

# Process Mining: Basic Terms



## INTRODUCTION

Process Mining technology extracts knowledge from event logs commonly available in today’s information systems, and provides a new means to discover, monitor, and improve business processes.

It can provide analysts with detailed information about the current “as-is” state and past performance of processes, as well as help improve and support business processes in competitive and rapidly changing environments.

This Glossary hopes to serve as a guide for consultants, business managers, and end-users and explains some of the most common terms used in Process Mining and their meaning.

Please note that this is not a definite list. We would love to hear your comments and suggestions of additions. Please write them to [marketing@minit.io](mailto:marketing@minit.io).

## Table of content

<b>Introduction</b>	<b>1</b>
<b>Context: How Process Mining works?</b>	<b>2</b>
<b>Core Process Mining Terms</b>	<b>3</b>
Activity	3
Active time	3
Attribute	3
Case	3
Case ID	3
Data Attribute/Attribute	4
Data refresh	4
Edge	5
Event	5
Event Log	5
Filtering	5
Financial Analysis	6
Frequency Analysis	6
Log	6
Performance Analysis	6
Process Animation	6
Process Clone	6
Process Compare/Comparison	7
Project Hub	7
Process Instance (PI)	7
Process Map	7
Process Mining	7

<a href="#">Process Discovery</a>	7
<a href="#">Process Statistics</a>	7
<a href="#">Process Variant</a>	8
<a href="#">Resource</a>	8
<a href="#">Social Chart</a>	8
<a href="#">Timestamp</a>	8
<a href="#">Views</a>	8
<a href="#">Waiting Time</a>	8
<b><a href="#">Related Terms</a></b>	<b>9</b>
<a href="#">Benchmarking</a>	9
<a href="#">Business Analysis</a>	9
<a href="#">Business Analyst</a>	9
<a href="#">Business Architecture</a>	9
<a href="#">Business Case</a>	9
<a href="#">Business Goal</a>	9
<a href="#">Business Need</a>	9
<a href="#">Business Objective</a>	10
<a href="#">Business Problem</a>	10
<a href="#">Business Process</a>	10
<a href="#">Business Process Re-engineering</a>	10
<a href="#">Business Process Management (BPM)</a>	10
<a href="#">Business Process Modelling Notation (BPMN)</a>	10
<a href="#">Data Mining</a>	10
<a href="#">Decision Analysis</a>	11
<a href="#">Domain</a>	11
<a href="#">Elicitation</a>	11
<a href="#">Enterprise Architecture</a>	11
<a href="#">Metadata</a>	11
<a href="#">Process</a>	11
<a href="#">Process Model</a>	11
<a href="#">Repository</a>	12
<a href="#">Root Cause</a>	12
<a href="#">Root Cause Analysis</a>	12
<a href="#">Stakeholder</a>	12

## Context: How Process Mining works?

An automated IT system facilitates the processing of invoices using a predefined process in which activities and their ordering is defined.

The activities executed in this system are recorded in an event log. The log contains records of actual executions of events on cases on a certain moment in time by a certain process participant. So, an (event) log contains information about process instances (e.g. cases) and the events that are performed on/for them.

Process Mining software such as Minit takes this event log and automatically pieces together the 'as-is' business process. Process Mining algorithms combine events, analyze their chronological order, correlations, frequency, and performance. Then the software visually demonstrates each process as an interactive process map containing all relevant information down to single case level. The software uncovers all the nuances of the process, statistical information, variants, process exceptions, unusual transactions, bottlenecks, deviations, and potential risks in processes.

# Core Process Mining Terms

## Activity

Activity is an action or task that can be performed for a process instance. The activities are the steps, or status changes, that happen in your process such as Order Created, Order Approved, Missing Information etc. For the purpose of analysis, it is necessary that the event log contains all activities that describe the relevant business milestones in the process. If the log contains activities that are not crucial (such as debug information) these can be filtered out by the process mining software. Activity is one of the three minimum data requirements necessary for Process Mining.

## Active time

Active Time is time spent on performing activities.

## Attribute

= Data Attribute

## Case

Case (i.e., a process instance) is one instance, one execution, of your process. For example, if you are looking at an Order to Cash process then handling of one Order is one case. A case can have one or more activities.

## Case ID

Case ID is a specific ID of one Case that connects the individual steps of a process instance from the beginning to the end, for example, a customer number, order number or patient ID. For example, in an Invoice Approval Process a Case ID may be the Invoice Number. It is one of the three minimum data requirements to make your event log viable for Process Mining. Every event in your process log must refer to one specific case, so that Process Mining software can compare several executions of the process to one another.

There may be more than one way to set up your case ID and the set-up of Case ID influences the process scope and determines where your process starts and where it ends.

For example, in a Customer Support process you could set up the case ID in two different ways. You could focus on analyzing the processing of a particular customer request, then customer request ticket number is the case ID. Or you may want to see the overall process for a customer as the process scope—the same customer may have gone through multiple support requests over time. Then the customer ID will be the Case ID.

## Data Attribute/Attribute

Data Attribute is any attribute recorded in the event log. Examples of attributes are: the amount of a purchase order, the name of the supplier, the name of the employee performing the activity, the patient's age etc. Activities, Case ID and Time Stamps are also attributes, but to avoid confusion they are mostly referred to by their respective names. These attributes are used in different types of analysis to provide context and important information that is necessary to answer the questions that you have about your process. For example, financial attributes are necessary for financial analysis. E.g. to measure the cost of unnecessary rework, add the costs per resource into your event log with working hour rates of your employees as an additional attribute.

By looking at the data attributes that are available, you may also get further ideas for your analysis. We recommend including all attributes that you find relevant because they can improve the value of your analysis.

Attributes can be used for filtering, to focus or split your data set, to compare the processes for different productions, channels, regions, etc. And they provide context in root cause analysis - when you discover interesting patterns in your process such as when delays are occurring mostly with one particular supplier, or in one region, or when one team does not follow the prescribed process.

If you include attributes that can change over time (like the organizational role of a person), get the right attribute value from the time that the activity was performed. For example, you want to know the role the person had when they did something in the process rather than the role that they have now.

To decide which additional attributes you should include beyond the minimum requirements (which are Case ID, Activity, Time Stamp), it is important to know the goals of your process analysis. Relevance of attributes depends on your domain and use case.

### Some typical additional attributes are:

- **Country, Branch** if you are analysing data from a large company that operates in different locations and want to compare processes in different branches or countries.
- **Product** if you want to compare performance of different product categories for example in a sales process.
- **Channel** through which a lead came in is often relevant for sales processes. In repair services new requests may come in through the dealer, the call center, or the web portal.
- **Partner/Affiliate** if you want to take a deeper look at the performance of different partners or affiliates.
- **Resource** – Person or Department that handled the activity is necessary for performance analysis, if you want to optimize the distribution of work between teams or eliminate inefficiencies at the hand-over points between organizational units.

## Data refresh

The purpose of data refresh is to be able to easily refresh the underlying data of a process analysis with new data. For example, if your process analysis started by importing a CSV file of Invoice Approval event log for the year 2016 and your customer hands over event log with an enhanced dataset for the year 2016 and first half of 2017, you might need to refresh the whole analysis that you already performed with the new data.

Within Process Mining software Minit Data refresh functionality does not depend on the type of data source, you can refresh data from file, SQL Server or ODBC driver.

## Edge

Edge/Path is directed line connecting two activities in a process map. Edge is created based on the chronology of activities in a case and represents successor/predecessor relationship between activities.

## Event

Event is a specific occurrence of an activity in time and in a specific process instance performed by a resource. Events having the same value of attribute Activity result into single Activity.

## Event Log

Event Log is a recording of a set of events. To be able to apply process mining techniques it is essential to extract event logs from data sources (e.g., databases, transaction logs, audit trails, etc.). In an event log, each event refers to a case, an activity, and a point in time. An event log can be seen as a collection of cases and a case can be seen as a trace/sequence of events.

Within a typical event log each row corresponds to an event. There are at least three columns:

- Case ID (patient id, order number, claim number etc.)
- Activity (approve, reject, request, send etc.)
- Timestamp (2015-08-18T06:36:40, etc.)

There may be other additional columns (attributes): resource, transaction type, age, costs, etc.

Event data may come from a wide variety of sources:

- a database system (e.g. patient data in a hospital)
- a comma-separated values (CSV) file or spreadsheet
- a transaction log (e.g. a trading system)
- a business suite/ERP system (SAP, Oracle, etc.)
- a message log (e.g. from IBM middleware)
- an open API providing data from websites or social media
- CRM, Service Management, BPM, LOB, Call Centres, and more.

The standard format for event logs supported by the majority of Process Mining tools is [XES \(eXtensible Event Stream\)](#). XES was adopted in 2010 by the [IEEE Task Force on Process Mining](#) as the standard format for logging events. Next to XES other target formats supported by Process Mining tools are [MXML \(Mining eXtensible Markup Language\)](#) and [CSV files](#). Minit Process Mining additionally supports import from SQL Server, Excel, Access, ODBC.

## Filtering

Filtering is the way to adjust the visualization of the dataset, for example, if the dataset contains data that do not fit into the concept of your analysis and contains process instances from the testing stage of the process. A filter is a rule or a set of rules determining which cases and/or events from the process will be later included in the process analysis. Filters are applied to the entire cases in case of case-level filters or they can modify cases by including/excluding events in case of event-level filters.

The overall result of the filtering is cases and/or events that match all the active filters. If necessary, the same filter type can be used multiple times (such as a filter for attributes). However, for some filters, their multiple uses do not make sense, or it could lead to an empty result set (such as combining several variant or timeframe filters).

## Financial Analysis

Financial Analysis focuses on the process from the point of view of the finance attributes – the flow of the money within the process. For example, you can find out the total costs/incomes about process activity, average values, costs/incomes about the case etc.

**Finance attributes** can be defined on three levels:

- **Per Case** – set for attributes holding the costs/revenues of your cases such as invoice amount, purchase order value, loan amount, etc.
- **Per Event** – in case your attribute holds fixed costs/revenues of specific events/activities such plumbing installation, outsourced marketing research, etc.
- **Per Resource** – for attributes specifying staff-hour rates of your employees or system resource performance related costs such as Cloud services paid at hourly rates.

## Frequency Analysis

Frequency Analysis visualizes the process (map) from the frequency point of view e.g. the number of times an activity occurred within the process or the number of its repetitions in a single case.

## Log

Log refers to the original log generated by the source system which records things that have happened. In order to be used within Process Mining software this needs to be converted to one of the supported formats.

## Performance Analysis

Performance Analysis visualizes the process (map) from the point of view of performance measures such as execution (active) time, waiting time, etc.

## Process Animation

Animation is used to visualize the development of the process over time. Animation is always shown over the displayed process map or social graph. The instructions displayed on the Process map screen equally apply to process map and social chart settings.

Within Process Mining software Minit animation settings are found in the left part of the screen under the Animation heading or the button.

## Process Clone

Process clone is a special variant of data refresh functionality with the only difference, that the result of it is not refreshed existing process and views, but a new process imported with the new data using the same import configuration. This feature allows the user to easily compare the newly created process clone with the original process, e.g. data from year 2016 with data from 2017 in the same structure. Available within Process Mining software Minit.

## Process Compare/Comparison

Process Compare allows an intuitive and efficient comparison of processes. You can compare the processes at the level of the process map, where the generated visualization allows you to identify the differences in the flow and frequency, performance or finance metrics. You can also compare in detail values of metrics and attributes at the level of activities. You can compare the views of various processes or different views generated from a single process (e.g. for different periods of time). You can also compare mined process to a modelled process (process blueprint) in a form of BPMN diagram to check the compliance.

## Project Hub

Project Hub is the main and starting screen of Process Mining software Minit. It gives you an overview of projects, imported processes and created process views and it allows you to manage them.

## Process Instance (PI)

= Case

## Process Map

Process map represents the behaviour of the process captured in data by means of activities. The displayed process map always corresponds with the mining result in the process, reflects filter settings above the data and chosen frequency or performance metrics. When filters are applied, mining is performed again, metrics are recalculated, and a new process map is generated.

## Process Mining

Process Mining is analyzing a business process based on an event log. Process Mining techniques and software extract information from event logs. From the audit trails of a workflow management system or from the transaction logs of an enterprise resource planning system a Process Mining tool can discover and monitor the process, support analysis, optimization and improvement.

Process Mining is closely related to BAM (Business Activity Monitoring), BOM (Business Operations Management), BPI (Business Process Intelligence), and data/workflow mining. Unlike classical data mining techniques, the focus is on processes and questions that transcend the simple performance-related queries supported by tools such as Business Objects, Cognos BI, and Hyperion.

## Process Discovery

Process Discovery is the creation of process model from event log using process mining algorithms.

## Process Statistics

Process Statistics provides a brief and easy-to-understand overview of the properties of the process undergoing analysis and contains the information about the process through summary charts and statistics.

## Process Variant

Process Variant is a unique path from the very beginning to the very end of the process. Different Process Variants show different variations in the process flow – specific sequences of activities as they happen during the process execution. By understanding the variants in the process, you can find out which patterns deliver a good (or bad) performance.

The frequency of a variant shows how frequent a specific execution pattern is and helps distinguish main stream variants from outliers and exceptions. By understanding the mainstream variants, it is possible to improve and enhance the normal process. By optimizing the process to reduce the exceptional variants, you can deliver a more consistent performance. Promote the well-performing variants for a better and more consistent process performance.

## Resource

Resource is any actor that can execute an activity, for example humans, the system itself or a web service. The Resource – Person or Department that handled the activity is necessary for performance analysis, if you want to optimize the distribution of work between teams or eliminate inefficiencies at the hand-over points between organizational units.

## Social Chart

Social chart represents the behavior of the process captured in data by means of resources. The displayed model always corresponds with the mining result in the process, reflects filter settings above the data and chosen frequency or performance metrics. When filters are applied, mining is performed again, metrics are recalculated, and a new social chart is generated.

## Timestamp

Timestamp is a time indication consisting of a date and possibly a time part. One or more timestamps per step (for example for the beginning and the end of an X-ray examination) are used to calculate the process sequence and to derive parallel process steps.

Timestamps are the third minimum requirement for Process Mining. At least one timestamp for each activity is necessary to be able to bring the events for each case in the right order. If you want to analyse activity durations, you need both a start and a completion timestamp for each activity.

It is possible to analyze a process even if Timestamps are missing from the log – provided all events are in chronological order. In such a case performance analysis is not be possible.

## Views

A view is the representation of a single process you can work with on your own or in cooperation with your colleagues within Process Mining software Minit. View stores all your settings for the analysis of the process. If you want to analyse one process from several viewpoints, just create several views.

## Waiting Time

Waiting time is time spent between activities (on edges).

# Related Terms

## Benchmarking

Benchmarking is a comparison of a decision, a process, a service, or a system's cost, time, quality, or other metrics to those of leading peers to identify opportunities for improvement; used to review products and/or services that are being delivered by your competitors.

## Business Analysis

Business Analysis is the practice of enabling change in the context of an enterprise by defining needs and recommending solutions that deliver value to stakeholders.

## Business Analyst

Business Analyst is any person who performs business analysis, no matter their job title or organizational role; a person who helps companies determine their objectives for meeting certain opportunities or addressing specific needs and then helps that company determine solutions and ways to meet those needs.

## Business Architecture

Business Architecture is the design, structure, and behaviour of the current and future states of an enterprise to provide a common understanding of the organization. It is used to align the enterprise's strategic objectives and tactical demands.

## Business Case

Business Case is the justification for a course of action based on the benefits to be realized by using the proposed solution, as compared to the cost, effort, and other considerations to acquire and live with that solution.

## Business Goal

Business Goal is a state or condition that an organization is seeking to establish and maintain, usually expressed qualitatively rather than quantitatively.

## Business Need

Business Need is a problem or opportunity of strategic or tactical importance to be addressed.

## Business Objective

Business Objective is an objective, measurable result to indicate that a business goal has been achieved. SMART objectives are specific and measurable results the business wants delivered by the project; long-term operational or strategic goals of the project.

## Business Problem

Business Problem is an issue of strategic or tactical importance preventing an enterprise or organization from achieving its goals.

## Business Process

Business Process is an end-to-end set of activities which collectively responds to an event and transforms information, materials, and other resources into outputs that deliver value directly to the customers of the process. It may be internal to an organization, or it may span several organizations.

## Business Process Re-engineering

Business Process Re-engineering is rethinking and redesigning business processes to generate improvements in performance measures.

Business Requirement is representation of goals, objectives, and outcomes that describes why a change has been initiated and how success will be assessed.

## Business Process Management (BPM)

BPM is a management discipline that determines how manual and automated processes are created, modified, cancelled, and governed.

## Business Process Modelling Notation (BPMN)

BPMN is graphical representation for specifying business processes in a workflow.

## Data Mining

Data Mining is statistical analysis technique to discover patterns and relationships in large volumes of data. Can be descriptive, diagnostic, and predictive. Predicting future actions using data analysis.

## Decision Analysis

Decision Analysis is an approach to decision-making that examines and models the possible consequences of different decisions and assists in making an optimal decision under conditions of uncertainty.

## Domain

Domain is the sphere of knowledge that defines a set of common requirements, terminology, and functionality for any program or initiative solving a problem.

## Elicitation

Elicitation is an iterative derivation and extraction of information from stakeholders or other sources. A more intentional set of tasks than simple requirements gathering. Receiving of information from stakeholders or other resources. The main path to discovering requirements and design information.

## Enterprise Architecture

Enterprise Architecture is a description of the business processes, information technology, people, operations, information, and projects of an enterprise and the relationships between them.

## Metadata

Metadata is a description of data to help understand how to use that data, either in terms of the structure and specification of the data, or the description of a specific instance of an object.

## Process

Process is a business activity that transforms data; it has a distinct beginning and end; a process may be high-level or detailed; an activity performed by the business that transforms information (data); a set of activities designed to accomplish a specific objective by taking one or more defined inputs and turning them into defined outputs.

## Process Model

Process Model is a set of diagrams and supporting information about a process and factors that could influence the process. Some process models are used to simulate the performance of the process.

## Repository

Repository is a real or virtual facility where all information on a specific topic is stored and is available for retrieval.

## Root Cause

Root Cause is the cause of a problem having no deeper cause, usually one of several possible causes.

## Root Cause Analysis

Root Cause Analysis is a structured examination of an identified problem to understand the underlying causes.

## Stakeholder

Stakeholder is group or individual with a relationship to the change, the need, or the solution.