

DSc. PhD. Edyta Brzychczy

 AGH University of Science and Technology, **Poland**

Abstract: Process modelling is a very important stage in a Business Process Management cycle enabling process analysis and its redesign. Many sources of information for process modelling purposes exist. It may be an analysis of documentation related directly or indirectly to the process being analysed, observations or participation in the process. Nowadays, for this purpose, it is increasingly proposed to use the event logs from organization's IT systems. Event logs could be analysed with process mining techniques to create process models expressed by various notations (i.e. Petri Nets, BPMN, EPC). Process mining enables also conformance checking and enhancement analysis of the processes. In the paper issues related to process modelling and process mining are briefly discussed. A case study, an example of delivery process modelling with process mining technique is presented.

Keywords: process modelling, process mining, event logs, Petri net, model discovery

1. INTRODUCTION

Business process management requires a deep knowledge about real realisation of processes in an organization. This knowledge enables understanding and analysis of business processes aiming in their improvements.

There are various possibilities to express knowledge about a process, one of them is to build a process model with a specific detail level based on available data sources. In the construction of the process model, different sources of information are usually used, i.e. process documentation (also from integrated management systems, procedures), observations and interviews are carried out, and more recently, event logs from organisation's IT systems are analysed with use of process mining techniques.

Process modelling is one of the most basic stages of the so-called Business Process Management cycle (Fig. 1). This cycle starts with process identification and process architecture creation. In the process modelling stage a process model ("as-is") is developed to map how the process has been realized in the organization. The current state of the process is documented in the form of one or several models.

The process model facilitates understanding of the process and enables communication between persons involved, helps to assess the complexity of the process and identify problems. Moreover, it is indispensable for further analyses aimed at increasing the efficiency of the process or its redesigning (Dumas et al., 2018).

In the process modelling various detail level of models can be developed, depending on data sources and management purposes.

In a case of industrial processes two main sources of event data for process modelling could be used: high-level events logs (with process activities) from ERP-type IT systems and low-level events logs (e.g. sensor data) from process monitoring systems, without a clear, direct reference to activities of the process being analysed. Low-level event data used for high-level models creation generally requires transformation before the process modelling stage. That transformation could be done with use of various data mining techniques (Günter and van der Aalst, 2006).

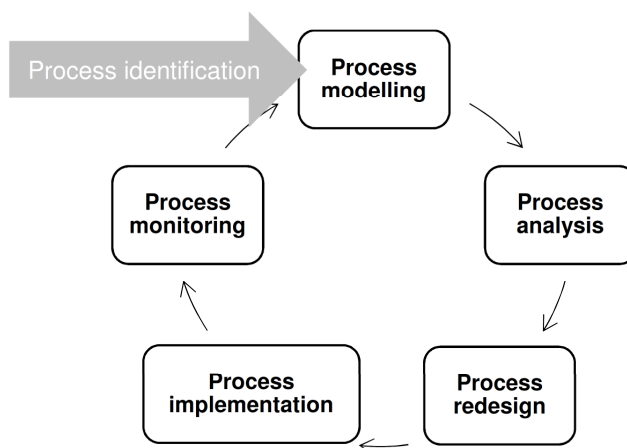


Fig. 1. Business process management cycle

Source: based on (Dumas et al., 2018).

The paper is structured as follows. In the Section 2 introduction to selected process modelling techniques and process mining is presented. Section 3 presents a case study and an example of process modelling based on event log from a delivery company. Discussion of the results is presented in Section 4, whereas conclusions are presented in the last section.

2. METHODOLOGY OF RESEARCH

There are many business process modelling techniques. The most well-known include (Aguilar-Savén, 2004; List and Korherr, 2006; Alotaibi, 2016): flow chart technique, data flow diagram (DFD), role activity diagrams (RAD), Gantt chart, The Integrated Definition for Function Modelling (IDEF), Petri nets, Object oriented methods (i.e. UML), Workflow languages (e.g. YAWL) as well as very popular nowadays – Business Process Modelling Notation (BPMN) (Object Management Group, 2013) or Event - Driven Process Chain (EPC) (Scheer et al., 2005).

Among the mentioned models one should pay special attention to models, that are a graphical and a precise mathematical notation, namely Petri nets.

Petri nets are the oldest and best investigated notation used for process modelling allowing expression of concurrency in a model (van der Aalst, 2016). Their structure consists of four types of elements: places (P), transitions (T), arcs (A) and tokens. An example of Petri net is presented in Figure 2.

The Petri net reflects the process structure, but does not reflect the dynamics of a net. The dynamics is defined by the so-called marking, that can be expressed by $N = (P, T, A, M_0)$, where M_0 is a function defined on the set of places, called the initial marking of the N net (Szyrka, 2008). Initial marking is a function that assigns to each place in the network a non-negative integer denoting the number of tokens in this place.

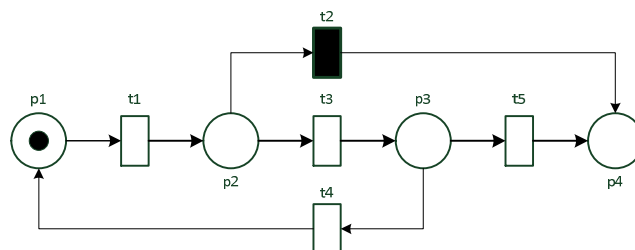


Fig. 2. A Petri net example

The Petri net has specific structural and behavioural properties that are very important and decisive for their suitability for modelling of real processes, including reachability, boundedness, limitation or liveness. For the modelling of processes, special kind of Petri nets is often considered, so-called workflow nets (WF-net) (van der Aalst, 2016).

In general, places of Petri net can represent process states or conditions as well as buffers or specific location (office, room). Transitions represent activities or events. Tokens can represent physical object (e.g. person, product), information, state or condition indicator.

The flow-oriented nature of processes and formalism of Petri nets makes them a natural candidate for the modelling and analysis of workflows (van der Aalst, 2009). Nowadays, Petri nets can be found as quite simple and intuitive notation for modelling of the processes popularised together with process mining techniques.

The idea of process mining can be briefly described as discovering knowledge from event data, available in modern business applications for the purpose of process models creation, monitoring and their improvement (van der Aalst, 2016).

The necessary element for conducting the process mining is an event log containing structured data on the process realisation. Each event in a log refers to a specific activity and concerns a specific case (process instance). Basic information on the process realisation includes: case number, activity name and event timestamp (Table 1). Often, event logs contain also additional information, e.g. about resources in the process (people or devices) that perform or initiate activities, and other data (e.g. cost related to activity).

Table 1.
An example of event log

Case Id	Timestamp	Status	responsible
1	2017-03-31T17:33:07	Order checking	Ann
1	2017-05-08T13:04:27	Order send to magazine	Ann
1	2017-05-16T09:22:15	Products completed	John
1	2017-05-16T09:25:30	Delivery accepted	John
1	2017-05-16T10:12:29	Send delivery	Bas
2	2017-05-15T09:03:29	Order checking	Ann
2	2017-05-15T09:30:23	Order rejection	Ann
3	2017-05-16T09:18:57	Order checking	Ann
3	2017-05-16T15:50:53	Order send to magazine	Ann
3	2017-05-16T16:07:48	Queued	Clemens
3	2017-05-22T17:37:57	Products completed	John
3	2017-05-22T18:13:35	Delivery accepted	John

The main tasks of process mining include (van der Aalst 2016): model discovery, conformance checking and enhancement (Fig.3).

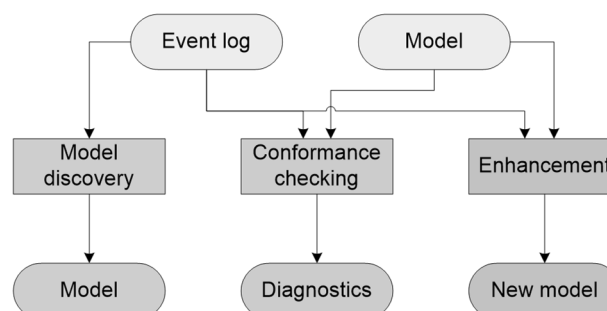


Fig. 3. Process mining tasks

Source: based on (The Process Mining Manifesto, 2012).

Process model discovery involves building a process model based on the event log without using *a priori* information. This analysis is based only on data from information systems at the company's disposal.

During the conformance checking tasks, the existing process model is compared to the event log from the process execution in order to detect deviations (differences) between the model and its actual realization.

Enhancement includes extended process analysis using additional information recorded in the event log (e.g. time or organizational perspective). Taking into account the timestamps from the event log the process bottlenecks can be found. Using the information about the persons carrying out the activities, one can also analyze the work of teams in an organization (e.g. using sociograms).

Process modeling is strictly related to process discovery task. In this task various algorithms for process model discovery can be used e.g. Alpha Miner (van der Aalst et al., 2004), Heuristic Miner (Weijters and Ribeiro, 2011), Inductive Miner (Leemans et al., 2013), Evolutionary Tree Miner (Buijs et al., 2014). These algorithms were implemented in ProM framework, widely known academic open-source software for process mining developed at Technical University of Eindhoven. The case study and modelling of selected process with ProM are presented in the next section.

3. PROCESS MODELLING – A CASE STUDY

A case study concerns the modeling of delivery process based on real data obtained from the courier company. The analyzed process (door-to-door, D2D) begins with the sender preparing a parcel. The parcel at the moment of receipt by a courier is scanned by him with a mobile device, then parcel receives the status: "Adopted by the courier from the client". If the courier does not have a mobile device, the parcel is taken from the client and receives the status given by the warehouse: "Adopted by warehouse from the courier". The next step is the admission to the warehouse and there are possible statuses: "Adopted by sending division" and "Adopted by delivering division". Sometimes it happens that a parcel sent in a sending division is also delivered by him. However, the status of "Adopted by sending division" is a much more frequent situation. After the "Adopted by sending division", the parcel is sorted and sent from the division, receiving the "Sorted" and "Sent from division" statuses successively to be adopted by the delivering division. The parcel with "Adopted by delivering division" status is sent directly for delivery, unless the delivering division is also the sending division. Parcels normally are received and prepared for delivery on the following day. In the case of direct acceptance of the parcel in the delivering division, it must be retained until the next day. Then it receives "Warehouse state inventory" status. The final stage is parcel delivery, expressed by "Adopted for delivery" status and "Delivered" status, that finalize the process (Czekaj, 2017). Theoretical model of analyzed process in BPMN notation is presented in Figure 4.

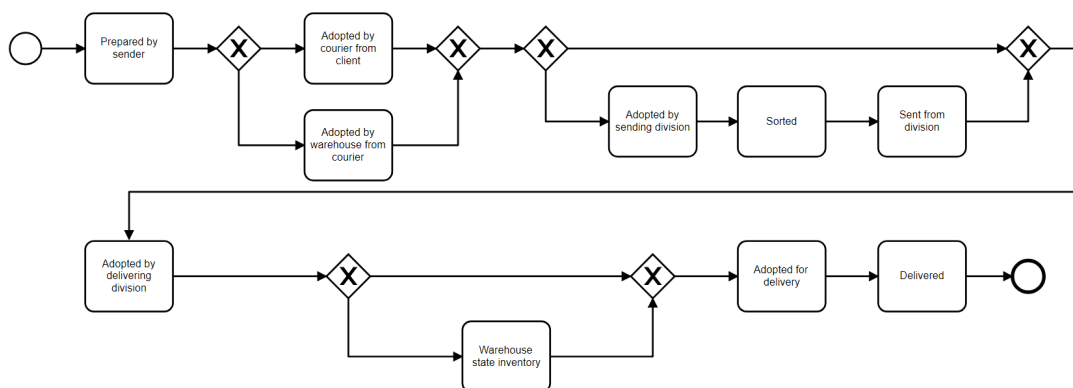


Fig. 4. Model of delivery process

The event log used for process modelling includes data on 5274 delivered parcels and 40947 events within one month of the division operation (Fig. 5).

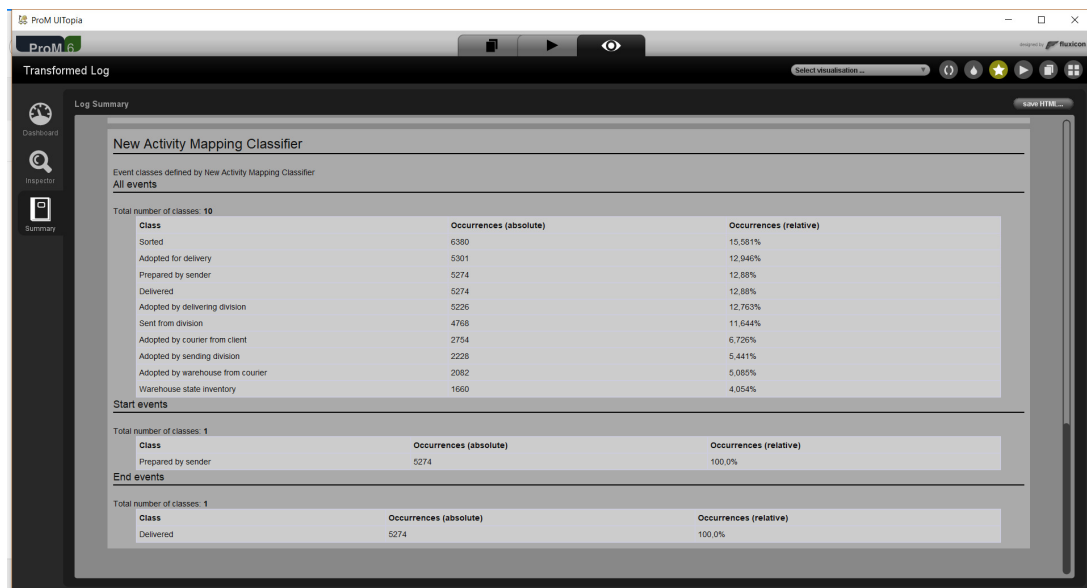


Fig. 5. Even log summary

The most frequent statuses are: “Sorted” (16%), “Adopted for delivery” (13%), “Prepared by sender” (13%), “Delivered” (13%), “Adopted by delivering division” (13%). The distribution of events shows that some parcels has been sorted more than once. All process instances has been started with “Prepared by sender” status and finished with “Delivered” status. In the event log 91 case variants are included, so more complex model than theoretical can be expected. Based on the event log, process model in a form of Petri net was generated (Fig.6).

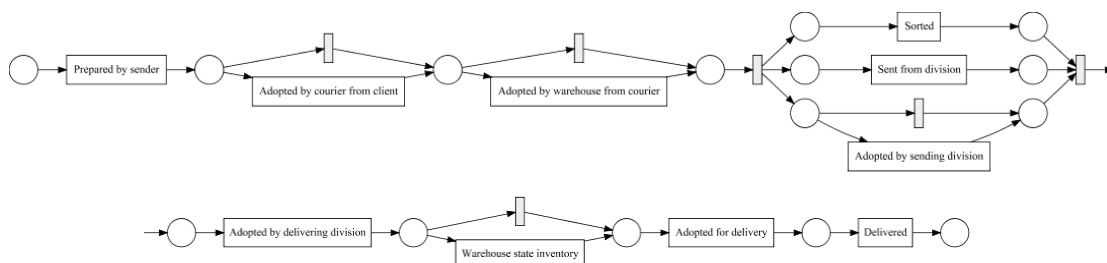


Fig. 6. Petri net model

Modelling of the process was done in ProM Lite 1.2 software using the Inductive Miner (IM) algorithm. The choice of IM algorithm guarantees that the resultant Petri net model is sound and deadlocks free. It is also one of the most frequently chosen algorithms for discovering process models due to best performance in terms of model quality criteria (Augusto et al. 2017). Discussion of the results is presented in the next section.

4. DISCUSSION

Comparison of the discovered model with theoretical process model shows clear differences between them, especially in the first (upper) part of the model.

Firstly, model allows situation that two alternative statuses “Adopted by courier from client” and “Adopted by warehouse from courier” are sequenced. Indeed, in the event log in 28 traces such inappropriate situation was recorded. This part of the model enables also execution of the process with disregarding of mentioned statuses.

Moreover, model enables various execution order of “Sorted”, “Sent from division” and “Adopted by sending division” statuses, that should be executed in a sequence. Alternative of statuses “Adopted by sending division” and “Adopted by delivering division” is not expressed in the presented model.

Main reason of such model construction is strictly related to event log content and variants of process execution, that are recorded there (Fig.7). There are many traces with unexpected order of statuses, that indicate process deviations from theoretical model.

Detail analysis of differences between theoretical process model and an event log can be done with conformance checking techniques, enabling identification of commonalities and discrepancies (Munoz-Gama, 2016).

Conformance checking can also be used for evaluation of discovered models. The model can be evaluated with regard to four competing quality criteria e.g. fitness, precision, generalization and simplicity (van der Aalst, 2016).

In a practice, evaluation of model is not a trivial task. However, in ProM there is possibility to calculate values of fitness, precision and generalization for a given model and an event log.

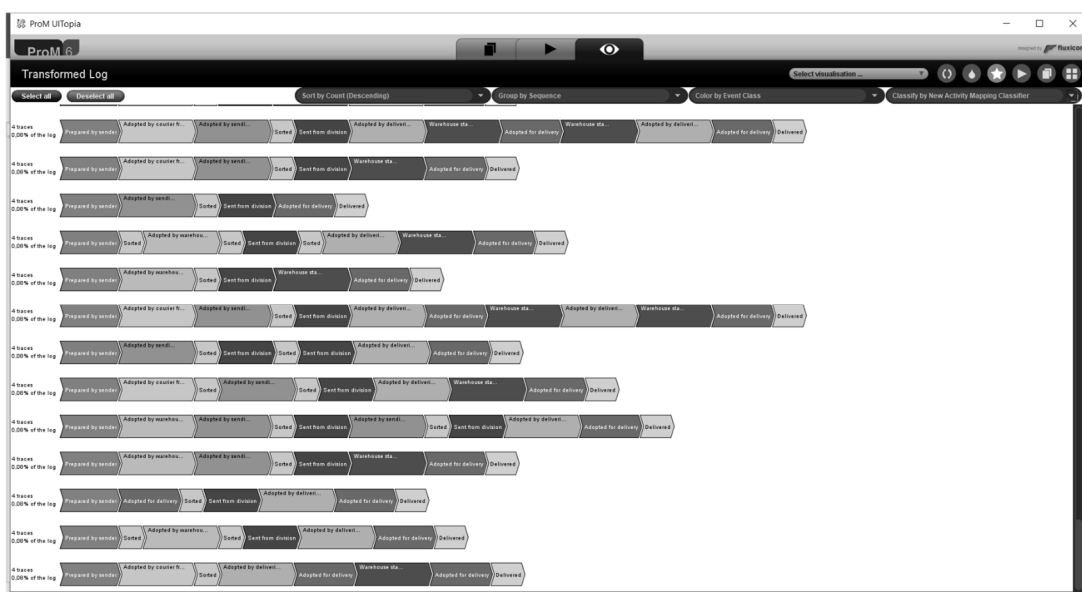


Fig. 7. An example of traces’ variants in the analysed event log

Discovered model is quite well general (0.99), it means that model has ability to preserve additional behaviour (not recorded in the event log). High generalization has unfortunately impact on precision (0.81) – model allows too much additional behaviour in comparison to the event log. Model has ability to replay behaviour recorded in the event log expressed by the high fitness value (0.96).

Real performance of the analyzed process differs from theoretical model due to inappropriate process execution which can be seen in the event log visualization (Fig. 7).

Conformance checking and deeper analysis of context information stored in the event log (also with use of process mining techniques) should provide more detail information about causes of problems in the process and its performance. These results can be used for further process improvement.

CONCLUSIONS

Process models are essential elements for business process management purposes. Their knowledge enables process analysis, monitoring and redesign activities in organization. The models can be expressed with use of various modelling techniques. Among them precise

mathematical models can be found, namely Petri nets. Formalism of Petri nets and their construction elements enable extended modelling and analysis of workflows.

In the paper as one of the data sources for process modelling, the event logs from IT systems were mentioned. Usage of event logs for process modelling is possible with process mining techniques and dedicated software.

Nowadays popularity of process mining techniques, that greatly facilitates automatic process modelling and its analysis, rapidly grows. Process mining enables process model discovery based on an event log, conformance checking and enhancement of the process by deep analysis of its various perspectives (organizational, case, time).

In the paper case study of the delivery process (D2D) and the process modelling with ProM software were presented. Discovered process model in a form of Petri net reflects real performance of the process recorded in the event log. Its comparison with theoretical model shows clear differences in process execution and gives insights about discrepancies related to the parcel delivery process.

ACKNOWLEDGEMENTS

This paper presents the results of research conducted at AGH University of Science and Technology – contract no 11.11.100.693.

REFERENCES

- Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework, *International Journal of Production Economics*, 90(2), pp. 129-149.
- Alotaibi, Y. (2016). Business process modelling challenges and solutions: a literature review. *Journal of Intelligent Manufacturing*, 27(4), pp. 701–723.
- Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F., Marrella, A., Mecella, M. and Soo, A. (2017). Automated Discovery of Process from Event Logs: Review and Benchmark. arXiv:1705.02288
- Buijs, J. C., van Dongen, B. F., and van der Aalst W.M.P (2014). Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *International Journal of Cooperative Information Systems*, 23(1) p. 1440001.
- Business Process Model and Notation™ (BPMN™) 2.0.2 Object Management Group, 2013 (<http://www.omg.org/spec/BPMN/index.htm>).
- Czekaj, S. (2017). Analysis of the parcels delivery process in a selected company using process mining techniques. BSc., AGH University of Science and Technology (in Polish).
- Dumas, M., La Rosa, M., Mendling, J. and Reijers, H. (2018): *Fundamentals of Business Process Management*, Berlin Heidelberg, Springer-Verlag.
- Günther, C.W and van der Aalst, W.M.P. (2006). Mining Activity Clusters from Low-Level Event Logs. BETA Working Paper Series, WP 165, Eindhoven University of Technology, Eindhoven.
- Leemans, S. J., Fahland, D. and van der Aalst, W.M.P (2013). Discovering block-structured process models from event logs containing infrequent behavior. In: *International Conference on Business Process Management*. Springer, pp. 66–78.
- List, B. and Korherr, B. (2006). An evaluation of conceptual business process modelling languages. In: *Proceedings of the 2006 ACM symposium on Applied computing (SAC '06)*. ACM, New York, USA, pp. 1532-1539.
- Munoz-Gama, J. (2016). Conformance checking and diagnosis in process mining – comparing observed and modeled processes. *Lecture notes in business information processing*, vol 270. Springer, Cham
- Scheer, A.-W., Thomas, O. and Adam, O. (2005). *Process Modeling using Event-Driven Process Chains*. In: M. Dumas, W. M. P. van der Aalst and A. H. M. Ter Hofstede, eds., *Process-Aware Information Systems: Bridging People and Software through Process Technology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Szpyrka, M. (2008). Petri nets in design and analysis of concurrent systems. Warszawa: WNT (in Polish).
- The Process Mining Manifesto* by the IEEE Task Force on Process Mining, In: F. Daniel, K. Barkaoui, S. Dustdar, eds., *BPM 2011 Workshops, Part I, LNBIP 99*, pp. 169–194. Berlin: Springer-Verlag, 2012.
- Van der Aalst, W.M.P. (2009). *Process-Aware Information Systems: Lessons to Be Learned from Process Mining*. In: *Transactions on Petri Nets and Other Models of Concurrency II, Lecture Notes in Computer Science*, vol. 5460, Berlin: Springer-Verlag, pp. 1-26.
- Van der Aalst, W.M.P. (2016). *Process Mining: Data Science in Action*. Berlin: Springer-Verlag.

- Van der Aalst, W.M.P., Weijters, A.J.M.M. and Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9), pp.1128-1142.
- Weijters, A. and Ribeiro J. (2011). Flexible heuristics miner (fhm). *IEEE Symposium on Computational Intelligence and Data Mining*, IEEE, pp. 310–317.