

# Towards a Comprehensive Methodology for Process Mining

Kiarash Diba

Hasso Plattner Institute, University of Potsdam, Potsdam, Germany  
{kiarash.diba}@hpi.de

**Abstract.** Process mining exploits data recorded in information systems of organizations to unleash insight and knowledge into their operational processes. As process mining techniques are reaching maturity, their applications are becoming more widespread across various domains. Therefore, more research on the methodological and practical perspective is required to steer and guide these applications to successful result. This position paper sketches the first steps to be taken toward a standard methodology for process mining.

**Keywords:** Process Mining · Process Mining Methodology · Process Mining Reference Model

## 1 Introduction and Motivation

Process Mining has evolved into a well-known technology to provide valuable insight into the underlying processes and workflows of organizations. During the recent years, many process mining techniques and algorithms have been developed and are reaching maturity and their applications have been investigated and proved valuable across variety of domains. Despite this level of maturity in techniques and algorithms, the broader process mining discipline has not yet matured. Although process mining projects involve various steps and activities from extracting and preparing required data to providing useful knowledge and insight, the entire spectrum of activities have not been thoroughly investigated. Instead most of the academic focus has been concentrated on the development and improvement of techniques and algorithms. Besides, most case studies and projects have been carried out in an unstructured and ad-hoc manner involving a great amount of manual and time intensive work and there is little guidance on conducting such projects successfully in both industrial and academic settings. Therefore, this work focuses on methodological aspects of process mining rather than on specific methods, in order to provide well-defined foundations for process mining. This not only helps practitioners and academics in conducting successful process mining projects, but positions process mining techniques into a broader spectrum of process-related knowledge discovery and sheds more light on steps that have received less research attention. Thus, inspired by related works in the field of data mining and practical experience, this work takes the initial steps towards a comprehensive standard methodology for process mining.

The remainder of this paper is structured as follows. The next section discusses related works and their limitations. Section 3 outlines the approach to be followed and steps to be taken for establishment of the comprehensive methodology, followed by a high-level overview of the initial developments of the methodology in section 4. Finally, section 5 concludes the paper.

## 2 Related work

Methodology provides the theoretical foundation for understanding which methods, set of methods, or best practices can be applied to a specific case, which is employed for the design, planning, implementation and achievement of project objectives [5]. Currently there are few works focusing on methodologies for process mining namely the L\* lifecycle model [1], Process Diagnostic Method (PDM) [3], its extension in healthcare domain [6], and PM<sup>2</sup> [4]. However, these methodologies are not comprehensive and suitable for every project and have a number of limitations. Besides, they have not been widely evaluated and applied in various projects. PDM has a narrow scope focusing on a limited number of capabilities of process mining. Besides, it neglects the importance of business considerations, planning and domain knowledge [4]. L\* lifecycle model primarily focuses on discovery of a single process model enriched with performance and resource information and therefore, it is more suitable for structured processes and narrowly scoped projects [4]. In addition, non of the two offer sufficient flexibility and iterations. The sequence of activities suggested are assumed to be followed rather strictly for every project which is rarely the case in complex projects. In different projects depending on different requirements, some steps might be skipped or performed in a different sequence. Although PM<sup>2</sup> addresses a few limitations of the previously mentioned methodologies, it can still be improved and extended with more flexibility, more detailed and specific steps, techniques, best practises and practical guidelines. Successful examples of methodologies can be found in the field of data mining where similar works such as CRISP\_DM [7] have been applied successfully for many years in variety of settings.

## 3 Proposed Approach

In order to construct a comprehensive methodology we will first define methodology and clarify what a methodology is and what it should contain and motivate the use and benefits of such methodologies. Then we will formally compare related work from both process mining literature and related fields such as data mining based on their structure, applicability and reputation. We will also analyze case studies and use cases from a structural point of view before establishing the methodology. In addition, we will make use of questionnaires among process mining experts both in academia and industry to consolidate the motivation and formation of the methodology. Afterwards, the methodology needs to be tested, evaluated and consequently adjusted followed by continuous refinements. The

next section provides an initial high-level overview of the methodology to be developed.

## 4 Outline of the methodology

A methodology should be able to be applied to specific projects in different context with different goals and requirements while remaining as generic as possible [5]. Therefore, the proposed methodology in this work will consist of different levels of abstraction each having different characteristics and different purposes. The highest level consists of general phases and stages involved in process mining projects. This high level view needs to be as generic as possible accounting for all possible scenarios and contexts process mining can be applied. The lower levels consist of more detailed generic and specific activities for each phase driven by the context of specific project goals and requirements. The lowest level of methodology involves an actual run of these activities for a specific project. This hierarchical nature of the methodology allowing flexibility and addressing the challenge of balancing genericity and specificity is one of the main features of the methodology.

In addition, the methodology contains a user guide with best practises, common approaches and techniques in order to guide the user with various challenges. The methodology describes the overall approach to extract knowledge and insight into processes and provide a roadmap to follow while planning and carrying out process mining projects, addressing two of the process mining challenges stated in the process mining manifesto [2] namely Improving Usability for Non-Experts, and Improving Understandability for Non-Experts. It also facilitates and encourages efforts for automation and reusability of process mining project flows and currently manual (or partially automated) and time consuming steps such as data extraction and preparation.

Process mining projects usually involve the following high level phases:

**Planning** which focuses on both the business and technical aspects of the project. In this phase project plan, requirements, objectives and available resources are identified and discussed and a concrete project plan is prepared.

**Data Discovery** containing data extraction and event log preparation. The journey here, which could be one of the most challenging parts of the project starts with identifying relevant data sources, finding, extracting, merging and cleaning the extracted data and leads to preparing an event log in required formats.

**Process Discovery** consisting of explorative analysis, process overview and control-flow discovery. Depending on requirements and the nature of the project, explorative or goal-driven, the steps taken in this phase vary. Initial insights and statistics into the process is gained which assist the following step of analysis. based on this initial insights project might go back to previous phases to modify and adjust plans or to collect additional data or to adopt different views on the data.

**Analysis** focusing on the main analysis and evaluation of the result. Different types of analysis can be performed and process mining techniques can be combined with data mining, statistics and other types of analysis to provide useful insight and knowledge and address the project objectives.

**Knowledge Transfer** phase which can be reporting diagnostics and improvement insights and/or preparing a monitoring system for operational support. Due to the iterative nature of process mining projects, there should be multiple iterations introduced between different phases.

## 5 Conclusion

This paper outlines an overview and the landscape of a comprehensive methodology for process mining. The prospective methodology will involve several hierarchical levels, consisting of high level phases to specific activities for each phase. In addition, a user guide and best practises and techniques will be included to facilitate successful projects in various settings. Continuous extension, refinement and evaluation need to be performed before and after establishment of the methodology to ensure generality, completeness and applicability.

## References

1. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer (2011)
2. van Der Aalst, W.M.P, Adriansyah, A., De Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., Van Den Brand, P., Brandtjen, R., Buijs, J. and Burattin, A.: Process mining manifesto. In International Conference on Business Process Management, pp. 169-194. Springer, Berlin, Heidelberg (2011).
3. Bozkaya, M., Gabriels, J., Werf, J.: Process diagnostics: a method based on process mining. In: International Conference on Information, Process, and Knowledge Management, eKNOW 2009, pp. 2227. IEEE (2009)
4. van Eck, M.L., Lu, X., Leemans, S.J. and van der Aalst, W.M.: PM <sup>2</sup>: A Process Mining Project Methodology. In International Conference on Advanced Information Systems Engineering, pp. 297-313. Springer, Cham (2015).
5. Irny, S.I. and Rose, A.A.: Designing a Strategic Information Systems Planning Methodology for Malaysian Institutes of Higher Learning (isp- ipta), Information System 5(1), (2005).
6. Rebuge, A., Ferreira, D.R.: Business Process Analysis in Healthcare Environments: a Methodology based on Process Mining. Information Systems 37(2), 99116 (2012)
7. Shearer, C.: The crisp-dm model: the new blueprint for data mining. Journal of data warehousing 5(4), 1322 (2000)