



# An automatized data extraction approach for process mining and business process analysis

Sina Namaki Araghi, Franck Fontanili, Elyes Lamine, Frederick Benaben

## ► To cite this version:

Sina Namaki Araghi, Franck Fontanili, Elyes Lamine, Frederick Benaben. An automatized data extraction approach for process mining and business process analysis. IFAC 2017 - 20th World congress of the International Federation of Automatic Control, IFAC, Jul 2017, Toulouse, France. p.9580-9584. hal-01924018

**HAL Id: hal-01924018**

**<https://imt-mines-albi.hal.science/hal-01924018>**

Submitted on 15 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An automatized Data Extraction approach for Process Mining and Business Process Analysis

Sina Namaki Araghi . Franck Fontanili . Elyes Lamine . Frederick Benaben

*\*Industrial Engineering Center, University of Toulouse-Mines d'Albi, Albi, France, (firstname.surname@mines-albi.fr)*

---

**Abstract:** In this paper we explain an abstracted version of our approach to provide a rapid business process modelling, and diagnosis for operational processes within organizations. This approach is oriented toward designing a set of tools which has three main functions; Tracking, Modelling, and Assessing. As a result, first we would gather the events automatically (Tracking), thanks to an Indoor Positioning Systems. Second, by using Process Mining we would be able to get the business process models (Modelling). Third, by using case-based, or heuristics algorithms and Discrete Event Simulation we want to propose a tool for process diagnosis and improvements (Assessing). This research project targeted three main fields which are Healthcare, Supply Chain Management, and Sport. This project is subjected as a doctoral thesis in a collaboration between Industrial Engineering Center of Ecole des Mines d'Albi-Carmaux and Maple High Tech in France.

**Keywords:** Process Mining, Discrete Event Simulation, Real Time Location System, Business Process Management

---

## 1. INTRODUCTION

Today's organizations in numerous areas either in service domain, or industry are competing in a dynamic business world, where each day customers can receive a new product or service with different and more improved characteristics. Thus an organization is obliged to provide a continuous improvement environment within its processes to be able to compete at this level. This goal is accessible by using a Process Aware Information System (PAIS) (Dustdar, 2006). Acquiring a realistic and relevant model of the real world processes is an essential prerequisite in any continuous improvement approach. Nowadays to reach that competence, it is usually useful to proceed with data extraction from multiple data bases in the information system. But current approaches rarely gathers all the data, which could be useful to build a process model. Some activities, particularly those, which are not automated or instrumented, such as manual tasks, are not always tracked or recorded. For instance, hospitals information systems do not gather all the data used to model the different pathways followed by patients. It is the same problem in a manufacturing workshop, or on a larger scale in a supply chain, where the different steps of operation, delay, transfer, and inspection of products are not systematically recorded. Therefore, when the data for modelling are not sufficient or even non-existent, it is a necessity to extract the information either by using Interview-Based, or Evidence-Based discovery. As it already had been discussed in previous research works, these two ways have various shortcomings in case of gathering the exact and clean information (*Fundamentals of Business Process Management*

/ Marlon Dumas / Springer, 2013). Equally important if these observations are collected, they could be misinterpreted due to the limitation of a sample which does not ensure a true representation of the entire population. Consequently, the modelling could be inconsistent compared to the real processes, followed by a risk of a wrong diagnosis and unsuitable improvement solutions. This is the first challenge regarding this research work. The next one is choosing a proper modelling language, which could represent the processes in relevance with their field. On the other hand, projects about the approach to transform a real process to a business model as accurately as possible are more unusual. The third challenge in this research work is providing a Business Process Analysis (BPA). This will focus on obtaining diagnosis and relevant key performance indicators. Regarding these challenges, we aim at developing a process mining tool, which would be able to extract data automatically from actual situations, generate business process models, and propose improvement alternatives. This tool will focus on operational processes within organization. In this order, we could provide a better organizational perspective for the enterprises, to understand their real time situation.

This paper is organized through four main sections. The first section aims at explaining the framework of a project, which is oriented toward a PhD thesis ("A Tool-Based approach for a Fast Business Process Modelling and Diagnosis"). In the next section, the "detailed proposal" is presented. It concerns the approach to overcome the described challenges. In the "state of the art" section, some results from previous works

are provided. Finally, “conclusion” describes the possible outcome of this project.

## 2. DETAILED PROPOSAL

This tool based approach aims at defining elements to automate, accelerate, and improve the Business Activity Monitoring (BAM) and Business Process Analysis (BPA). The extracted processes must be an applicable and dynamic model of reality. In figure 1, you can find an IDEF 0 model which shows the targeted approach in this paper. This approach is based on three main functions labelled as Tracking, Modelling, and Assessing. These three functions are naturally linked, and will be run in an integrated way by using the relevance tools. In this order we will be able to generate a solution for each described challenge.

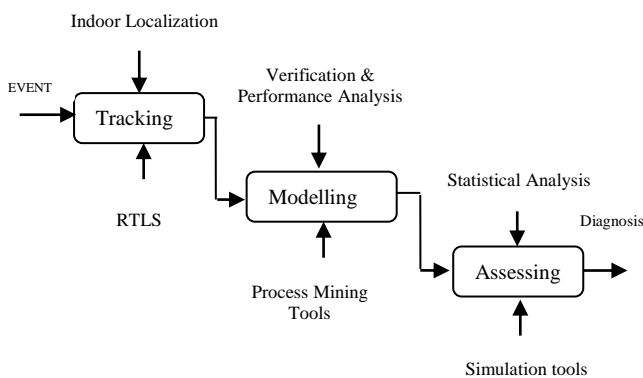


Fig. 1. The IDEF 0 representation for each step of the research work

### 2.1 Tracking

The three targeted areas (healthcare, supply chain, sport) are oriented toward multiple processes, which contain numerous entities and activities. Acquiring a dynamic awareness of their elements is necessary, to improve the outcome of these processes. RTLS (Real Time Location System) is a system dedicated to support the Tracking function and could be an answer on automatized data extraction. RTLS is an adapted technology for indoor localization. Its main function is similar to GPS (Global Positioning System) for locating a mobile object (People, or any other Physical Objects), which wears a radiofrequency tag. However, RTLS uses indoor sensors for tracking objects, while GPS uses satellites. In addition, RTLS uses different algorithms based on signal triangulation. While using this technology we will be able to overcome the first obstacle, which consists in adapting and testing this technology to automatically generate a clean and detailed log file. This log file could consist of all the necessary entities for us to trace and model the processes. Figure 2 displays our vision for the two primary functions (Tracking and Modelling). This figure shows different activities regarding extraction of data from multiple data bases and indoor localization sensors, and finally generating the model. Besides automated tracking of physical flow, one of the biggest challenges in this work would be to enhance

the robustness of the gathered log files. It is necessary to infer multiple rules to control the process, such as routing or prioritization rules, dynamic assignment of staff, use of extra resources, etc. Also there would be a need to merge different types of data to be able to track the elaborated processes. Consequently, we need to use various types of Data Mining algorithm to have a proper log file, which displays the reality.

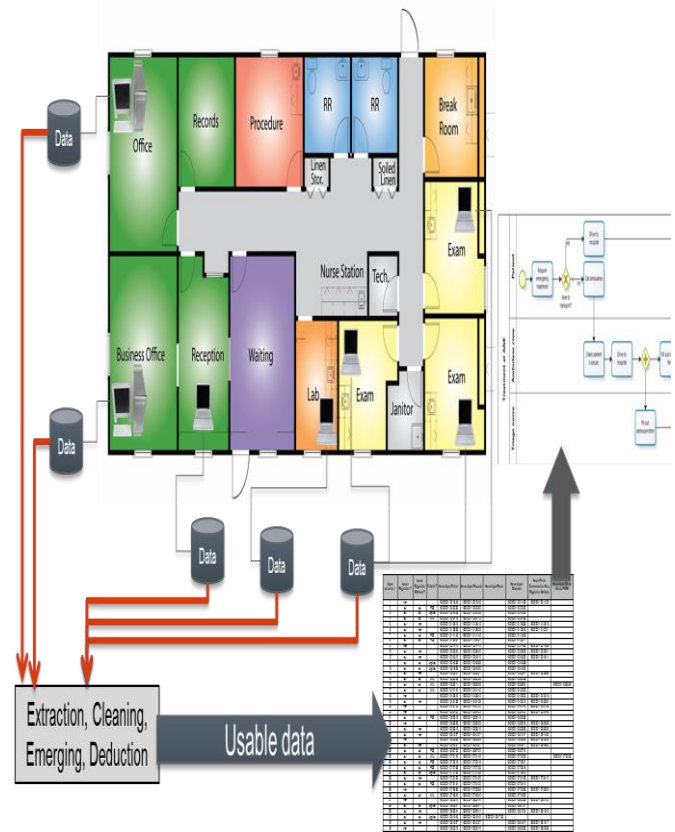


Fig. 2. An illustration for the two primary functions of the approach (Data extraction, Modelling).

### 2.2 Modelling

At the second phase of this research work we aim at extracting information from the event logs. This idea is based on discovering activities while using a Process Mining tool. As a proof of concept, for beginning of this work we will concentrate on tools which have already made their ways into the process mining market, such as ProM<sup>1</sup> and Disco<sup>2</sup>. To improve the execution of processes, we would like to couple Business Process Management and Lean Management. The customer orientation of Lean is similar to that of BPM (4). These two approaches would aid us to improve the satisfaction level of users, and to enhance the process flow and agility of the organization. The prospected modelling language for this approach would be BPMN 2.0 (Business Process Model and Notation) which has many competencies regarding our objectives. BPMN is widely used, because one

<sup>1</sup> [www.processmining.org](http://www.processmining.org)

<sup>2</sup> <https://fluxicon.com>

could show the flat control flow perspective, sub processes, data flows, and resources within one BPMN diagram. Moreover, it allows the users to build a conventional and comprehensive process models (Kalenkova et al., 2015).

### 2.3 Assessing

It's been cited that one of the main shortcomings in process mining tools is disability to provide compatible diagnosis and improvement alternatives for process models (Rojas et al., 2016). Process Mining focuses on extracting knowledge from data generated and stored in corporate information systems in order to analyse executed processes. In the healthcare domain, process mining has been used in different case studies, with promising results. Accordingly, we have conducted a literature review of the usage of process mining in healthcare. The scope of this review covers 74 papers with associated case studies, all of which were analysed according to eleven main aspects, including: process and data types; frequently posed questions; process mining techniques, perspectives and tools; methodologies; implementation and analysis strategies; geographical analysis; and medical fields. The most commonly used categories and emerging topics have been identified, as well as future trends, such as enhancing Hospital Information Systems to become process-aware. This review can: (i) provide a useful overview of the current work being undertaken in this field; (ii) help researchers to choose process mining algorithms, techniques, tools, methodologies and approaches for their own applications; and (iii) highlight the use of process mining to improve healthcare processes (8). Thus, strengthen this point in business intelligence tools is a priority in this research work. For instance in healthcare processes, we would like to propose different algorithms in order to discover the bottlenecks, thanks to the key performance indicators, such as longest activity, critical path, distance that each patient takes, etc. So the first step of our work concerns the study of existing algorithms and results, which they could provide. Finally we would like to present our new algorithms to generate diagnosis for each process. We also aim at using a discrete event simulation approach. Consequently, the user will be able to execute the process and see its weaknesses before implementing the process in real world.

### 3. Applicability

This work would like to target three different areas. First, it would be healthcare. With the importance of efficiency in this area many healthcare organizations are under major pressure to reduce their costs and enhance their services. On the other hand, these organizations are troubled with vast amount of data, caused by the improvement in using information systems to record all the internal events. Previously these data were gathered by using interview-based discovery. Now we would like to use indoor localisation to generate the log files, and step by step (based on figure 1) try to provide insight for each process. As a result, this approach would help healthcare organizations to enhance their information system to become process-aware (Rojas et al., 2016). Eventually, they could reduce their costs, simplify the elaborated processes, find the bottlenecks, and visualize the

process paths. Thus, it could improve the patient safety and satisfaction level.

The second area is Supply Chain Management. Nowadays, this subject is attached to data mining and process mining. The difference between supply chain and healthcare is that in healthcare we trace the map of one object which is a patient, however in supply chain we are gathering data about multiple objects which are either assembled or disassembled in the chain among different actors. So, this high frequency of changes and fluctuations, caused by vast number of elements in logistic processes could be an evidence for us to target this area (Becker et al., 2017).

The third area would be Team Sports. This area is oriented toward enormous amount of collected data for each player during games, or in training sessions. Many sports as basketball, soccer, volleyball, and American football feel the need to have a perspective about their strategies at each action. It could be very useful to have a tool which in correspondent with indoor localization system provides insights for team staffs about the outcomes of their plays, such as the distance that has been covered by each player, efficiency in the motion of players, and blind spots of each position.

Eventually, figure 3 presents our goal in this research work, which is to apply a Business Process Intelligence (BPI) approach to gain improved processes from raw event logs. Based on this approach the tracking step and preparation of log files would transfer the raw information into insight for organization. Then, by creating process models and choosing relevant KPI's, one could obtain a proper diagnosis. Finally, one could evaluate the performance of processes thanks to KPI's, and by using simulation we can run several scenarios in order to find the best way of executing processes. We can consider process mining as a tool which represents the BPI (van der Aalst and Weijters, 2004).



Fig. 3. The Business Intelligence approach

## 4. STATE OF THE ART

In this section we would like to provide some results from the previous works in this area by answering these questions:

1. What is the advantage of using indoor localization for extracting data?

2. Why is necessary to use a Business Process Management approach?
3. How Process Mining could help?
4. What are the advantages in using Discrete Event Simulation?

(i) To answer the first question needless to say that the very first step to execute a process mining tool is to use a log file. It has been cited (Suriadi et al., 2017) in many use cases the major problem was to gather detailed log file which corresponds to the reality. Therefore, using indoor positioning could be an applicable solution for our problem. This way we could get the data automatically with exact level of details that we wanted (Miclo et al., 2015).

(ii) Business Process Management is oriented toward operational processes. It aims at extracting information from raw data and providing a map of the operational processes which could help us to gain a perspective of our organizations. Moreover, based on some citations (Aalst and P, 2013) using BPM and Process Mining could help us in discovering, automatizing, monitoring, and analysing the processes. On the other hand, BPM could be defined as an applicable approach while using simulation. Eventually these could be the reasons for us to consider BPM as a helpful approach in our research work.

(iii) The enterprises are targeting the idea of correlating the strategical management level (Decision) and support management level by executing the operational processes in an efficient way. This could be feasible by having a better vision on these processes. The very first step could be extracting data from information systems and event logs to capture the operational business processes as they are being executed (van der Aalst and Weijters, 2004).

(iv) In our review on previous works on process mining tools (Aalst and P, 2013), it has been noted that many tools are considering the discovering, and conformance checking activities, but the last section which is enhancement, has been neglected. To solve this problem, we propose to use Discrete Event Simulation. As a result we would be able to get an insight about the outcome of each process. Albeit, it has been mentioned because of the lack of experiences in process simulation area, we don't see many usage of these tools. On the other hand, enterprises believe this kind of tools simplify the reality and they are not applicable. But in contrary, providing compatible diagnosis and a perspective of reality for each process could be pragmatic.

To summarize, using Indoor Localization System and Process Mining could help service, entertainment, or industry organizations to visualize the processes, bottlenecks, and interactions among all the entities within processes (such as medical staff, tools, and patient) (Fernandez-Llatas et al., 2015). However, based on our observation, it is rare to find a tool which could complete the Business Intelligence approach (figure 3). As a result, the enhancement of processes still is time consuming and challenging for organizations.

## 5. CONCLUSION

The presented paper subjected to describe the main characteristics of a PhD thesis which aims at defining an approach to identify components to automate, accelerate, and improve the design of virtual processes, which must be a virtual and dynamic copy of the real process. Thanks to this research work, improvement solution will be assessed before being implemented on the real system. It will also provide interesting opportunities in terms of agile and predictive control in the near future. This topic is close to ongoing research work on Industry 4.0 (Rennung et al., 2016).

## ACKNOWLEDGEMENT

This research work is supported by Maple High Tech Company. We are thankful to our colleagues specially Mr. Ludovic Tancerel who is supporting us by his great knowledge on Indoor Localizations Systems.

## REFERENCES

1. Becker, T., Lütjen, M., Porzel, R., 2017. Process Maintenance of Heterogeneous Logistic Systems—A Process Mining Approach, in: Freitag, M., Kotzab, H., Pannek, J. (Eds.), *Dynamics in Logistics, Lecture Notes in Logistics*. Springer International Publishing, pp. 77–86.
2. Dustdar, S., 2006. Marlon Dumas, Wil M. van der Aalst, Arthur H. ter Hofstede, *Process-Aware Information Systems: Bridging People and Software Through Process Technology*, September 2005, Hardcover, 432 pp., ISBN: 0-471-66306-9, US \$99.95. *Inf. Process. Manag.* 42, 1396–1397. doi:10.1016/j.ipm.2006.01.005
3. Fernandez-Llatas, 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, 29821–29840. doi:10.3390/s151229769
4. *Fundamentals of Business Process Management | Marlon Dumas | Springer*, 2013
5. Kalenkova, A.A., van der Aalst, W.M.P., Lomazova, I.A., Rubin, V.A., 2015. Process mining using BPMN: relating event logs and process models. *Softw. Syst. Model.* doi:10.1007/s10270-015-0502-0
6. Miclo, R., Fontanili, F., Marquès, G., Bomert, P., Lauras, M., 2015. RTLS-based Process Mining: Towards an automatic process diagnosis in healthcare, in: *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. Presented at the 2015 IEEE International Conference on Automation Science and Engineering (CASE), pp. 1397–1402. doi:10.1109/CoASE.2015.7294294

7. Rennung, F., Luminosu, C.T., Draghici, A., 2016. Service Provision in the Framework of Industry 4.0. *Procedia - Soc. Behav. Sci.*, 13th International Symposium in Management: Management During and After the Economic Crisis 221, 372–377. doi:10.1016/j.sbspro.2016.05.127
8. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D., 2016. Process mining in healthcare: A literature review. *J. Biomed. Inform.* 61, 224–236. doi:10.1016/j.jbi.2016.04.007
9. Suriadi, S., Andrews, R., ter Hofstede, A.H.M., Wynn, M.T., 2017. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Inf. Syst.* 64, 132–150. doi:10.1016/j.is.2016.07.011
10. van der Aalst, W.M.P., Weijters, A.J.M.M., 2004. Process mining: a research agenda. *Comput. Ind., Process / Workflow Mining* 53, 231–244. doi:10.1016/j.compind.2003.10.001