

Process Mining for Detailed Process Analysis



Mieke Jans and Manal Laghmouch

1 Introduction

Processes are an important part of a company's daily operations. They are the core to creating value for the end user, both internally and externally. Where organizations are functionally organized around departments such as purchasing, marketing, production, sales, and finance, it is the processes across these departments that ensure smooth operations. For example, the production process will grind to a halt if no goods are purchased, resulting in a purchase need. This information stems from the production planning where a purchase request is registered, but will not just stay within this department. Presumably the purchasing department will take over and place an order with a specific supplier. This is then shared with the warehouse as well as the production department and the finance department. Based on this information, each department will take further action (e.g., approve orders, receive goods, pay invoices). In this example, a clear start and end point can be defined, along with a fixed set of activities performed to achieve a certain goal: registering the need to purchase, registering a purchase order, approving the purchase order, receiving the purchased goods, receiving the invoice, and paying the invoice. This is a typical example of a business process that runs across the various functional departments within an organization, as most processes do.

In general, a process is initiated by a particular need and ends by fulfilling that need. The purchasing process example starts with the need to purchase goods and

M. Jans (✉)

Research group Business Informatics, Hasselt University, Hasselt, Belgium

School of Business and Economics, Maastricht University, Maastricht, Netherlands

e-mail: mieke.jans@uhasselt.be

M. Laghmouch

Research group Business Informatics, Hasselt University, Hasselt, Belgium

e-mail: manal.laghmouch@uhasselt.be

ends with the goods being purchased. To fulfill the need, a set of activities is performed in a logical sequence. A business process is typically defined as follows:

A business process is a set of activities performed in a coordinated manner with a specific business goal in mind.

The defined sequence of activities is part of an organization's set of processes, which are often modelled and portrayed in a process model. A process model depicts how the process (according to the organization) should run.

The remainder of this chapter is structured as follows. In Sect. 2, we discuss process modelling and analysis by describing the modelling languages that are commonly used and introduce the field of Business Process Modelling and Process Mining. Section 3 describes the core principles of process mining and the required input for process mining analyses. Section 4 explains how internal and external auditors can use process mining in practice. In Sect. 5, we conclude this chapter.

2 Process Modelling and Analysis

In this section, we look at fundamental concepts related to process modelling and analysis. We start by explaining the relevance of process modelling and analysis by introducing model-based process analysis. Next, we describe the types of process modelling languages that are commonly used. We conclude this section by introducing a family of data-driven process analyses: process mining.

2.1 *Model-Based Process Analysis: Business Process Management*

Business processes present how things should be handled within an organization. They represent the organization's daily operations and form the basis of improvement opportunities. The discipline that focuses on designing, executing, analyzing, monitoring, and improving business processes is Business Process Management (BPM) (Dumas et al., 2018).

Business Process Management (BPM) focuses on the modelling, implementation, monitoring, and improvement of business processes. It provides a structured methodology with the goal of continuously improving processes: the BPM life cycle, as shown in Fig. 1. The BPM life cycle consists of five activities, after identifying the business processes that are present in the organization. The first activity is *process discovery*. This leads to uncovering the actual process, as performed within the organization. This is classically based on available documentation and interviews. Process discovery within BPM results in an as-is process model. It is important to identify this model to perform the next step: *process analysis*. Process analysis generates insights on possible process improvements. These improvement options

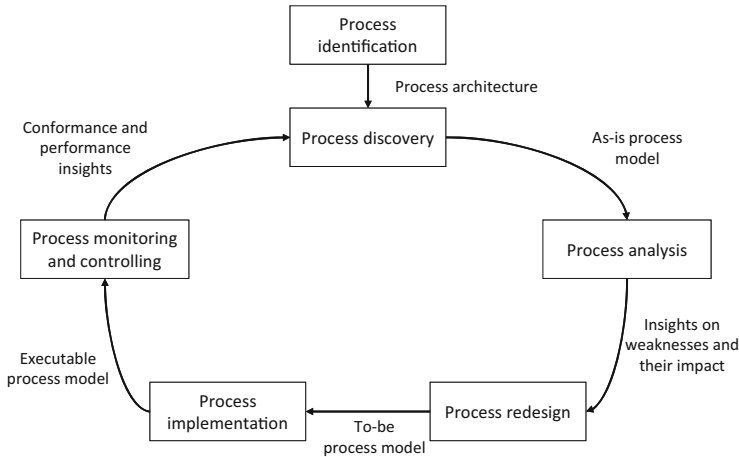


Fig. 1 BPM lifecycle. (Source: Fundamentals of BPM, 2nd ed., 2018, Dumas et al.)

can relate to both increased operational efficiency and better coverage of potential risks. Based on these insights, a *process redesign* follows: the process design is reviewed and adjusted where possible. A new process model is born: we call this the to-be process model. The adjustments associated with the to-be process are implemented in the next step. Both the configuration of the information system and the instructions to the parties involved are adjusted to the new process design. When this new process is put into use, the next step will be *process monitoring and controlling*. This activity will generate new insights on top of the existing documentation. This can be used as input for a new cycle that starts with the mapping of the current as-is process model.

In the traditional interpretation of the BPM lifecycle, there is a striking separation between process models on the one hand and process data on the other. The activities *process discovery*, *analysis*, *redesign*, and *implementation* are often based on process models (in textual or graphical form). In contrast, the *process monitoring and control* activity is often data-driven: key figures of the process are monitored and analyzed. In the purchasing example, this could have referred to the number of open orders and blocked invoices. However, there is no default interaction between the process models and the automatically generated process data. Data on how a process is actually executed is often not taken into account.

Process analyses based on documented process descriptions and interviews are called model-based analysis techniques. Although these techniques provide interesting insights, they have a number of limitations. For example, the quality of the analysis depends on the quality of the available process descriptions and a model-based analysis does not provide valuable insights when the models do not match reality. A mismatch between the model and reality can have several causes. A model is an ideal image of reality or a guiding tool. Consequently, a model is often a simplistic representation of a desired situation in which, unlike in reality, no

exceptions occur. In addition, processes can change unintentionally over time and (rather subjective) discussions with different people involved in the process can result in different models. All these elements call for a new approach to analyze business processes: an objective and realistic approach—process mining. Before getting into the topic of process mining, we discuss two different ways of modelling business processes.

2.2 Process Modelling Languages: Procedural vs. Declarative

A well-organized enterprise models the core processes needed to achieve an organization’s business goals. Process modelling has numerous advantages; it creates a picture of where a company places its emphasis. Furthermore, it provides guidance to actors involved in the process and focuses on the goals to be pursued. The design of well-thought-out processes prevents operational inefficiencies on the one hand and integrates desirable control mechanisms on the other. Through these two elements, correctly followed processes translate into value creation as well as risk reduction (Dumas et al., 2018).

Broadly speaking, there are two different approaches to capturing a process model. A process can be described procedurally or declaratively. A *procedural approach* means that all possible process executions are specified exactly in the model. An example of a procedural process model can be found in Fig. 2. Figure 2 is an example of a procedural process model for a purchasing process (in BPMN). This model specifies that there are only three different process paths for the execution of a purchase:

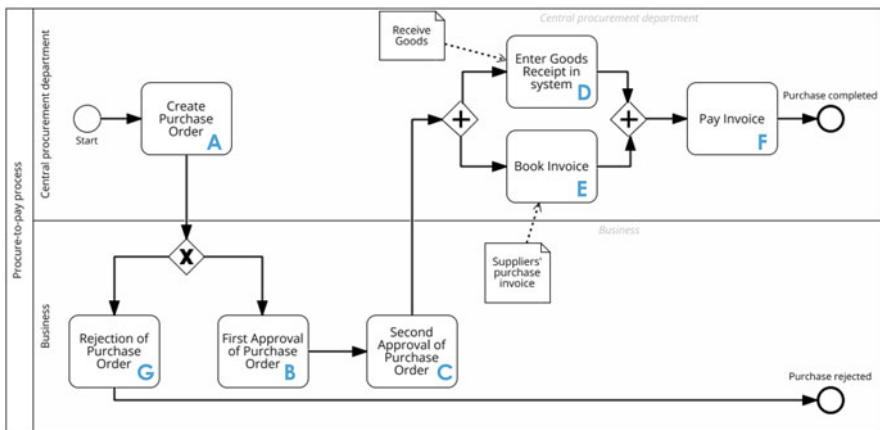


Fig. 2 Example of a procedural process model for a purchasing process

Table 1 Sample list of rules to describe a buying process

If ...	, then ...
“Pay invoice” is present	“Book invoice” is present
“Approve order form—2nd level” is present	“Approve order form—1st level” is present
“Reject order form” is present	“Book invoice” is not present
“Approve order form—1st level”	Executor “Approve purchase order—1st level”
and	≠
“Approve order form—2nd level” are present	Executor “Approve purchase order—2nd level”

- A-B-C-D-E-F
- A-B-C-E-D-F
- A-G

Any other implementation is a violation of the process model. Often, a procedural approach fits well with the modelling of highly structured processes.

Several modelling languages exist within the procedural approach. Although, in the past, flowcharts and EPC models have found their way into business, there are numerous drawbacks associated with these types of models. These drawbacks are mainly about the ambiguous model interpretation and the specific language dependence of software (Dumas et al., 2018). As a solution, a standard was developed for procedural process modelling: Business Process Modelling and Notation. The BPMN standard was created by the Object Management Group (OMG), which is an independent party that develops system-independent standards for computer systems.¹ Process models drawn up according to this standard are easy to interpret. At minimum, a process consists of activities in rectangles, arrows, and additional semantics to indicate relations, like parallelism and choice relationships. For example, Fig. 2 contains a parallelism of activities D and E, indicated by a diamond with a plus, and a choice after activity A, shown as a diamond with an X.

The second way to describe a process is through a *declarative approach*. In a declarative process model, relationships between activities are determined by rules. An example of such a rule is as follows: “the activity *register order* always takes place before the activity *approve order*.” The basic principle of declarative modelling is that a process may be executed in any way, given that certain rules are followed. Rather than capturing the process in fixed paths from start to finish, the total set of rules then defines the process. Opposed to a procedural approach, a declarative approach is recommended for less structured, flexible business processes. Table 1 provides an example of a set of rules that could describe the process from Fig. 2. Depending on how many rules are included in such a set, the process is defined more or less constrained. For example, working with partial (and therefore multiple) deliveries would violate the process model in Fig. 2, but would not violate the rules in Table 1.

¹<https://www.omg.org/bpmn>

A declarative process model is a set of business rules that describe the constraints that a correct process execution should adhere to. A framework that formalizes and standardizes the declarative modelling approach is DECLARE. With DECLARE, a set of formalized constraints using Linear Temporal Logic can be written (Pesic et al., 2007; Van der Aalst et al., 2009).

2.3 Data-Driven Process Analysis: Process Mining

Process mining is an umbrella term for all data-driven process analysis techniques. It brings together the disciplines of data mining and BPM to gain insights into business processes. Process mining allows analyzing a set of data, in particular to better understand operational processes and enterprise activities. The input of a process mining analysis is an event log. Such an event log contains the automatically generated data during the execution of a process. It is comparable to an audit log which is structured in a specific way. This log is used to obtain a realistic representation of the actual process during the *process discovery* activity. Unlike a traditional approach to this activity in the BPM lifecycle, an approach via process mining shows the actual process performed along with the process variants that took place (instead of the desired (normative) process model) (Van der Aalst, 2016).

Process mining techniques that are relevant within the audit can be divided into two groups: “process discovery” and “conformance checking.” Process discovery embraces techniques that discover process models from structured process data. These techniques start from the data stored in the information system during process execution to discover process patterns in this data. The discovered patterns are then visually represented in a process model. This provides an objective representation of the actual process performed and can be used to identify improvement opportunities. Conformance checking goes a step further by testing the conformance of the actual process against the normative process model or against business rules. When conformance is determined by comparing discovered actual process executions to a normative process model, the result is an overview of mismatches between the actual and the normative process model. When conformance consists of checking business rules, the result is an overview of transactions that do not conform to the business rules. Figure 3 visualizes process discovery and conformance checking. You can think of these two types as the core of a process mining analysis.

2.4 Six Phases Within a Process Mining Analysis

While performing a process mining analysis as part of an audit, there are typically six phases that are completed.

Every process analysis starts with *the construction of an event log*. An event log is a specific structured data file that minimally consists of case identifiers, related

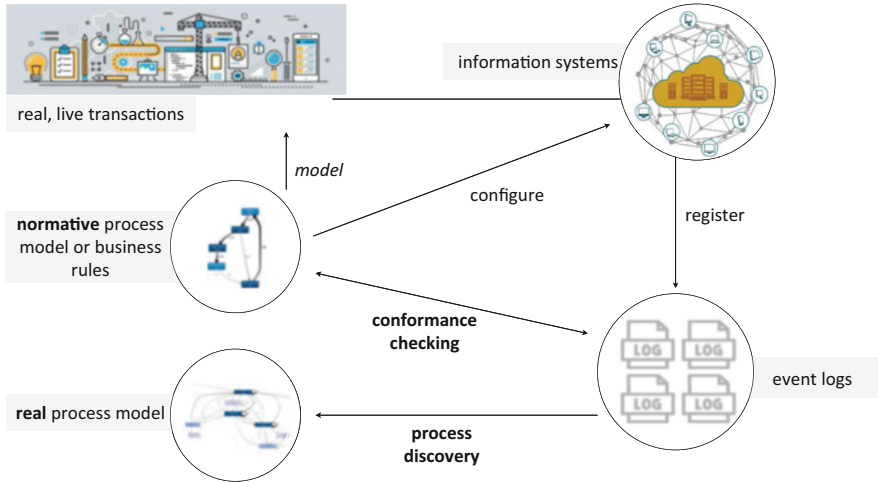


Fig. 3 The core types of process mining

activities, and the timestamps of transactions in a certain process. This will be discussed in more detail in the next section.

When an event log is available, an analyst will apply *process discovery* to discover the actual process in the form of a process model. The output of process discovery can provide initial insight into how structured a process is or is not.

To gain more insights, a *conformance check* is usually performed. This can be done in two ways: either actual process executions are compared to a (procedural) *process model* or to a *list of rules* (declarative approach).

Once there is more insight into how often and where in the process deviations are made from the prescribed process, a *variant analysis* can be performed. A “variant” is a path that is followed for at least one process execution. The variant analysis allows for a closer examination of the different ways in which the process has been followed in reality. For example, an information system may have logged 1000 purchases, 600 of which have followed a similar path (and thus belong to the same variant of the process). Let’s call this path variant A. The remaining 400 purchases follow a different path and belong to variant B. For example, variant A looks like this: <"create purchase order", "approve purchase order—1st level", "approve purchase order—2nd level", "register goods receipt", "book invoice", "pay invoice">. Process executions in variant B, on the other hand, follow a different path of activities: <"create purchase order", "reject purchase order">. Variant A and B represent two ways the process was carried out. In reality, there will be many more variants than stated in the previous example. An examination of the different variants of the process will provide many insights. A variant analysis can, amongst others, answer questions on which process path is most frequently followed and on how many different ways the process was carried out.

Finally, a process mining analysis ends with a *case analysis*. During this analysis, specific characteristics of certain transactions are examined in depth. For example,

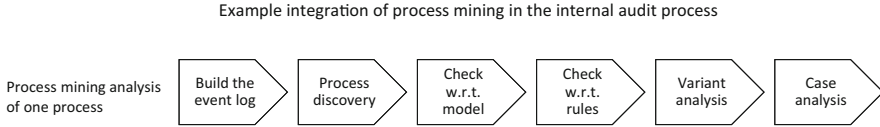


Fig. 4 The phases of a process mining analysis during an audit

one can zoom in on the process performances linked to a certain document type, a certain supplier, a certain period, or a certain level of materiality. This is an in-depth analysis at the level of a subset of transactions.

Figure 4 visually depicts the six elements of a process mining analysis.

3 Requirements and Core Principles of Process Mining

In this section, we will look more closely at what is minimally required to engage in process mining and what the basic principles of process mining algorithms are.

3.1 *The Event Log*

The most important step of a process mining analysis is to collect the right data in the right format, the event log. This section describes what information should be contained in the event log and which structure is required. As briefly mentioned in the introduction, an event log is a structured file that contains all relevant data of a process. In other words, it is a log of events (also called actions) that make up the process and forms the input for a process mining analysis. It is therefore important to know what it takes to build a high-quality event log.

An event log combines data that may come from different information systems in an organization. The raw data from these systems, as automatically stored during process executions, is the starting point of the event log. Often this data is stored in different systems or at least in different tables that are connected via references. Combining the relevant data, selected from thousands of tables, requires a lot of effort, time, and expertise. Think for example of the business processes supported by a SAP[©] or Oracle[©] ERP system. Since such an implementation can consist of tens of thousands of tables, some knowledge is needed to know where to find the right process related data. Identifying relevant event data and converting it to a structured event log is not a task without effort. Therefore, it is important to work with someone who has the right knowledge about the information systems and data in the company to build the event log. Nonetheless, in this chapter we give the basics so that you can assess whether a process analysis based on event data might be possible within a certain organization.

Table 2 Process mining terminology

Concept	Description	Example
Process	A systematic, structured series of actions carried out with the intention of achieving a particular goal.	The delivery process in a postal company.
Case	An instance (execution) of a process.	The execution of the delivery process of goods to Mrs. Lize Kelders who received her goods on April 15, 2021.
Activity	An action to be performed in the process.	The creation of a delivery note in the system.
Activity instance	The specific execution of an activity at a specific time for a specific case.	The creation of delivery note D1001 in the system on April 4, 2021 for delivery 1001.
Event ^a	An atomic part of an activity instance.	The departure of delivery 1001 (the arrival of delivery of 1001 would be a different event; both events are part of the activity instance “delivery 1001”).
Time	The time related to an event.	The 30th of January 2021
Resource	The performer of an event.	Mr. Jan Thomas.

^a In administrative business processes, the events and activity instances are often the same. For example, actions such as “approve,” “register,” or “book” do not have separate start or end moments. Therefore, in this chapter we often use the term “event” to refer to an activity instance, as this increases readability

The construction of the event log depends on the type of questions you are trying to answer through a process mining analysis. For the example purchase process, on the one hand you might want to answer questions about the flow of purchase orders over time. On the other hand, you might also be interested in the flow of purchase invoices to gain insight into the company’s invoicing. Although both questions sound similar, they require different event logs because they look at the process from a different perspective: from the perspective of purchase orders and from the perspective of purchase invoices.² To better understand the content and construction of an event log, it is important to become familiar with terminology used in process mining. Table 2 lists the most important terms related to process mining and an event log with a description of their meaning.

Table 3 shows, for clarification, an excerpt from an event log of a sales process. The given event log consists of the following six columns: “Case ID,” “Event ID,” “Timestamp,” “Activity,” “Resource,” and “Value (in €).” Each row in the event log represents an event and belongs to the execution of a particular case, which in this example is a sales order. The given excerpt shows the events of three cases. Case 1 consists of four events that are already arranged chronologically. Event 51425446101 describes the creation of a sales order (activity) with a value of € 2000 by Jan (resource) on 13 April 2021 at 12:00:00 (time). Linking all events for

²For details on the technical construction of an event log, please refer to the report “From Relational Databases to Valuable Event logs” by Prof. Mieke Jans of Hasselt University in Belgium, freely available on LinkedIn.

Table 3 An excerpt from an event log of a sales process

Case ID (Order ID)	Event ID	Timestamp	Activity	Resource	Value (in €)
1	51425446101	13/04/2021 12:00:00	Register sales order	Jan	2000
	45458455102	14/04/2021 09:23:12	Approve order	Ines	–
	45689454103	19/04/2021 21:34:33	Send goods	Lies	–
	45856655104	20/05/2021 08:23:11	Receive payment	Ahmed	1700
2	45545886291	15/04/2021 17:25:00	Register sales order	Jan	30,000
	77496564292	16/04/2021 01:35:12	Approve order	Ines	–
	85452211297	16/04/2021 22:00:01	Approve order	Peter	–
	85625444301	17/04/2021 07:43:15	Send goods	Lies	25,000
3	87,312,542,221	11/03/2021 09:34:54	Register sales order	Marie	1500
	87,312,542,222	26/03/2021 09:37:11	Receive payment	Ahmed	1500

case 1 results in one specific execution of the sales process, from the point of view of the sales order. The process started for case 1 on April 13, 2021 with the creation of an order by Jan and ended with the receipt of payment on May 20, 2021. The sequence of the four listed activities in this specific order reflects one process variant. In this excerpt, this variant is not repeated. However, it might emerge later in the event log that this variant is the most frequent variant of all the process executions.

At a minimum, an event log contains information about the case IDs, activities, and related times (columns 1, 3, and 4).³ Based on these minimum requirements, process mining is able to represent the real flow of actions over time. In addition, an event log can contain additional information about events, such as the resource and value in this example. In a process mining context, these properties are called attributes. You can add as many attributes as desired to the event log. Note that the more attributes you add, the larger (broader) the event log becomes. It is therefore recommended to only include those attributes that add value to your process analysis. As a consequence, it is important to, as a preparatory step, unambiguously identify business questions that the process analysis should answer.

³ A standard has been developed for event logs, XES (<https://xes-standard.org>). This format is system and software independent and is used by most process mining tools on the market as a log format. Often there is also the possibility to load the log as a csv file and the tool itself converts this into an xes file.

To make an initial assessment of whether or not a process mining analysis is possible, it is important to ask the following questions:

- Is it possible to designate a case to follow through the process?
Note that a “purchase” or “sale” is not a suitable candidate, as this is usually not stored as a separate entity in the information system. Underlying documents such as a “purchase order” or an “invoice” may be more appropriate.
- Is it possible to link the activities that make up the process execution of the case to the chosen case?

Only if these two questions can be answered positively, you can further consider a process mining analysis to answer your business questions. For more information on this specific topic, we refer the interested reader to Jans (2019) and Jans et al. (2019).

3.2 *Process Discovery*

After the event log is built, process mining analyses can be performed. As already mentioned, process discovery is the first analysis that is performed. It aims to represent the process as it was actually performed within an enterprise. An event log of one specific process is the input for such process discovery analysis. Based on the event log, process discovery can then discover a set of process models that together reflect the actual business process. In what follows, the mechanism behind process discovery algorithms is explained, as well as the possible outputs.

3.2.1 **The Mechanism Behind Process Discovery**

As described in Sect. 2.1, an event log consists of at least three elements: a case ID, an activity, and the time an event was recorded. These three data points are necessary to visually represent the flow of a process. Figure 5 visually represents the mechanism behind process discovery in a simplified way. On the left side of the figure, an event log consisting of three cases is represented. The cases in the log go through a number of activities. To explain the process discovery mechanism as simply as possible, the figure abstracts from the time when the activities occurred. For this example, it may be assumed that the activities are arranged in chronological order. A process discovery algorithm starts by identifying the path of each individual case. For example, case 1 follows the path $\langle A, B, C, D, D, E \rangle$. Case 2 follows a different path, which is $\langle A, B, C, D, E \rangle$ and the path of case 3 looks like this: $\langle A, C, B, D, E \rangle$. Finally, the algorithm combines the paths and learns patterns that can be visualized. Each of the previous three paths starts with activity A. Then, activities B and C follow activity A. In two cases, B follows first and then C. In the other case, it is the other way around. It is inferred that the order of these two activities is of secondary importance. After the execution of B and C, activity D takes place, either twice or not. The process ends in all observed paths with activity E. The combination

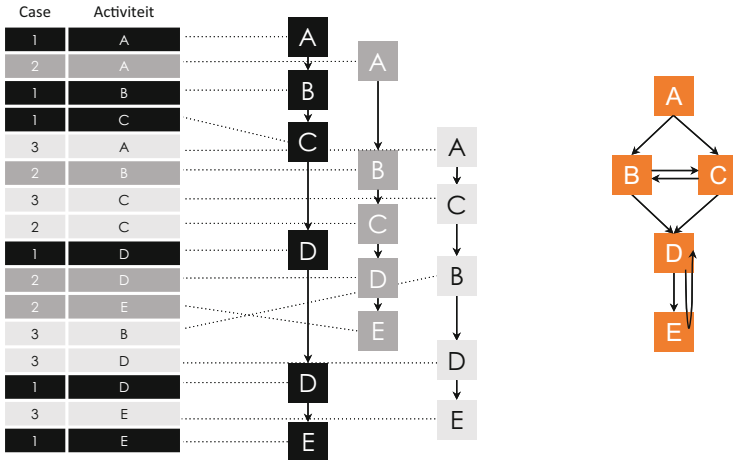


Fig. 5 A simplistic illustration of the mechanism behind process discovery. (Source: Process Mining Book at <https://fluxicon.com>)

of these discovered patterns results in a process model that reflects the process that is actually followed. Figure 5 shows this process model in orange.

3.2.2 Levels of Abstraction

Process discovery outputs a set of process models that together represent the behavior captured in an event log. Although a set of process models are the output of process discovery, they are often presented “on top of each other” in one process view (as in Fig. 5). The end user can determine the level of abstraction of the given process model. You can compare this principle with a dynamic road map that can zoom in or out. To get a general idea of how a route network is structured, an overview of the most frequently used roads, usually motorways, is sufficient. An abstraction is then made of other roads that do exist and are used, but less intensively. Following this analogy, a high level of abstraction is also desirable in an initial introduction to the process. Getting an understanding of how the process works in most cases is sufficient. If the end user is interested in more details, then a low level of abstraction better suits the analyst’s needs. In our road map analogy, local roads, and possibly even bike lanes are included in the map.

Figure 6 shows two levels of abstraction of the same process, departing from the same event log. The model on the left is a more high-level representation of the process: it is more abstract than the model on the right. Depending on the purpose of the process mining analysis, one level of abstraction fits better than the other. Throughout the analysis, the abstraction level can be changed by zooming in and out on the process (process mining software allows for easy variation in abstraction level).

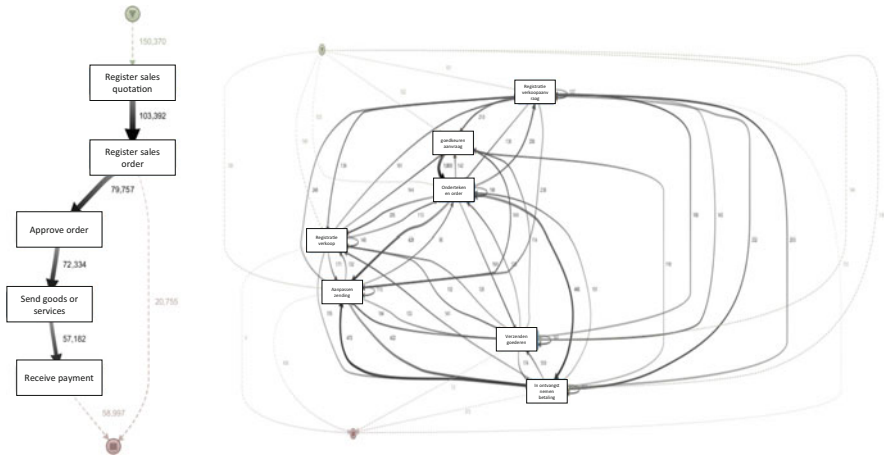


Fig. 6 Two different levels of abstraction of process discovery output

3.2.3 Output

The most commonly used visualization in process mining software is the “process map.” It shows the activities in rectangles and connects these with arrows if one activity (in the event log) is followed by another activity. The more often this relationship is observed, the thicker the arrow. This is called a “directed graph.” Although the modelling language is very intuitive, it consists of ambiguous relationships. Take for example the process as discovered in Fig. 5. It is not clear whether after activity A both activity B and C follow, or whether only B or C is sufficient, or whether many repetitions of B and C must follow. The core of the problem with this process representation lies in not being able to represent parallelism and choices unambiguously. On top of that, there may be combinations of arrows in the model that are not actually present in the event log. Given these shortcomings, an output according to the BPMN standard is a good addition.⁴

3.3 Conformance Checking

Whereas process discovery provides insights about the actual processes within an organization, conformance checking can identify where the actual process matches or deviates from prescribed procedures or business rules. Conformance checking compares actual process behavior (as recorded in the event log) with procedures, either in the form of a process model or business rules. Through this comparison, process deviations are identified. Identified process deviations can result from two

⁴More and more software packages are providing this functionality.

causes: on the one hand, they can be exceptional cases that require a different approach than the standard process execution. On the other hand, a process deviation can be the result of errors or fraud. To determine deviations, there are two possibilities, as mentioned before: a test of the actual behavior (as contained in the event log) against the normative process model or against business rules. Both possibilities are briefly explained.

An event log and a normative process model are required to perform a conformance check against a model. There are several approaches to technically perform a conformance check, but we limit ourselves here to a description of the underlying principle. For a check against the model, each case in the log is played back on the process model representing the desired process. For each case, it is determined whether or not it conforms to the model. A case that deviates from the model is a case whose activities do not run through the model completely or incorrectly. The amount of detail given as output depends on the technique used.

There are naive and advanced techniques to perform a conformance check. The naive techniques only show which cases deviate from the model. Advanced techniques go a step further by providing additional information about where exactly things go wrong and why that step is identified as a process deviation. Thus, the output of a conformance check against a normative model is a list of deviating cases or a list of more detailed process deviations. To illustrate, take a case—order 201—in which the activity “*send invoice*” is missing. A naive check will indicate that order 201 is not conforming to the model, while an advanced check will indicate that there was no invoice sent to the customer for order 201.

In addition to a check against a normative process model, the actual process behavior from the event log can be compared with business rules. For this, the business rules must be converted to a formal language.⁵ The set of business rules forms the declarative process model. The event log is then tested against the set of rules. If a case from the event log violates a business rule, then that case does not conform to the declarative model. The advantage of a check over rules is that it is known exactly why a case deviates: a case is not compliant with the process because rule X and Y were violated. Furthermore, the analyst can establish the rules that are of principle interest to check. By its nature, this approach leans close to the work of a financial auditor.

4 Process Mining in the Audit

The insights flowing from a process mining analysis form a good basis for improving business processes in terms of efficiency and risk. It gives a view on the level of control an organization has over its operations. Given the auditor’s responsibility to understand a client’s environment when performing a risk assessment, process

⁵The declarative language LTL is appropriate for this purpose.

mining can be a good support (Jans et al. (2013), Jans et al. (2014)). In what follows, we discuss how process mining can support the audit.

Since the added value of a data-driven process analysis is broadly applicable, process mining is often implemented by the organization itself as part of the internal audit (Chiu and Jans, 2019). The internal auditor has more resources and time to perform a comprehensive analysis, possibly even on a continuous basis, than the external auditor. Moreover, for the external auditor the investment of a process mining analysis is relatively large compared to a total audit engagement. We will therefore first discuss how process mining can be incorporated within the internal audit. However, the principles are the same for the external auditor. Finally, we expand on the use of the process mining implementation by the external auditor.

4.1 Process Mining and the Internal Audit

An internal audit is an indispensable element for checking the internal organization of a company. The internal auditor systematically examines whether the working methods and business processes of the company are efficient and under control. This usually consists of five steps. The internal auditor starts by drawing up a multi-year audit plan. Afterwards, the internal auditor plans the process audit, carries out the audit and communicates the results. Finally, the results are followed up. The most comprehensive implementation of process mining is one that is woven into all of these steps. In what follows, we describe what this might look like. This includes an abstraction of the size of the investment required to accomplish this.

During the *planning the audit schedule* for the next few years, an event log could be created of each process in the company. These logs can be analyzed via process discovery to get a global overview of how structured the processes are. To do this, the different logs are examined with a fixed set of parameters of the algorithm (such as the level of abstraction). A process model that looks very orderly will likely represent a process with less risk than a process model with many more activities and arrows. In addition to the visual aspect, a number of metrics can be listed and calculated to determine the schedule. Examples could be: the number of cases, the number of different variants per 100 cases, and the number of variants needed to cover 80% of the event log. All of this information together can assist the auditor in creating the audit schedule.

While *planning the process audit*, the process discovery analysis from the previous phase can be repeated for the selected process. In this phase, the discovered process can be examined in more detail. Where in the previous step only the structure of the process was looked at, attention is now paid to the logic shown in the discovered process model. Does this process broadly correspond to the expectations? What paths and activities emerge when we lower the level of abstraction? Without seeing much detail, an arrow between “Create order form” and “Pay invoice” can already be an indication to investigate further in a later stage.

In addition to process discovery, an initial conformance check can be performed against the procedural model. This provides an initial, general view of the anomalies that occur. Are there possible explanations for variants that occur frequently in the event log, but that conflict with the model? Anomalous variants that occur frequently but for which the auditor can formulate an explanation are, for example, variants that can be expected in the occurrence of partial deliveries. Whether these are actually partial deliveries will need to be verified later during the audit. All these aspects are included in the plan of the audit to be performed. It is a different scenario if the auditor sees a deviating variant for which no explanation can be formulated. This too will be included in the audit plan, but perhaps with more investigative actions.

After planning, the *audit* will be *conducted* and the event log will be examined in more detail. The triggers for further analysis that were uncovered in the previous phase will now be included. Indeed, the deviations from the process model must be cleared out: are they alarming deviations or are they logical process variations? To do this, the auditor will iteratively develop explanations for the identified deviations and then verify these explanations (Jans and Hosseinpour, 2019). In the example of partial deliveries, the auditor may approve the variants that include multiple deliveries, noting the assumption of partial deliveries. Separate analyses can verify that controls such as the 3-way match have worked effectively. Another anomalous example might be the absence of a goods receipt. One possible explanation for this is that it includes services, not goods. To clear this up, the auditor can take all the variants (and their cases) in which a goods receipt is missing and then test whether these cases were services. The cases for which this was the case will be “cleared,” while for the other cases another explanation should be sought. Perhaps there are certain suppliers for whom other agreements apply or where the delivery takes place at a different place and this is not recorded in the system? For each explanation, the auditor will try to clarify the discrepancies based on the data. This will require a combination of variant analysis, case analysis, and rule checking. If no explanation can be given for certain deviations, the case will be included on a list of potential anomalies.

During the *communication of the results*, visual support for process mining analyses will play a particularly powerful role. The different phases of the internal audit and how process mining can support it are visually summarized in Fig. 7.

4.2 *Process Mining Interaction Internal and External Audit*

In addition to providing support for the internal auditor, a process mining analysis can add value for the independent external auditor. By analogy with Fig. 7, Fig. 8 shows how process mining can support the external auditor. Here, it may or may not be possible to build on what the internal auditor provides.

By using process discovery to visually represent the actual business processes, along with a first comparison of the log data against the desired process model, the

Internal audit

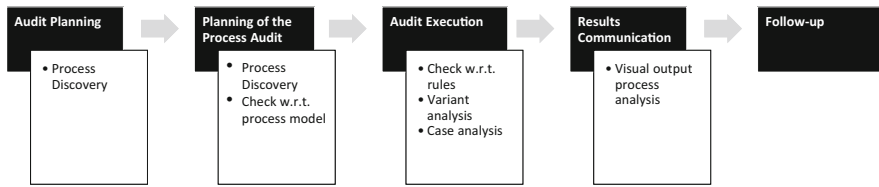


Fig. 7 Integrating process mining into internal audit

External audit

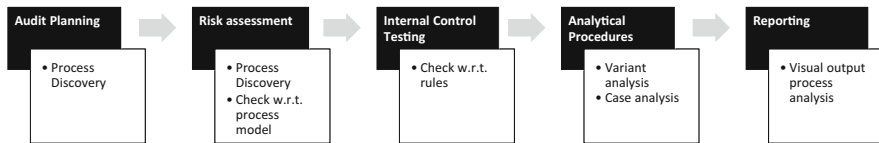


Fig. 8 Integrating process mining into external audit

auditor can obtain an initial overview of the business processes. This can serve as support for the planning phase and risk assessment work.

Consistent with conducting the internal audit, the external auditor will address exposed nonconformities. Given the external auditor’s focus on financial reporting, a different emphasis may be placed in the deviations to be examined. For example, repeated approvals of the same voucher will generate interest from an efficiency standpoint, but perhaps not from an audit standpoint. Despite a potentially different selection of deviations, the approach to clarify them is similar to what has already been described for the internal auditor. This will require a combination of a review of process executions against business rules, variant analysis and case analysis. Rule testing is well suited as a control test. Indeed, each control mechanism can be formulated as a rule: “if..., then....” For example, “if a receipt is created, then it is approved later.” Variant and case analysis are used to answer more targeted questions and lean closely towards substantive controls.⁶ Examples include reviewing transactions of a specific person, process executions in which manual activities have taken place and activities outside of working hours.

As with the internal audit, communication will take place, supported by the visual output of the analyses. An important aspect in this is that the findings are based on objective data and that they are easily transferable if the right graphics are used. Figure 8 summarizes the audit process supported by process mining.

If the auditor wishes to expand on the findings of the internal auditor, he or she must, as with any other audit, build in a number of checks regarding the quality and

⁶However, the split between control testing and substantive testing is no longer strictly applicable if the full data set is used to verify the operation of a control.

completeness of the information provided. In the context of process mining, there are the following specific points of interest:

- If the event log is provided, the auditor should check the underlying script and
 1. verify that no errors have been made
 2. check and take into account the underlying assumptions and filters used
- What type of systems were consulted to build the event log? Are these systems well managed in terms of access and control? Can information from these systems be relied upon?
- If discovered process models or anomalies are provided: what algorithm was used with what settings (which parameters are used)? Is there a script to replicate (and check) this analysis? Which normative model or set of rules was tested?

4.3 Practical Applications and Available Software Tools

While research in the field of process mining continues to increase, there is an increased adoption of process mining in companies. The Task Force on Process Mining (<https://tf-pm.org>) aims to promote process mining, publish scientific research on it, establish standards, and organize workshops. Furthermore, the Task Force keeps up-to-date information on developments within the process mining domain and publishes event logs, introductory and other videos, and case studies from the industry. For an up-to-date overview of practical process mining applications and software tools, we refer you to the website of the Task Force on Process Mining, where the aforementioned information is available under the tab “resources.”⁷

5 Conclusion

Business processes are at the core of a well-functioning organization. They reflect how information should flow and what actions should be taken to achieve business goals. Because processes reflect the functioning of a company, including the implemented control mechanisms, they are a good starting point for gaining an insight into the environment in which financial reporting occurs.

The discipline of Business Process Management (BPM) focuses on managing and improving business processes within an organization. It is often partially adopted during an audit. Traditionally, BPM starts from an analysis technique that mainly relies on interviews and consulting existing process documentation. Based on the

⁷Although there is not an exhaustive list of tool vendors, many have their own introduction video on this site.

(derived) process models, insights are gained about how the company manages its processes and whether it is in control.

Although the traditional BPM approach can provide valuable insights into an organization's processes, it is limited to analyzing prescribed procedures in the form of normative process models. Normative process models do not describe the actual processes within an organization but rather propose an ideal image, a procedure that should be followed. As a result, the quality of the process analysis depends on the quality of the models and the extent to which the model matches reality. This is because the actual process executions often contain situations that are not included in the prescribed processes. Depending on how much these exceptions occur, there is a small or a large(r) mismatch between reality and the process models that form the basis of analyses.

To ensure that process analysis leads to correct insights, process mining can be applied. Process mining is a collective name for all data-driven process analysis techniques that start from an event log. It combines the strengths of the BPM approach with data analysis techniques to gain insights into the actual business processes. Process mining allows us to analyze the entirety of recorded activities to understand the business processes better. More specifically, process mining techniques provide insights into the ordering of activities, the timing of activities, and the actors involved in the actual process.

Every process mining analysis starts with the collection of data. Data from one or more sources are combined to build an event log. An event log contains data about one specific business process and is therefore used as the input for process mining analyses. A process mining analysis for one process usually includes the following six steps: (1) building an event log, (2) process discovery, (3) a conformance check against a process model, (3) a conformance check against a set of rules, (5) a variance analysis, and (6) a case analysis.

Process discovery and conformance checking are the two main types of process mining that are relevant to auditing. Based on an event log, a process discovery analysis can reveal the actual business process. This actual process is then usually visually presented in a process model. Conformance checking goes a step further by comparing the process behavior from the event log with a normative process. This normative process can either take the form of a procedural model or a set of business rules. Plotting the recorded activities against the norm leads to the identification of process deviations. Process deviations can either indicate exceptional cases, or errors and fraud. Filtering out the second group currently remains a challenge, as with all data analysis approaches in the context of an audit.

After performing a process discovery and conformance check, a variant or case analysis may be of interest. A variant analysis is an examination of the different ways in which the actual process was executed. A case analysis takes a close look at a specific subset of transactions by analyzing certain characteristics in depth. An example are the transactions performed on a particular day or by a particular department or person.

The insights generated from a process mining analysis provide a sound basis for improving business processes in terms of efficiency and risk. This broad view of

processes ensures that the insights generated are relevant to both the internal and external auditor. This chapter discussed how the various process mining analyses can support both auditors, as well as the concerns of the external auditor if he or she wishes to elaborate on the analyses of the internal auditor.

References

- Chiu, & Jans. (2019). Process mining of event logs: A case study evaluating internal control effectiveness. *Accounting Horizons*, 33(3), 141–156. <https://doi.org/10.2308/acch-52458>
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. (2018). *Fundamentals of BPM* (2nd ed.). Springer.
- Jans, M. (2019). Auditor choices during event log building for process mining. *Journal of Emerging Technologies in Accounting*, 16(2), 59–67.
- Jans, M., & Hosseinpour, M. (2019). How active learning and process mining can act as Continuous Auditing catalyst. *International Journal of Accounting Information Systems*, 32, 44–58.
- Jans, M., Alles, M., & Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems*, 14(1), 1–20.
- Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review*, 89(5), 1751–1773.
- Jans, M., Soffer, P., & Jouck, T. (2019). Building a valuable event log for process mining: An experimental exploration of a guided process. *Enterprise Information Systems*, 13(5), 601–630.
- Pesic, M., Schonenberg, H., & Van der Aalst, W. M. (2007, October). Declare: Full support for loosely-structured processes. In *11th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2007)* (p. 287). IEEE.
- Van Der Aalst, W. M. (2016). *Process mining: Data science in action*. Springer.
- Van der Aalst, W. M., Pesic, M., & Schonenberg, H. (2009). Declarative workflows: Balancing between flexibility and support. *Computer Science-Research and Development*, 23(2), 99–113.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

