

Fernandez et al., (2015) use a similar approach working with indoor location systems to collect data on different business objects in order to draw conclusions concerning the corresponding processes. For this second step, the authors also work with process mining techniques.

3.3 Evaluation

Particularly from a reading of the work of Zeng et al., (2013), the question arises of how the present paper will distinguish from previous research. Synchronization of workflow models has already been described extensively in literature. Hence, this article primarily differs, on the one hand, with respect to the investigated area, as here company internal processes are investigated. On the other hand, in our approach, the different workflow models from the single application systems shall be integrated and not just synchronized. Thus, we aim at establishing an end-to-end business process that is interconnected across the different system boundaries. To achieve this goal, it is for example suitable to work with the shared resources approach (Chapter 3.2), with the respective data being tracked – in our case – by RFID systems. In choosing this solution, it is not necessary that both systems use the same denomination or ID for one and the same object (Chapter 4).

4 CROSS SYSTEM PROCESS MINING

In order to achieve the target to extract reliable and encompassing models of process execution, it is nowadays necessary to link information from multiple information systems. This is due to the fact that business processes involve different functional areas within a company which are usually using highly specialized software applications in order to fulfill their tasks. As a result hereof, companies' application landscapes involve many different application systems, each executing a little part of the overarching process (Fuerstenau and Glaschke, 2015).

4.1 Requirements and Objectives of the Procedure

Established techniques like process mining are conceptualized to evaluate information from single information systems. Typically, for this purpose, log files from one information system are analyzed. In this procedure, firstly log entries are grouped according to their affiliation (e.g. by assorting all log entries that belong to a certain PID), secondly dependencies and chronological consecutions are analyzed and thirdly and finally all the aforementioned bits of information are aggregated in a workflow model.

In order to face now the challenge to model complex cross-system processes, the concept of process mining shall be complemented by a linking technology in order to be able to model complete business processes that are not limited to a single information system. This linking technique shall allow for a combination of the different sub-process models that could be extracted from the different information systems participating in the overarching process.

One essential precondition of this approach is that each information system that forms a part of the overall process is connected with the RFID infrastructure. This is necessary for the PID from the information system to be passed to the RFID infrastructure.

Another necessary prerequisite for this solution is that the PIDs from the different information systems are unique, i.e. that they are working with different number ranges. Only in this way, each process instance can be identified unequivocally.

Besides, also the given RFID tag must identify the objects unequivocally by ignoring the sharing or merging of RFID tag numbers.

Thus, our aim is to facilitate the integration of various log files and the therein contained process instances as illustrated in Figure 2.

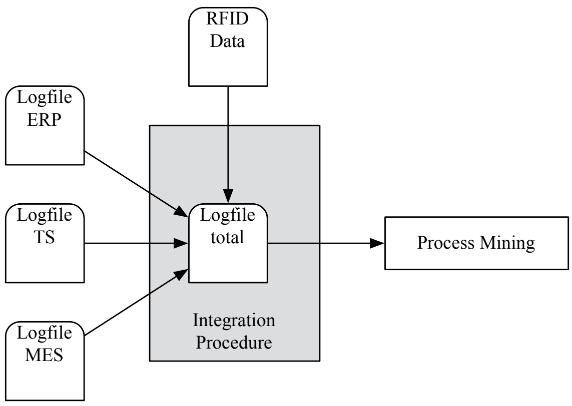

Figure 2: Fundamental procedure for log file integration.

The combination of information system log files with RFID data should reveal the overarching process model which is composed of single system specific process models that are combined in the appropriate order. One of the major challenges in this context is to identify corresponding sub-processes in the different information systems. In order to face this challenge, a distinct linkage between the different sub-process models is needed.

Just like in case of process mining, configurations have to be made within the information systems for the created log files to provide an uninterrupted description of the process instances. To achieve this, each generated instance has to pass the current process ID to the RFID infrastructure at known interconnection points. This PID is then saved on the respective RFID tag. To make this possible, scanning processes must be performed at specific process steps. Ideally, these scans will take place at the beginning and end of a process.

In order to realize the linkage between the different information systems, two kinds of physical components are needed. The first component are containers that physically accompany the different sub-process steps through the entire process. This kind of containers is set up in the first process, which is executed in the ERP system. At this point, an empty container is equipped with an RFID-tag that has a unique ID.

The second component is the aforementioned infrastructure of RFID gates, which is used to scan the containers accompanying the different process steps.

Figure 3 illustrates how the individual scanning processes are registered as entries on the RFID tag. To allow for this kind of documentation, the involved application systems have to be customized accordingly. This customization procedure is already standard in case of process mining applications (Van der Aalst, 2011b). This technique, however, additionally requires that when scanning the RFID tags, the data is passed and stored. This also requires the appropriate configuration of the participating application systems.

Yet, this solution proposal is just one variant. Another method would consist in passing the unequivocal RFID tag ID to the information systems to be saved there. In this case, the tag ID would have to be written to the logs of the application systems. Another possible approach would be to log the identity within the RFID architecture. As here the respective functionality needed to log the scan processes is given, this variant will be favored in the context of our validation.

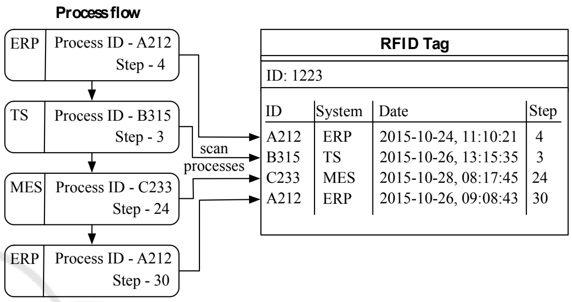

Figure 3: Data structure on the RFID tag after scanning.

4.2 Approach for an Integration Procedure

Although the here presented solution proposals seem to be convincing, there is a significant gap between the processing through information technology and reality. Whereas in an information system tasks can already be performed without the necessity of a concrete processing object with RFID tag to exist, in reality and as shown in figure 4, there will always be overlapping time entries in the log files.

As strategy for our integration procedure, we therefore have to take the following steps: The process steps will have to be passed to the result log following their chronological order and in doing so,

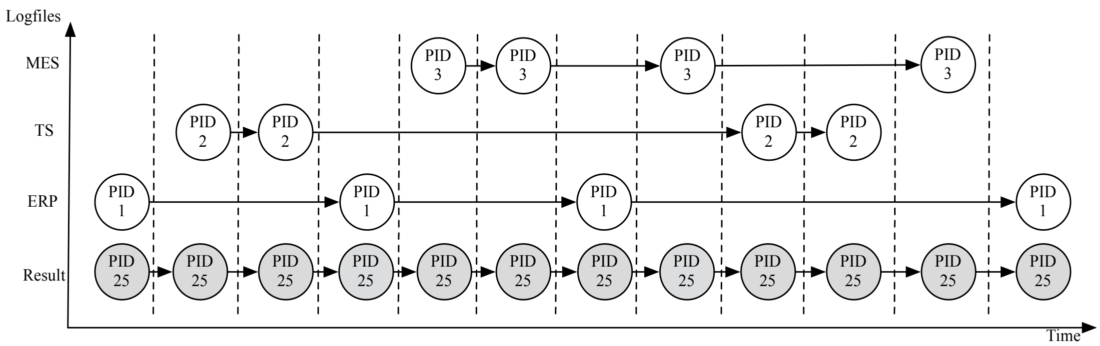

Figure 4: Basic procedure for log file integration (for one process instance).

the PIDs from the application systems will have to be replaced by the PID on the RFID tag. For this purpose, the algorithm goes through the entries on the RFID tag, systematically replacing the PIDs in the log files, to afterwards copy the lines into one new contiguous result log file.

The process mining procedure which further processes the result log file then uses the corresponding algorithms in order to reconstruct the workflow with the aid of the time entries. Associated examples and algorithms for this procedure are presented by de Medeiros et al. (2003). For the validation of this approach, an appropriate data model is needed which is illustrated in Figure 5. Each RFID reader is assigned to one place and one application system. The RFID readers generate log entries, which, in turn, are assigned to just one business object. In our case, it is the read data and the written data that is logged. Then, the software passes the reading and writing processes to the assigned application system.

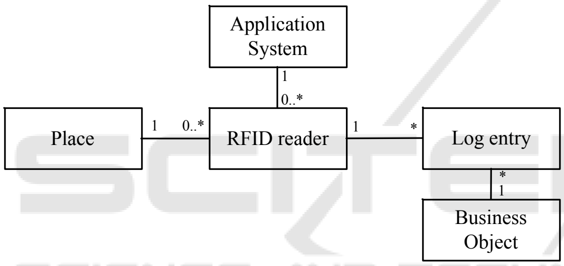

Figure 5: Data Model of RFID Management Software.

Furthermore, we developed a process to be used for validation that can also serve as reference for the obtained result models. This process was designed to reflect the typical production process of a manufacturing company and it comprises the following phases:

To begin with, one order with several items results in different manufacturing jobs in the ERP system. Next, the corresponding components are commissioned and the production container is packed. Afterwards, the order is scanned for the first time and is passed on to the transport system (TS), which moves the container to the first free workstation. Subsequently, several production steps are carried out in MES (Manufacturing Executin System). At different process steps, the container is scanned for identification purposes. In order to measure the product properties, the data is handed over to a quality management system and the measurement results are documented. After this step, the container is transported to the warehouse or distribution center.

Even though the approach seems to solve a range of problems with regard to the cross-system process mining, it has to be remarked that this type of aggregation cannot be used for all types of processes. The applicability is limited to processes where the physical parts, that are subsumed in the container, are scanned at all of the involved information systems or the data is passed, at least once, to each of the participating information systems of the process. This is due to the fact that the combination technique relies on a briding mechnism based on part handover.

5 VALIDATION

In order to validate the concept for cross system process modeling, several steps are necessary, which need to be executed subsequently.

As a first step, an appropriate model composition procedure needs to be realized in order to integrate the different log files into one overarching log file. At this point, and similar to the proceeding in process model generation, the distinction of relevant and irrelevant information determines the quality of the composed overarching process model. This composition process is supported by the information from the RFID management software and, in particular, by the unique IDs that are allocated by the software component.

As a second step with regard to validation, the experiment itself needs to be set up. Therefore, a production setting is needed which allows for the representation of the formerly defined process. This stage involves physical components, such as some kind of production machines, as well as the related software infrastructure, which is composed of the different systems mentioned in the process. Finally, the RFID management software as well as the RFID infrastructure including the containers have to be set up.

For the validation of our approach, a university Industry 4.0 Laboratory is used. This laboratory is equipped with several application systems like, among others, an ERP and a MES system from the company Asseco Solutions AG. Additionally, different infrastructure components like robots, transportation systems and a RFID infrastructure are available. The manufacturing process can thus be simulated through configuration of the components or it can be actually performed (figure 6).

The RFID infrastructure consists of stationary and portable RFID readers with which RFID tag information can be read and written. For the simple configuration of the RFID infrastructure, we have a

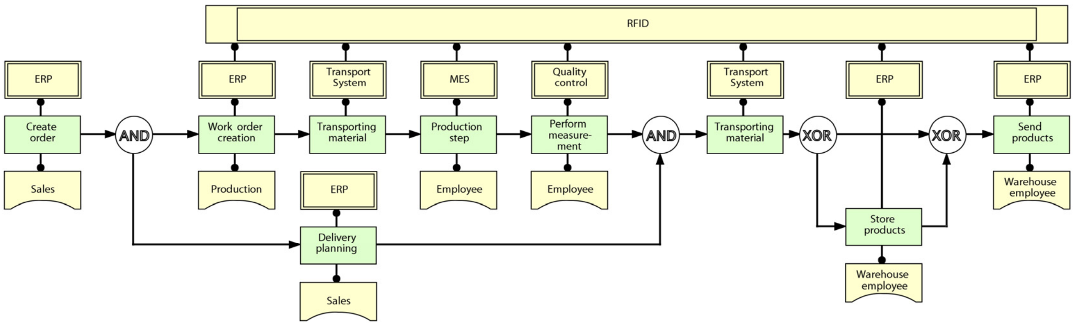

Figure 7: Example process with RFID support.

management software that is used as middleware in the laboratory plant.

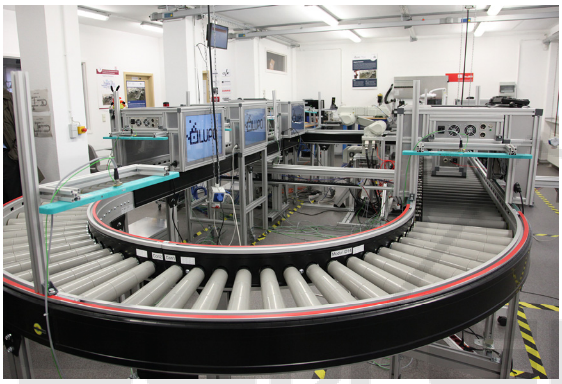

Figure 6: Physical validation environment.

Besides, delivery planning takes place while the production is in progress. Based on the thereby elaborated scheduling list, the TS decides where to transport the products (warehouse vs. distribution center). These steps are also logged and get assigned the PID from the ERP (Enterprise Ressource Planning) system.

For validation the process should be executed several times with different variations. Especially the different mining algorithms from the process mining software (Claes and Poels, 2012) will here have to be tested and the respective results will have to be compared. The validation will be successful when the elaborated models actually reflect the process shown in Figure 7.

6 CONCLUSIONS AND OUTLOOK

In this paper, it could be demonstrated that a simple and accessible method can close the gap between the different systems and processes. For this purpose, the concept of shared resources (Zeng et al., 2013) was applied, which is already used in cross-organizational approaches. Furthermore, it could be shown that the RFID technology does already provide a solution concept. Thus, we have succeeded in presenting a cross system integration procedure.

If the results are achieved within the laboratory environment, the question arises of whether the approach will also function in practice and in more complex environments with massive parallel operations. All in all, the question remains of what factors are responsible for the quality of results. Here, on the one hand, the quality of the log files from the application systems will definitely have to be taken into consideration and, on the other hand, the number of iterations needed to identify parallel and independent process steps will be of importance. The validation results will be published in further works.

In this paper, a simple setting was chosen to demonstrate the benefits of the technology combination of process mining and RFID. Future work must continue to follow this trend. It is possible to track more than one and different business objects. The application system provides further information, which can be used for the enrichment of process model analysis. An example - in this case products were anonymous, but they can be categorized in future according to the product category. So it is possible to structure the mined process model.

To follow the idea of end to end processes we have to consider the supplier processes and other support processes like maintenance. Therefore laboratory environments with real applications systems and physical components can help, to establish the techniques in practice by showing practical solutions and related benefits.

REFERENCES

Abad, I., Cerrada, C., Cerrada, J. A., Heradio, R., Valero, E., 2012. *Managing RFID sensors networks with a general purpose RFID middleware.* Sensors, 12(6), 7719–7737.

Accorsi R., Stocker T., 2012. *On the Exploitation of Process Mining for Security Audits: The Conformance Checking Case*. ACM Symposium on Applied Computing. doi:10.1145/2245276.2232051.

Claes, J., Poels, G., 2012. *Process mining and the ProM framework: An exploratory survey*. Business Process Management Workshops. Springer Berlin Heidelberg.

De Medeiros, A. A. K., Van Dongen, B F, Van der Aalst, WMP., Weijters, A.J.M.M., 2003. *Process mining: extending the α-algorithm to mine short loops*. On the move to meaningful internet systems, Springer Berlin Heidelberg, 389-406.

Fernandez-Latas, C., Lizondo, A., Monton, E., Benedi, J. M., Traver, V., 2015. *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems.* Sensors, 15(12), 29821-29840.

Finkenzeller, K., 2015. *RFID-Handbuch: RFID Handbook: Fundamentals and practical applications of transponders, contactless smart cards and NFC. (in german)* Carl Hanser Verlag, Volume 7.

Fuerstenau, D., Glaschke, C., 2015. *Weighting of Integration Qualities in IS Architectures: A Production Case.* Paper presented at the ECIS 2015, Germany.

Glaschke, C., Gronau, N., 2015. *Automated Process Model Discovery - Limitations and Challenges*. Paper presented at the International Symposium on Business Modeling and Software Design.

Gonzalez, H., Han, J., Li, X., 2006. *Mining Compressed Commodity Workflows from Massive RFID Data Sets.* 15th ACM International Conference on Information and Knowledge Management, ACM, 162-171.

Gronau, N., 2016. *Business Process Management in industry and administration.* (in german), Gito Verlag, Berlin, 2016.

Gronau, N., 2015. *Trends and Future Research in Enterprise Systems.* Lecture Notes in Business Information Processing, Volume 198, 271-280.

Jakkhupan, W., Archint, S., Li, Y., 2012. *Business process analysis and simulation for the RFID and EPCglobal Network enabled supply chain: A proof-of-concept approach.* Network and Computer Applications, 34(3), 949-957.

Krishna, P., Husalc, D., 2007. *RFID Infrastructure*. IEEE Communications Magazine, 45(9), 4-10

Maddern, H., Smart, P. A., Maull, R. S., Childe, S., 2014. *End-to-End Process Management: Implications for Theory and Practice.* Production Planning & Control, 25(16), 1303-21.

Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S., 2007. *A Design Science Research Methodology for Information Systems Research*. Journal of Management Information Systems. 24(3), 45-77.

Thiede, M., Fuerstenau, D., 2016. *The Technological Maturity of Process Mining: An Exploration of the Status Quo in Top IS Journals.* Paper presented at the MKWI 2016, Germany.

Van der Aalst, WMP., Accorsi R., Ullrich M., 2012. *Process Mining*. Retrieved. (in german) April 29, 2015, from http://www.gi.de/nc/service/informatiklexikon /detailansicht/article/process-mining.htm

Van der Aalst, W., 2011a. *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer.

Van der Aalst, W., 2011b. *Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior.* IEEE Transaction on Service Computing, 6(4), 525-535.

Vargo, S., Lusch, R., 2004. *Evolving to a New Dominant Logic for Marketing*. Journal of Marketing, 68(1), 1-17.

Wen, J., Mingyang Z., Zhiying W., 2015. *Activity Recognition with Weighted Frequent Patterns Mining in Smart Environments.* Expert Systems with Applications, 42, 6423-32.

Yingfeng, Z., Jiang, P., Huang, G. Qu, T., Zhou, G., Hong, J., 2012. *RFID-Enabled Real-Time Manufacturing Information Tracking Infrastructure for Extended Enterprises.* Journal of Intelligent Manufacturing, 23, 2357-2366.

Zeng, Q., Sun, S. X., Duan, H., Liu, C., Wang, H., 2013. *Cross-Organizational Collaborative Workflow Mining from a Multi-Source Log.* Decision Support Systems 54, 1280-301.

Zhou, W., Piramuthu, S., 2010. *Framework, strategy and evaluation of health care processes with RFID.* Decision Support Systems, 50(1), 222-233.