# Process Mining in a Nutshell

1

Lars Reinkemeyer

**Abstract**

Fundamentals such as event logs, cases, activities, and process variants are explained. Concrete examples show how Process Mining can be used for business transparency and value. Allowing full transparency based on event logs, the implications of this important change—away from perception based towards a fact-based process management—are discussed. The metaphor of an MRT is used to explain possibilities, benefits, and limitations of Process Mining.

"Process Mining is a process management technique that allows for the analysis of business processes based on event logs." This definition by Wikipedia embraces the unique approach of Process Mining to allow the analysis of any business process based on digital traces captured in event logs. An event log is a collection of events which have taken place in order to perform a business process. Each event refers to a specific activity that took place at a certain moment in time and can be assigned to a unique case. An event log consists—as a minimum requirement—of a Case ID as a numeric identifier, an Activity as a specification of which activity has taken place and a Timestamp for the precise time of every action taken (see Fig. 1.1).

Further attributes can be added to provide further information about the specific activities. In a corporate environment, event logs are digital traces which are stored for any business activity in databases such as ERP, CRM, SRM, MES, or PLM systems. Each customer offer, order, invoice, etc., is processed in a database where it

L. Reinkemeyer (✉)
University of California, Santa Barbara, Santa Barbara, CA, USA
e-mail: reinkemeyer@ucsb.edu



L. Reinkemeyer (ed.), *Process Mining in Action*,
https://doi.org/10.1007/978-3-030-40172-6_1

**Fig. 1.1** Event log

| Case ID | Activity | Timestamp |
|---|---|---|
| 1 | Create ticket | January 2, 3:15 PM |
| 1 | Screen ticket | January 2, 3:32 PM |
| 1 | Repair - simple | January 10, 9:45 AM |
| 1 | Close ticket | January 10, 11:34 AM |
| 2 | Create ticket | January 2, 4:04 PM |
| 2 | Screen ticket | January 2, 4:05 PM |
| 2 | Repair - complex | January 3, 1:38 PM |
| 2 | Close ticket | January 4, 9:23 AM |
| 3 | ... | |

leaves digital traces. Detailed descriptions of procedures on how to build valuable event logs from data that is stored in relational databases are publicly available.[1]

In order to "mine" a process, the digital traces are identified, extracted, and visualized in a form which reflects the actual process flow, thus providing transparency regarding the sequence of activities as they have actually taken place. Processing time and sequence of the events provides a complete picture of each case, allowing to trace process flows, understand delays, separate loops, and identify complexity drivers.

Figure 1.2 shows the principle of Process Mining to visualize simple and complex process flows in the form of different variants: the left picture presents a simple, standardized process flow. The right picture presents a more complex process flow with different variants, reflecting multiple options for how the process can be performed.

While people tend to design and think in the form of simple process flows (left image), reality tends to be more complex with multiple variants (right image). In this respect, it is commonly differentiated between "To-Be" and "As-Is" processes:

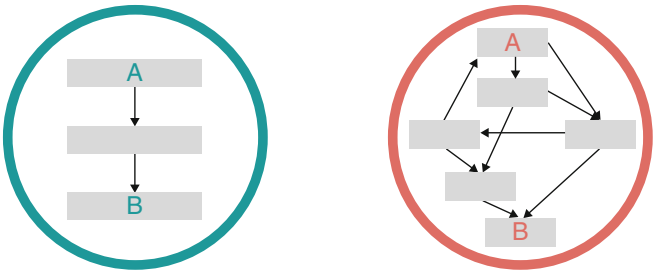


**Fig. 1.2** Principle of process variants

[1]Example: http://businessinformatics.be/2016/12/20/from-relational-database-to-valuable-event-logs-for-process-mining-a-procedure/

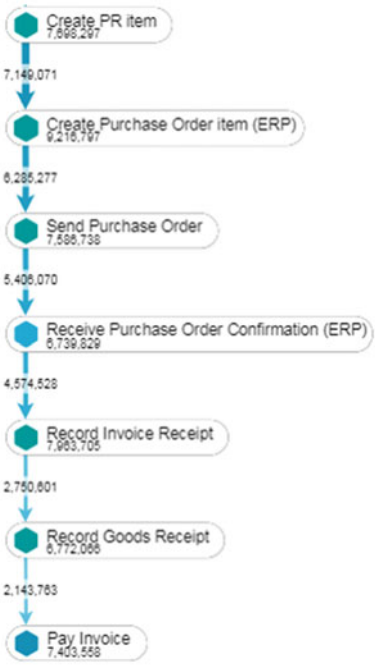


**Fig. 1.3** Most common process variant

To-Be presents the ideal, perfect process flow without friction, as it is typically designed in theory. As-Is on the other hand presents the actual process flow with all deviations and complexities, which occur in real-life operational processes.

Process Mining allows to visualize any process complexity with the option to drill down from the most simple variant to more complex visualizations of multiple process variants until the ultimate visualization of all process variants, thus reflecting a full picture of activities as they were actually performed.

Figure 1.3 shows an example of the most simple = most common variant of a Purchase to Pay (P2P) process, from "Create PR item" (create Purchase Requisition item) to "Pay Invoice." In this most simple variant, the most common activities are shown in their actual sequence. The numbers under the activities indicate how often the activity has occurred (e.g., "Create PR item" occurred 7,698,297 times), while the numbers on the connection arrows indicate how often an activity was directly followed by the subsequent activity (e.g., in 7,149,071 cases the activity "Create PR item" was directly followed by the activity "Create Purchase Order (PO) item"). As the activity "Create PR item" occurred more often than indicated in the connection to the next activity "Create PO Item," it is an indication that there were other subsequent activities. The activity "Create PO item" occurred 9,216,797 times, indicating that many of these activities did not have the activity "Create PR item" as a direct predecessor, but some other activity, which is not shown in this most common variant.

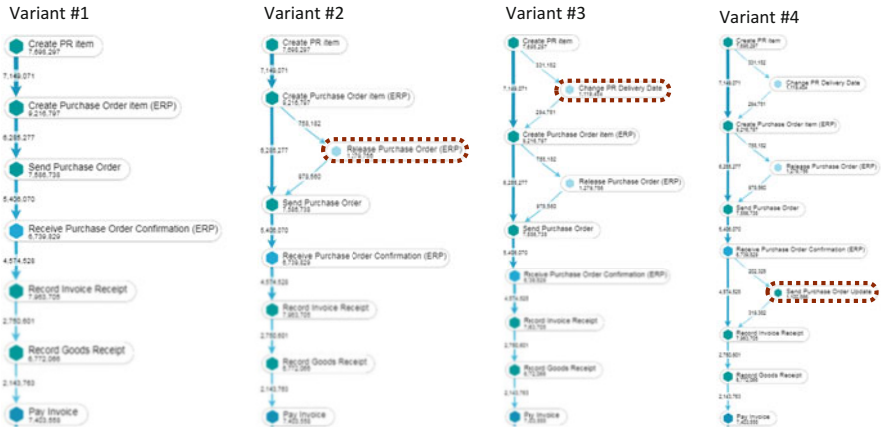


**Fig. 1.4** Process variants, example for a P2P process

Process Mining allows to drill into any process step by step, showing further variants with further activities. Figure 1.4 shows the four most common process variants of a P2P process, each disclosing one additional activity. Variant #1 has been described above. Variant #2 shows the second most common variant, where the additional activity "Release PO" has been logged with 1,279,756 cases. Variant #3 shows the third most common variant, where the additional activity "Change PR Delivery Date" has been logged with 1,119,484 cases. And Variant #4 shows the fourth most common variant were the additional activity "Send PO update" has been logged with 1,100,566 cases.

As an alternative to the number of subsequent activities (as shown on the connection arrows) the time required between two activities can be displayed. Calculated as average or median time, this provides insight into which process steps have long durations and might be the reason for delays or bottlenecks.

*Key Learning #2: Process Mining Allows Full Transparency Based on Event Logs*

Based on these principles, Process Mining allows to drill down step by step into all actual process variants until all cases are included, thus providing an exact picture of the full complexity of the actual processes. Figure 1.5 samples in an excerpt the full complexity of an order to cash (O2C) process, which has several hundred thousand different process variants. The insight allows full transparency to how customer orders are actually processed throughout the organization.

The fun part starts with the surprising insights, which can be gained for specific use cases and which—in almost all cases—reveals numerous surprises even for experienced process experts. The excerpt of O2C gives an idea about the insight into

**Fig. 1.5** Excerpt of process complexity, example of an O2C process

complex global processes, thus allowing the identification and a thorough understanding of operational (in)efficiencies. Based on these insights, dedicated measures for process improvements can be defined and deployed and impact can be continuously measured. Dedicated use cases allow to filter those activities and variants which are of special interest, such as delay, rework, or effort driver. Examples of surprising process variants, which were only discovered with Process Mining, have been presented by multiple companies and include the following samples:

- In procurement, more than ten approval steps for the purchase of a single computer mouse could be identified. The different approvers were not aware how many other colleagues also had to approve the same PO.
- In accounting, more than six activity loops were necessary to clarify single payment blocks due to insufficient documentation and authorization rights.
- In accounting, Process Mining allowed to identify cases, where a daily batch process deleted entries which had been previously keyed in manually by human users.
- In logistics, the reason for late deliveries could be discovered by understanding the end-to-end (e2e) throughput time. Insights allowed to identify that selected freight forwarders didn't operate on the date which was requested for delivery.

Process Mining allows a fundamental change in how to work on process analytics and optimization: a traditional project requires extensive observation and manual documentation of actual process steps. Depending on the invested time and effort, a more or less representative number of observations can be collected and prepared, e.g., in Excel or PowerPoint. Based on extrapolations, some assumptions can be deduced in order to estimate actual process flows. However, this approach has obvious limitations, as it only allows to draw a rough picture of the actual reality which is open to biased interpretation and neglects the technical possibility to get full, fact-based transparency.

With increasing data processing power, access to large amounts of event logs, and performant visualization tools, these limitations have been lifted. Today, insights are possible based on every single activity, thus providing a comprehensive picture of all processes, even with millions of single activities. Visualizations allow aggregated views as well as detailed views, thus providing a concise picture of which can be used throughout the whole organization, by top management to operational employees. As a single source of truth, the provided transparency is consistent across all organizational levels.

Process Mining changes the way work is done, as it allows for unanimous objectivity: while previous process analytics projects had to rely on an excerpt of reality and were open for interpretation, today's insight and transparency are based on the entity of event logs and thus show indisputable, actual process flows. In daily business, this represents an important change, away from a perception-based towards a fact-based discussion. The adoption of this new kind of insights typically follows an evolution: as a first reaction, process experts tend to challenge the results and try to find examples for faults. And experience shows that the forgiving for human error is much higher than the forgiving for computer error. Once the "credibility" has been accepted, the focus shifts towards understanding the results and comparing these with current reports and Key Performance Indicators (KPIs) (Fig. 1.6).

Experience shows that once the Process Mining results are commonly accepted, less time is spent on discussing *what* is happening, but the focus of discussion shifts towards *why* things are happening. The analysis of reasons for delays and loops is the first step towards thorough process understanding and improvement. Based on the in-depth insights, actions are defined, deployed, and continuously tracked.

Process Mining allows to drive process understanding and process optimization deeper into the fabric of the organization, as it provides valuable insights on all organizational levels and enables all players to contribute. In many companies the process responsibility is formally assigned to a central process owner, who is in charge of process design, improvement, and efficiency. But operational experts, who execute process activities with specific domain know-how and "live" the process every day, have a significant impact on process efficiency and must support to drive the change. The head of procurement might be the process owner for the P2P

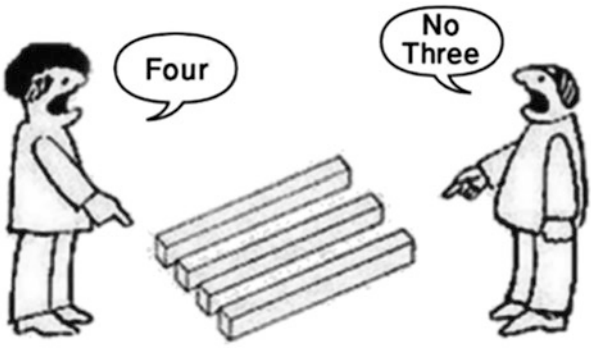


**Fig. 1.6** Perception

process, but the operational procurement team executes activities and can make immediate improvements. As Process Mining provides insights into all organizational levels, it fosters a decentralization of process improvements and a new collaborative approach as process understanding is faster absorbed throughout the organization, with employees on all levels being enabled to contribute to sustainable improvements. And the unbiased data foundation can facilitate a better collaboration across organizational silos, as the EDP use case in Chap. 15 presents. For BMW (use case in Chap. 11), Process Mining has even become a game changer, as it has substituted the traditional process analytics approach.

The key value of process insights is not that it allows to strive for perfect standardization of processes or add more activities and make flows more complex. On the contrary, insights allow to focus on those variants and activities which are effort drivers and of no value. "Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away" (Antoine de Saint-Exupéry). In an operational project, the majority of process variants and activities are usually of little interest, as there is "nothing left to take away," which means that all activities and variants are necessary. Focus must be on identifying the exceptions, which cause delay and additional effort with unnecessary activities and variants. With the new insight the process expert is enabled to drill down and decide which process activities and variants should be avoided or substituted on the strive towards process perfection. Perfection can also be defined as standardization, as the example of Uber in Chap. 10 shows, where Process Mining has been used for global process standardization.

The analogy with X-Ray or Magnetic Resonance Tomography (MRT) pictures can be used to explain the principle of Process Mining, as shown in Fig. 1.7: while MRT as a medical tool allows to screen the human body, Process Mining as a digital tool allows to screen business processes. While providing wholistic pictures of bodies/processes, these provide a quick diagnosis of conspicuous issues, detailed analysis, and targeted remediation. The technology (aka MRT machine/Process

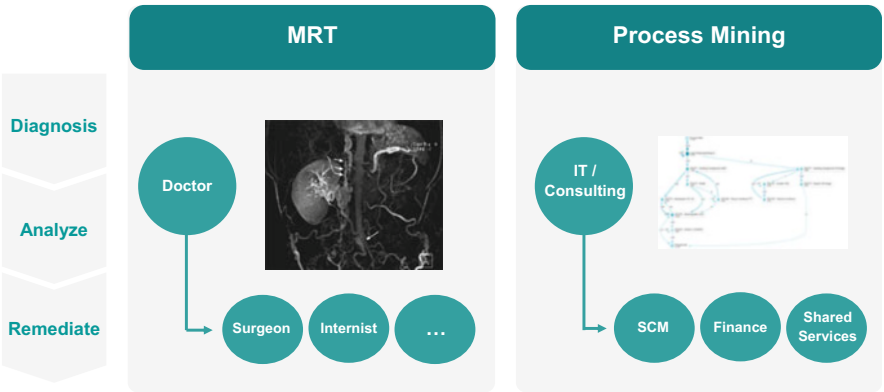


**Fig. 1.7** MRT versus process mining

Mining application) provides insight which can be used by experts (aka medical therapist/process expert or consultant) to support the owner (aka patient/process owner).

Throughout the years, this metaphor has been sharpened with several further aspects:

- *Purpose*: MRT technology allows amazing insights, but it needs to be applied for a predefined Purpose which has to be agreed with the patient. Process Mining allows amazing process insights, but it needs to be properly adapted jointly with the process owner for a predefined Purpose. Use cases are a common vehicle to define Purpose jointly with the process owner prior to applying the tool.
- *Therapy*: MRT makes no sense if the patient is not prepared to undergo a therapy once issues have been identified. Process Mining makes no sense if the process owner is not prepared to adopt the results and set up a project to remediate issues which have been identified.
- *Usability*: Using an MRT machine requires training and education. Using a Process Mining tool requires training and education, i.e., to bridge the gap between technology and business to build meaningful reports which are easy to use and valuable.
- *Investment*: Procurement and operations of the MRT equipment is a significant investment. Procurement and operations of Process Mining software usually also come as a significant investment, unless open-source software is used. In any case, investments for deployment and training of this digital tool should be considered as they typically will account for at least 50% of overall cost.
- *Future readiness*: Investment in MRT technology should be future proof, the system should be ready to be upgraded with technological innovation such as Artificial Intelligence (AI). The Process Mining application provider should provide a clear vision on how the tool will be further developed in respect to innovation, scalability, and keeping it future proof.
- *Vision versus reality*: MRT technology promises exciting use cases with AI, e.g., with the automatic detection of conspicuous features and proposal of appropriate therapies. Similarly Process Mining promises exciting use cases, such as automatic detection of digital traces, AI for proactive user support, and Machine Learning (ML) to optimize processes. However, operational experience shows that many of those promises still must prove their value in the face of domain experts.
- *Compulsory Screening*: Medical prevention requires compulsory screening for prevention, or to identify serious issues at an early stage. Process Mining should similarly be applied as a compulsory, preventative screening in order to assure "hygienic" efficient processes. Given today's data and technology, a process owner should rather explain why Process Mining is *not* used for process optimization instead of discussing the value of process transparency and screening.