

# Towards a Framework for Data Enhanced Process Models in Process Mining

Jonas Cremerius

Hasso Plattner Institute, University of Potsdam, Potsdam, Germany  
jonas.cremerius@hpi.de

**Abstract.** Understanding and improving business processes have become important steps towards success for organizations. Getting insights about a process is not only based on the control flow, but also on the data generated within the process. Today, almost every process step generates data, especially in the health care domain. So far, most analyzes inside a process model are limited to time analyzes and identification of decision logic. However, various attributes can be linked to events, such as the number of abnormal lab values derived from a lab test, which could be displayed in the process model. This can help to explore the process with domain experts, where they can choose the attributes of interest for each event and observe their influence on the process, i.e. the influence of abnormal lab values on the treatment process. Therefore, the interplay of event attributes and the control flow can be observed directly in the process model.

**Keywords:** Process Mining · Process Enhancement · Process Model Extension

## 1 Introduction

Business process models play a central role in exploring and analyzing the organization's business processes. With the help of process mining, process models can be derived from real-world process execution data [3]. Process mining is often conducted in the healthcare sector, as hospitals are becoming increasingly aware of the need to improve their processes [8]. Despite the increasing availability of data, adequate support for displaying or analyzing event attributes in a discovered process model is still lacking. So far, process enhancement inside the discovered process model is limited to time analyzes and decision logic, whereas organizational aspects are separately analyzed [15]. This is also represented in today's process mining tools. Fluxicon Disco<sup>1</sup> provides time analyzes only, whereas Lana Labs<sup>2</sup>, Celonis<sup>3</sup>, and ProM<sup>4</sup> do not provide attribute analyzes

<sup>1</sup> <https://www.fluxicon.com/disco/>

<sup>2</sup> <https://www.lanalabs.com/>

<sup>3</sup> <https://www.celonis.com/>

<sup>4</sup> <https://www.promtools.org/>

inside the process model. Only Process Diamond<sup>5</sup> allows displaying one attribute, which must be the same for all events. In several treatment processes, such as the diagnosis and treatment of heart failure, a large amount of data is generated [1]. Example events include “Analyze Lab Values” and “Treat Patient in Intensive Care Unit (ICU)”. For both events, a different set of attributes is produced, such as the number of abnormal lab values and the Glasgow Coma Scale (GCS). Considering the effect of these attributes on a process may provide insights about a certain diagnostic or treatment. As current process mining tools are not capable of exploring a process in respect to attribute values, we propose a respective functionality that enables business analysts and domain experts to identify conspicuous behavior not only in the control flow, but also in the attribute values.

## 2 Related Work

Process Enhancement is the extension or improvement of an existing process model using information about the actual process recorded in some event log [3]. There exist several approaches which analyze the activity duration and waiting time in a process model to identify, for example, bottlenecks [10, 14]. Additional analyzes inside the process model include identification of constraints at decision points [7, 11]. Further, Process Enhancement involves the organizational analyzes by looking at the assignment of human resources to process activities [12]. Enhancement is also conducted offside the process model where, for instance, different process characteristics are correlated with each other [5, 13]. Case and event attributes have been used in filtering process variants and clustering of traces. For instance, [4] clusters traces to discover process variants by different case and event attributes, such as age and the body part of an image analyzes.

## 3 Research Objective

None of the approaches mentioned above looked into displaying event attributes directly in the process model. So far, they are limited to time analyzes and identification of decision points. Our research aims to fill this gap and provides a framework to link event attributes with their respective event in the process model. Today’s data sets, such as MIMIC-IV, include various event attributes, such as the number of abnormal lab values [2]. As processes can get complex, different attributes might be interesting for the events. An example process is displayed in Fig. 1, which illustrates a simple treatment process. The events have their attributes attached in the process model. The first two events “Perform X-Ray” and “Perform CT” show the frequency of findings in specific body regions (heart and lung). After that, the lab values are analyzed, which show the frequency of abnormal lab values observed for each lab value. Then, the treatment is conducted,

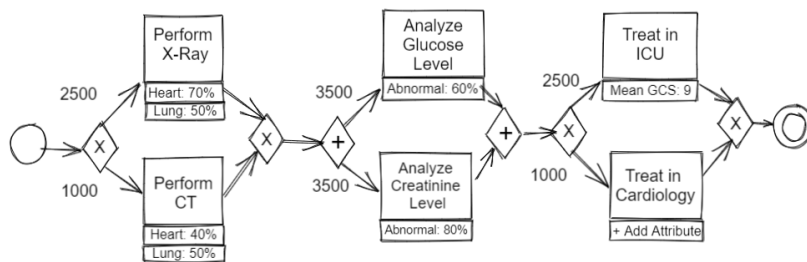
---

<sup>5</sup> <https://www.processdiamond.com/>

which can happen in the Intensive Care Unit (ICU) or Cardiology, where the mean Glasgow Coma Score (GCS) is shown. This process can be similar for several diseases on this abstraction level. Therefore, the insights regarding the control flow might be limited. However, the event attributes can be different, as the glucose level is more interesting for type 2 diabetes and the creatinine level for kidney disease [6, 9]. This could help to assess how meaningful analyzing a specific lab value for several diseases is. Furthermore, the effect of an attribute on the process can be explored with the help of a process model. For example, if the process of patients with an abnormal glucose level is of interest, one could just click on the attribute, triggering filtering according to the lab value. Then, not only the change in the process flow, but also in the event attributes can be seen, such as the mean GCS or the frequency of findings in the heart region of an image analyzes, which might be different for patients with an abnormal lab value. This could reveal novel insights, as one might not have thought that patients with an abnormal lab value also have a high prevalence of findings in the heart region of an X-ray.

We want to enable this kind of analyzes in the process model, which could lead to a more comprehensive exploration of processes together with a domain expert. The framework defines how different types of attributes are displayed in the process model and which computations can be performed on them, such as minimum, maximum, or mean. As process attributes can be different depending on the application context, we want to enable the domain experts to choose the attributes of interest. Nevertheless, an attribute recommendation system could be implemented, which helps to choose the appropriate attributes for an activity based on machine learning or descriptive statistics. Additionally, the framework could help to highlight events sharing the same attributes, such as different lab values or medical imaging techniques. Therefore, the following research questions need to be answered:

- How can event attributes be displayed in the process model (categorical vs. continuous variables)?
- How can the framework help to gain new insights about the process (detection of process variants, dependencies between attributes, etc.)?



**Fig. 1.** Process model with data attributes displayed for each event. CT, Computed Tomography; ICU, Intensive Care Unit; GCS, Glasgow Coma Score

## 4 Conclusion

This position paper discusses the need for looking at the inclusion of event attributes directly in the process model. With the increasing data availability, a more comprehensive view of the process is possible and different process variants can be explored. As process models can be used as a means of communication between process analysts and domain experts, incorporating event data in the model has the potential to improve that communication by illustrating the control flow and event attributes in one place.

## References

1. Acute and chronic heart failure guidelines, <https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Acute-and-Chronic-Heart-Failure>
2. Mimic iv, <https://mimic-iv.mit.edu/>
3. van der Aalst, W.: Process Mining. Springer Berlin Heidelberg (2016). <https://doi.org/10.1007/978-3-662-49851-4>, <https://doi.org/10.1007/978-3-662-49851-4>
4. Hompes, B., Buijs, J., Aalst, W., Dixit, P., Buurman, J.: Discovering deviating cases and process variants using trace clustering (11 2015)
5. de Leoni, M., van der Aalst, W.M., Dees, M.: A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems* **56**, 235–257 (Mar 2016). <https://doi.org/10.1016/j.is.2015.07.003>, <https://doi.org/10.1016/j.is.2015.07.003>
6. Levey, A.S., Perrone, R.D., Madias, N.E.: Serum creatinine and renal function. *Annu Rev Med* **39**, 465–490 (1988)
7. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Measuring the precision of multi-perspective process models. In: *Business Process Management Workshops*, pp. 113–125. Springer International Publishing (2016). [https://doi.org/10.1007/978-3-319-42887-1\\_10](https://doi.org/10.1007/978-3-319-42887-1_10), [https://doi.org/10.1007/978-3-319-42887-1\\_10](https://doi.org/10.1007/978-3-319-42887-1_10)
8. Martin, N., De Weerd, J., Fernández-Llatas, C., Gal, A., Gatta, R., Ibáñez, G., Johnson, O., Mannhardt, F., Marco-Ruiz, L., Mertens, S., Muñoz-Gama, J., Seoane, F., Vanthienen, J., Wynn, M.T., Boilève, D.B., Bergs, J., Joosten-Melis, M., Schretlen, S., Van Acker, B.: Recommendations for enhancing the usability and understandability of process mining in healthcare. *Artificial Intelligence in Medicine* **109**, 101962 (2020). <https://doi.org/https://doi.org/10.1016/j.artmed.2020.101962>, <http://www.sciencedirect.com/science/article/pii/S0933365720312276>
9. Olokoba, A.B., Obateru, O.A., Olokoba, L.B.: Type 2 diabetes mellitus: a review of current trends. *Oman Med J* **27**(4), 269–273 (Jul 2012)
10. Rogge-Solti, A., van der Aalst, W.M.P., Weske, M.: Discovering stochastic Petri nets with arbitrary delay distributions from event logs. In: Lohmann, N., Song, M., Wohed, P. (eds.) *Business Process Management Workshops*. pp. 15–27. Springer International Publishing, Cham (2014)
11. Rozinat, A., Mans, R., Song, M., van der Aalst, W.: Discovering simulation models. *Information Systems* **34**(3), 305–327 (May 2009). <https://doi.org/10.1016/j.is.2008.09.002>, <https://doi.org/10.1016/j.is.2008.09.002>
12. Schönig, S., Cabanillas, C., Jablonski, S., Mendling, J.: A framework for efficiently mining the organisational perspective of business processes. *Decision Support Systems* **89**, 87–97 (Sep 2016). <https://doi.org/10.1016/j.dss.2016.06.012>, <https://doi.org/10.1016/j.dss.2016.06.012>
13. Schönig, S., Ciccio, C.D., Maggi, F.M., Mendling, J.: Discovery of multi-perspective declarative process models. In: *Service-Oriented Computing*, pp. 87–103. Springer International Publishing (2016). [https://doi.org/10.1007/978-3-319-46295-0\\_6](https://doi.org/10.1007/978-3-319-46295-0_6), [https://doi.org/10.1007/978-3-319-46295-0\\_6](https://doi.org/10.1007/978-3-319-46295-0_6)
14. Wynn, M., Poppe, E., Xu, J., ter Hofstede, A., Brown, R., Pini, A., van der Aalst, W.: Processprofiler3d: A visualisation framework for log-based process performance comparison. *Decision Support Systems* **100**, 93 – 108

(2017). <https://doi.org/https://doi.org/10.1016/j.dss.2017.04.004>, <http://www.sciencedirect.com/science/article/pii/S0167923617300623>, smart Business Process Management

15. Yasmin, F., Bukhsh, F., Silva, P.: Process enhancement in process mining: A literature review (12 2018)