Scaling Process Mining to Turn Insights Into Actions

Wil M.P. van der Aalst[1] and Josep Carmona[2]

[1] Process and Data Science (PADS), RWTH Aachen University, Germany
wvdaalst@pads.rwth-aachen.de
[2] Universitat Politècnica de Catalunya, Spain
jcarmona@cs.upc.edu

Abstract. This final chapter reflects on the current status of the process mining discipline and provides an outlook on upcoming developments and challenges. The broader adoption of process mining will be a gradual process. Process mining is already used for high-volume processes in large organizations, but over time process mining will also become the "new normal" for smaller organizations and processes with fewer cases. To get the highest return on investment, organizations need to "scale" their process mining activities. Also, from a research point-of-view, there are many exciting challenges. On the one hand, many of the original problems (e.g., discovering high-quality process models and scaling conformance checking) remain (partly) unsolved, still allowing for significant improvements. On the other hand, the large-scale use of process mining provides many research opportunities and generates novel scientific questions.

Keywords: Process Mining · Execution Management · Process Management

1 Process Mining: Overview and Summary

The chapters in this book illustrate the broadness of the process mining discipline. The interplay between data science and process science provides many challenges and opportunities [1]. In this book, we aim to provide a comprehensive overview. There are many dimensions to characterize the 16 earlier chapters.

- *Theory-driven* versus *application-driven*.
- *Backward-looking* (e.g., process discovery and conformance checking) versus *forward-looking* (e.g., simulation and predictive process analytics).
- *Simple* control-flow-oriented event logs versus *complex* object-centric event data considering *different types of objects* and attributes.

In the first chapter of this book [3], we started with Figure 1 showing a 360 degrees overview of process mining. The subsequent chapters have been focusing on different parts of the pipeline depicted in Figure 1. The initial chapters focused on *process discovery*, starting with creating a simple Directly-Follows Graph (DFG) followed by a range of alternative, more sophisticated, techniques. As shown, process discovery is an important topic, but also very difficult [1]. Event data do not contain negative examples and the positive examples typically only cover a fraction of all possible behaviors. Mixtures of choice, concurrency, and loops make process discovery a *notoriously difficult task* with many *trade-offs*. Also, process models may be used for different purposes.

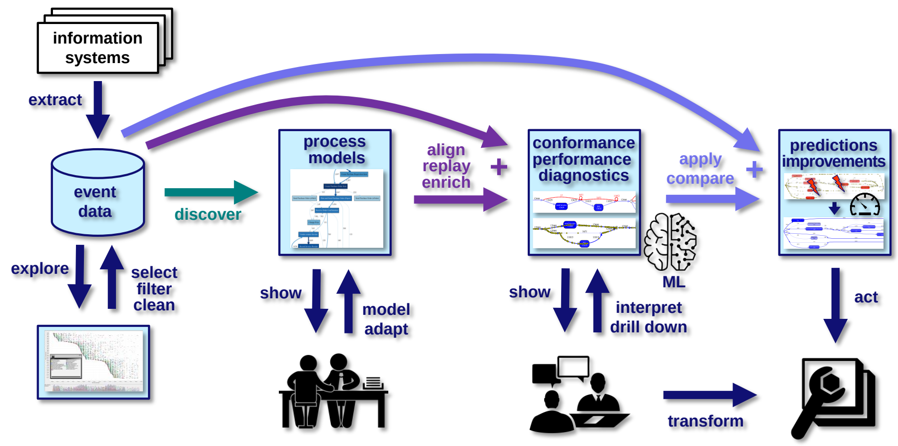

Fig. 1. Process mining uses event data extracted from information systems to provide insights and transparency that, ultimately, should lead to process improvements (i.e., process redesign, improved control, and automation).

After discovery, the focus shifted to *conformance checking* [1, 5]. Here the input consists of both modeled and observed behavior. For example, a multiset of traces is compared with an accepting Petri net. Surprisingly, state-of-the-art conformance checking techniques tend to be more demanding than discovery techniques (from a computational point of view). Computing alignments corresponds to solving optimization problems that grow exponentially in the size of the model and the length of traces in the event log.

Several chapters discussed the importance and complexity of data extraction and preprocessing. Later chapters focused on practical applications and more advanced topics such as model enhancement, streaming process mining, distributed process mining, and privacy-preserving process-mining techniques.

Figure 2 shows another overview of the building blocks of a successful process mining solution. The top of Figure 2 shows examples of application domains where process mining can be used. In this book, we elaborated on applications in healthcare, auditing, sales, procurement, and IT services. However, process mining is a generic technology that can be used in any domain.

In the remainder of this concluding chapter, we take a step back and reflect on the developments in our discipline. Section 2 discusses the inevitability of process mining, but also stresses that concepts such as a Digital Twin of an Organization (DTO) are still far away from being a reality. Section 3 explains that it is important to scale process mining. Finally, Section 4 provides an outlook also listing research challenges.

2 Process Mining as the New Normal

Although process mining has proven its value in many organizations, it is not so easy to create a convincing *business case* to justify investments [1]. The reason is that process

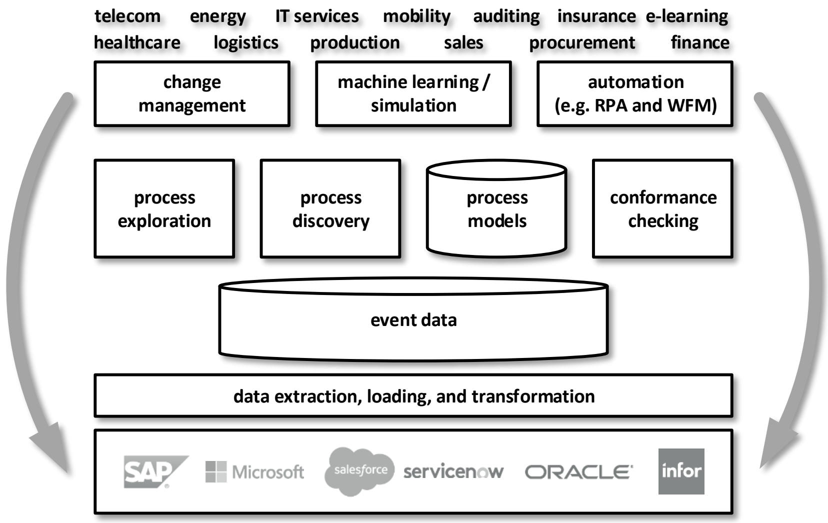

Fig. 2. Process mining can be used in any application domain. However, it may be non-trivial to extract accurate event data and turn process mining results into actions. Change management and automation play a key role in realizing sustained improvements (as indicated by the two arcs closing the loop).

mining will most likely reveal performance and compliance problems, but this does not imply that these are automatically solved [8]. Financial and technical debts are well-known concepts. However, most organizations tend to ignore their *Operational Process Debts* (OPDs). OPDs cause operational friction, but are difficult to identify and address. Although process mining results are often surprising, they typically reveal OPDs that were known to some, but not addressed adequately. Making these OPDs visible and transparent helps to address them.

In [2], the first author coined the term *Process Hygiene* (PH) to stress that process mining should be the "new normal" not requiring a business case. Just like personal hygiene, one should not expect an immediate return on investment. We know that activities such as brushing our teeth, washing our hands after going to the toilet, and changing clothes are the right thing to do. The same applies to process mining activities, i.e., process hygiene serves a similar role as personal hygiene. People responsible for operational processes need to be willing to look at possible problems. Objectively monitoring and analyzing key processes is important for the overall health and well-being of an organization. Process mining helps to ensure process hygiene. Not using process mining reflects the inability or unwillingness to manage processes properly. Fortunately, an increasing number of organizations is aware of this.

Although process mining is slowly becoming the "new normal", most organizations will *not* be able to use the forward-looking forms of process mining. As long as the extraction of event data, process discovery, and conformance checking are challenging for an organization, it is unlikely that machine learning and other forward-looking techniques (including artificial intelligence and simulation) will be of help. Terms such as the *Digital Twin of an Organization* (DTO) illustrate the desire to autonomously manage, adapt, and improve processes. Gartner defines a DTO as “a dynamic software model of any organization that relies on operational and/or other data to understand how an organization operationalizes its business model, connects with its current state, responds to changes, deploys resources and delivers exceptional customer value”. Creating a DTO can be seen as one of the grand challenges in information systems, just like autonomous driving in mobility. However, just like the development of self-driving cars, the process will be slow with many minor incremental improvements.

3 Scaling Process Mining

One of the main conclusions in [6] is that process mining needs *scale* to be most cost effective. Organizations need to aim for the *continuous* usage of process mining for *many processes* by *many people*. Initially, process mining was primarily used in process improvement projects. In such projects, a problematic process is analyzed to provide recommendations for improvement. Since data extraction is often the most problematic step, such projects often struggle to get results quickly. Moreover, the “end product” of such a project is often a just a report. To improve the process, change management and automation efforts are still needed. Therefore, traditional process mining projects struggle to realize a good Return on Investment (ROI).

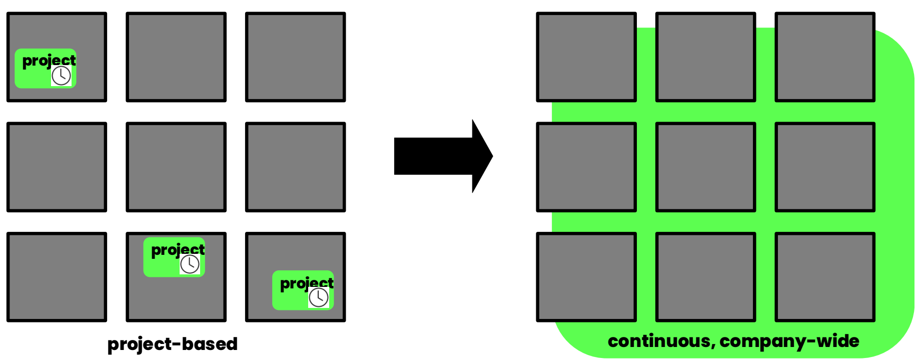

Fig. 3. Scaling process mining to maximize the benefits.

Therefore, process mining should not be seen as a project, but as a *continuous company-wide* activity as shown in Figure 3. There are several reasons for this.

- If data extraction is done properly, the initial efforts are high, but this can be repeated without much extra work. Once the data extraction pipeline is realized, it is possible to *continuously* produce process mining results based on new event data.
- Process mining is a *generic* technology. Hence, investments in software and people should be spread over many processes and organizational units. For example, an insurance company that has multiple products (e.g., different types of insurance) and

multiple offices (in different countries and cities) should not limit process mining to one product or one location.

- Organizational change often requires commitment from many stakeholders. Therefore, results should be *visible* for *all* involved in the process. If performance and compliance problems are only visible to a small group of experts, it is difficult to realize durable behavioral changes. Many improvement projects fail because people slip back into old ways of working after some time.

Compare process mining for an organization to creating weather forecasts for a country. It does not make any sense to create a weather forecast for just one city on a particular day. Investments only make sense if one is able to create a weather forecast for any city on any day. Similarly, process mining is most effective when applied to many processes continuously.

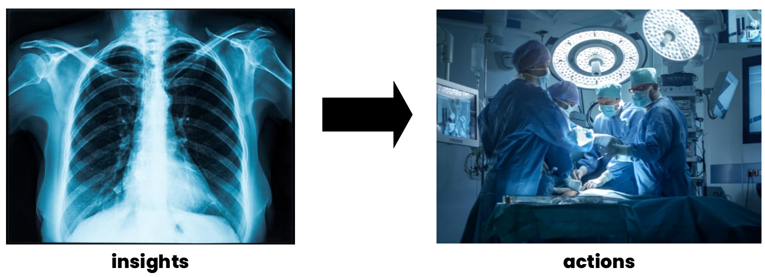

Fig. 4. Turning insights into actions.

As part of scaling process mining, it is essential that insights are turned into concrete improvement actions. This is illustrated in Figure 4. Process discovery and conformance checking can be seen as creating detailed X-ray images to detect problems and find root causes [1]. However, the value of an X-ray image is limited if it is not followed by interventions and treatment, e.g., surgery, chemotherapy, diet, and radiation therapy. Therefore, commercial process mining vendors are combining process mining with automation, e.g., Robotic Process Automation (RPA) and low-code automation platforms like Make.

4 Outlook

How will the process mining discipline and market evolve? Most analysts expect the usage of process mining to grow exponentially in the coming years. Given the growing availability of event data and mature tools, there is no reason to doubt this. To predict the evolution of methods, techniques, and software capabilities, it is good to take another look at the *process mining manifesto* [7] written by the *IEEE Task Force on Process Mining* in 2011. The process mining manifesto lists the following eleven challenges.

- *C1: Finding, Merging, and Cleaning Event Data*

– *C2: Dealing with Complex Event Logs Having Diverse Characteristics*
– *C3: Creating Representative Benchmarks*
– *C4: Dealing with Concept Drift*
– *C5: Improving the Representational Bias Used for Process Discovery*
– *C6: Balancing Between Quality Criteria such as Fitness, Simplicity, Precision, and Generalization*
– *C7: Cross-Organizational Mining*
– *C8: Providing Operational Support*
– *C9: Combining Process Mining With Other Types of Analysis*
– *C10: Improving Usability for Non-Experts*
– *C11: Improving Understandability for Non-Experts*

There has been substantial progress in the areas covered by these challenges posed over a decade ago. For example, we now have comprehensive sets of publicly available benchmarks (C3) and we much better understand the different quality criteria (C6). Thanks to the over 40 commercial process mining tools, it is now much easier to apply process mining (C10) and understand the diagnostics (C11). Due to the many approaches combining process mining and machine learning, there has been major progress with respect to C8 and C9. Nevertheless, most of the challenges are still relevant and even basic problems such as process discovery and conformance checking have not been completely solved.

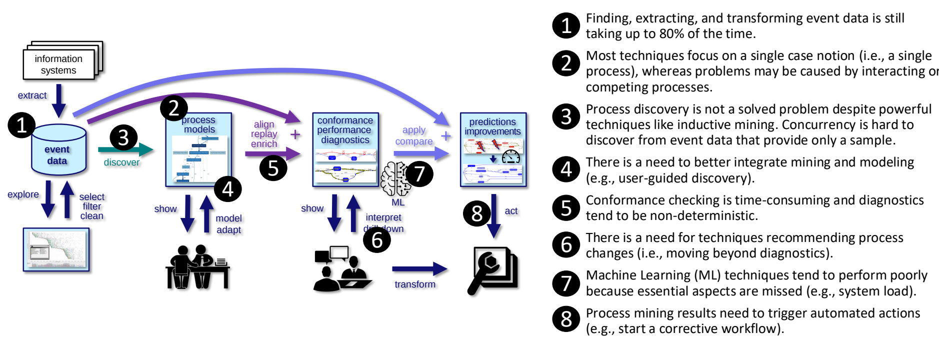

Fig. 5. Process mining challenges in focus in the next five years.

Figure 5 annotates the overview diagram with some of most relevant challenges for the coming years. There is quite some overlap with the eleven challenges in [7]. For example, finding, extracting and transforming input data is still one of the main challenges when applying process mining in practice. Approaches such as object-centric process mining [3, 4] try to make this easier by storing information about multiple objects in a consistent manner and allowing for process models that are not limited to a single case notion. Figure 5 also shows that there are still many open problems when it comes to basic capabilities such as process discovery and conformance checking. Figure 5 also lists challenges that were not discussed in [7]. For example, how to better combine algorithms and domain knowledge to create better process models (*user-guided discovery*) and suggest improvements. There is also an increased emphasis on using process mining results to automatically trigger improvements (*action-oriented process mining*).

We hope that this chapter and book will inspire both academics and practitioners to work on these important challenges. The process mining discipline is rapidly developing and there is still room for original and significant contributions.

Acknowledgments
Funded by the Alexander von Humboldt (AvH) Stiftung and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2023 Internet of Production – 390621612.

References

1. W.M.P. van der Aalst. *Process Mining: Data Science in Action*. Springer-Verlag, Berlin, 2016.
2. W.M.P. van der Aalst. Development of the Process Mining Discipline. In L. Reinkemeyer, editor, *Process Mining in Action: Principles, Use Cases and Outlook*, pages 181–196. Springer-Verlag, Berlin, 2020.
3. W.M.P. van der Aalst. Chapter 1 - Process Mining: A 360 Degrees Overview. In W.M.P. van der Aalst and J. Carmona, editors, *Process Mining Handbook*, volume ?? of *Lecture Notes in Business Information Processing*, pages ??–?? Springer-Verlag, Berlin, 2022.
4. W.M.P. van der Aalst and A. Berti. Discovering Object-Centric Petri Nets. *Fundamenta Informaticae*, 175(1-4):1–40, 2020.
5. J. Carmona, B. van Dongen, A. Solti, and M. Weidlich. *Conformance Checking: Relating Processes and Models*. Springer-Verlag, Berlin, 2018.
6. G. Galic and M. Wolf. *Global Process Mining Survey 2021: Delivering Value with Process Analytics - Adoption and Success Factors of Process Mining*. Deloitte, 2021. `https://www2.deloitte.com/de/de/pages/finance/articles/global-process-mining-survey-2021.html`.
7. IEEE Task Force on Process Mining. Process Mining Manifesto. In F. Daniel, K. Barkaoui, and S. Dustdar, editors, *Business Process Management Workshops*, volume 99 of *Lecture Notes in Business Information Processing*, pages 169–194. Springer-Verlag, Berlin, 2012.
8. L. Reinkemeyer. *Process Mining in Action: Principles, Use Cases and Outlook*. Springer-Verlag, Berlin, 2020.