Process Mining
Let data describe your process

Process mining is a growing discipline which leverages from the omnipresence of data for fact-driven analysis of processes. Process mining allows organizations to diagnose process bottlenecks and vulnerabilities based on reality rather than assumptions about the reality.

Elham Ramezani Ph.D., ir. Marijn Nagelkerke and drs. Paul Kromhout

Elham Ramezani Taghiabadi PhD
is a senior consultant at KPMG Risk Consulting.
Ramezani.Elham@kpmg.nl

Ir. Marijn Nagelkerke RE
is manager at KPMG Risk Consulting.
Nagelkerke.Marijn@kpmg.nl

Drs. Paul Kromhout RE CISA
is a director at KPMG Risk Consulting.
Kromhout.Paul@kpmg.nl

Introduction

The purpose of companies and organizations is creating value that satisfies their stakeholders. And it is within business processes where this value is actually created. If companies or organizations are not successful in creating value in a reasonable time, then they will not stay in business for a long time.

A sustainable growth of companies and organizations will require periodic changes or enhancements in their processes and continuous monitoring of their performance. However, this is easier said than done. Sometimes improvement projects do not deliver meaningful improvements to companies and no matter how close a company monitors its key performance indicators (KPIs), the overall impact on the business operation remains limited.

Often, when defining an improvement project, the leadership team will meet, put together their ideas for an improvement initiative, and then kick off the initiative. Accordingly, KPIs are defined and monitored. However, many companies put the emphasis on poorly selected improvement points. Consequently, when changes are made, the results achieved are disappointing or not as satisfying as expected. So, if this is not the right way to go about things, then what should companies actually be doing?

In this article, we explain how we can use process mining to answer the above mentioned question. First, we would like to give an overview of different types of analysis and their application that can be applied using process mining technology. We briefly discuss different process mining techniques and focus on two domains in which process mining could be leveraged. For each domain, we show a case from an engagement where we used process mining. Next, we will discuss a methodology to conduct a process mining project. Finally, we conclude by discussing some lessons learned and success factors in a process mining project.

What is process mining?

Process mining provides new ways to utilize the abundance of information about events that occur in the world

surrounding us. These events such as 'open door', 'approve loan', or 'create order' can be collected from the underlying information systems supporting a business process or sensors of a machine that performs an operation or a combination of both. We refer to this as 'event data'. Event data enable new forms of analysis, facilitating process improvement and process compliance. Process mining provides a novel set of tools to discover the real process, to detect deviations from the desired process, and to analyze bottlenecks and waste. Process mining is generic and can be leveraged to improve processes in a variety of application domains. It can be applied for various processes such as purchase-to-pay, order-to-cash, hire-to-retire, and IT management processes. It can be leveraged in any industry sector and has already been applied in various sectors, such as banking, consumer products, healthcare, insurance, professional services, public sector, and logistic services.

Process mining bridges the gap between traditional process analysis (e.g. simulation and other business process management techniques such as lean management and six sigma) and data-centric analysis techniques such as machine learning and data mining.

An input for all process mining techniques is event data which records the information about the execution of business processes. Process mining techniques can be categorized under four main activities: *process discovery, conformance checking, enhancement* and *process analytics.*

1. *Process discovery* uses logged event data from the executions of business processes and produces a model reflecting the actual behavior recorded in the logged data without using any information about the process. The model reflecting the actual process is a descriptive model, not to be used to steer and control the actual process, but rather to aim at capturing reality. Event data usually contains a unique identifier (called case ID). The case ID is selected according to the perspective of an analysis. Examples of case IDs are: an order number, a customer number, a ticket ID, a patient number, etc. In addition, event data should contain the recorded execution of activities and their timestamp. Case ID, activity name and timestamp are considered to be the minimum requirements for process discovery. Process mining is less feasible when an IT system does not record this information. Event data can contain additional information such as the person who executed an activity or the quantity of an order, or the age of a patient. Adding additional information allows for deeper and more precise analysis of a process, such as describing how different resources in a process interact with each other.

2. *Conformance checking.* By using conformance checking, the actual process (which can be based on a model obtained via process discovery) can be compared with event data and used to show deviations or alternative process paths.

3. *Process enhancement* includes extending or improving an existing process model using information about the actual process recorded in event data. For example, by extending a process model with performance information related to time or cost, or repairing the process model according to actual executions of the process recorded in the event data.

4. *Process analytics.* In addition to the above three process mining activities, other analysis techniques can be applied in the context of event data and process models, such as data mining techniques or visual analytics (e.g. histograms of events per case), of which the results can be used to understand process models with additional aspects, predict future (e.g., the remaining flow time and probability of success) and recommending suitable actions.

Applications of Process Mining

Process mining can be applied for both process improvement projects and audit engagements.

A business process is a chain of activities that are connected to each other to achieve an objective. There is often a significant gap between the officially documented process flow and what actually happens in these processes. Using process mining, we can visualize the actual process execution, and can detect inefficiencies, bottlenecks, and deviations from a desired process. After detecting and prioritizing improvement areas, improvement initiatives can be taken. Process mining can be used to perform a fact-driven analysis of the process and detect improvement areas.

Process mining can be used in a continuous manner. After resolving one bottleneck, the focus shifts to resolving the next inefficiency. Process mining should not be seen as a one-time project, rather a continuous process improvement technique. Therefore, we encourage a sustainable approach in which an expert in-house team continuously monitors processes and improves them at a gradual pace.

Process mining is not only used for post-mortem analysis of processes. It can also be used to predict future executions of a process (e.g. the remaining flow time and probability of success) and recommending suitable actions.

Process mining integrated in the audit

Process mining can also be used during different stages of the audit:

1. during walkthroughs;
2. used as a basis for sampling;
3. used for compliance checking.

Process mining used during walkthroughs

Do you go through time consuming interviews and meetings with clients to understand their processes? There is a better way! One of the first steps in any audit engagement, is understanding the client's business processes via walkthroughs. However, this approach is not very efficient for giving a comprehensive picture about how processes are actually executed. In addition, one cannot be sure whether the picture obtained is complete. Often, different departments at the client only have knowledge about the parts of the process that pass through their department and auditors need to contact different people to get a relatively complete understanding of the entire process. Furthermore, during walkthroughs, only very typical ways of performing a process are discussed and in many cases exceptions are not captured. Using process mining, this step can be shortened and the quality of walkthroughs can be drastically improved. Using process mining, an auditor can see the overview of business processes executed in reality together with all the details and possible exceptions.

Process mining used for sampling

Process models discovered during walkthroughs can be used for better and smarter sampling. That is, cases that have higher operational risk can be detected and used for further analysis.

Process mining used for compliance checking

Using process mining, in particular conformance checking, the deviations from a desired process path or a compliance rule can be detected. The entire population of historic data can be checked instead of limited analysis done on sample data (100% sample data).

Case: Improving business processes and perform a post implementation process review

This use case describes the use of process mining at a financial services company that implemented a new contract management system (CMS) in three countries. CMS facilitates the financial process from quotation through to termination. This system is workflow driven and provides event logging in separate tables. The two most relevant datasets we used were the Contract Lifecycle and Credit Lifecycle. Both data sets hold the actual event logging at a contract level and credit limit level. Both datasets have been analyzed using process mining, however this use case is limited to the Contract Lifecycle.

The dataset contained all contract information such as contract ID, activities that were performed, the resources who performed these activities, status of the contracts, and the country code. 14,500 contracts were analyzed with over 90,000 events divided over three countries. For this analysis, process discovery was used at the product level per country. The process flows were different between the three countries.

Figure 1 illustrates the discovered process models related to the different countries. As can be seen in Figure 1 [B], most contracts follow the same process flow (the thick black line). In country A (shown in Figure 1 [A]), the 'initialized' event in the flow is almost equally split between the next two events 'completed' and 'sent'. For about 1790 cases these two events in the process never actually happened. Interestingly one of the these events ('completed') is skipped for the majority of the contracts for the third country (Figure 1 [C]).

Similarly, other differences between the three countries were discussed. The company did not expect such results. Note that, the company implemented a standard back office system in three countries with the intention to standardize the processes for equal products (except for some small localization). However, our analysis revealed that different countries did not use the system in the same way. In addition, almost all the countries used the system differently from the designed processes.

The company further used the provided insights to standardize the processes between the three countries. Consequently, our root-cause analysis revealed that detected differences were caused by insufficient training that resulted in copying the activities as performed in the old system to the new system. The system confi-
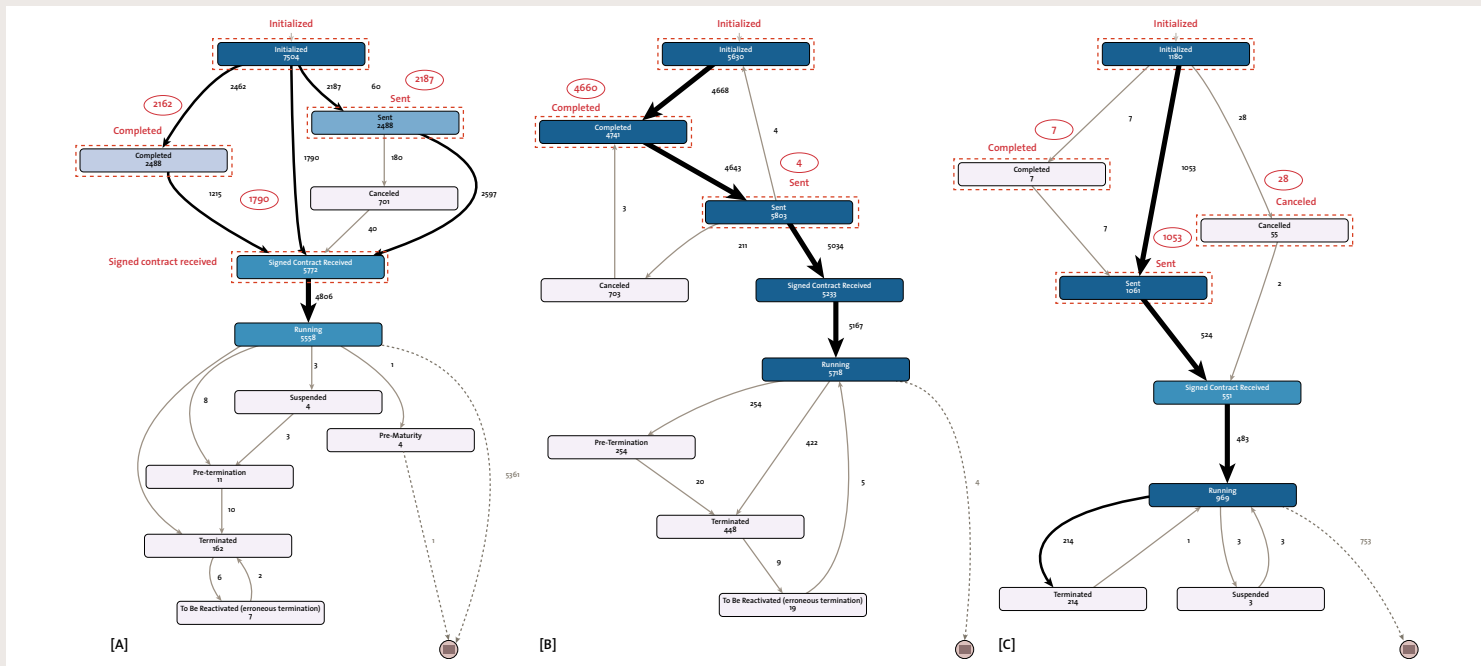
Figure 1. Process mining analysis revealed that a single process is performed differently in three different regions of the same company.

Figure 2. Detailed performance information related to a deviating process path.

Contract Life Cycle	Country 1	Country 2	Country 3
From Initialized to Running	0%	1.5%	2.4%
From Running to Terminated *	0%	0.4%	0.1%

Table 1. **Violations frequency per country related to breaches of four eyes principle for two pairs of activities (Initialized, Running) and (Running, Terminated).**

guration (process flow and authorization set up) did not prevent the user from doing this either. Figure 2 shows detailed information related to a deviating process path. As can be seen this process path have been executed for 16 contracts. On average, the time spent on this process path is about 13.7 weeks.

Detecting deviations between the desired process flow and the actual process flow triggered us to perform an in-depth risk assessment. As a result of this analysis, the deficiencies in the implementation of the internal

control framework were detected. As part of testing internal controls, process mining was used to test the four-eyes principle between successive events. Table 1 shows the frequency of breaches of four eyes principle per country for two pair of activities: (Initialized, Running) and (Running, Terminated).

Using process mining, all these violations of the four-eyes principle were identified. The violating contracts were then filtered to be analyzed further. Diagnostics were provided on the contracts that had been violated. In addition to the provided diagnostics, the system configuration was re-designed to align the configuration with the desired process models. The authorization set up was re-defined to ensure compliant internal controls and this provided concrete action points to improve the process executions. The company gave positive feedback on the following items:
• they could see and understand the actual business process on contract level in an intuitive and easy visualization;

- they learned how different countries execute a single process;
- they received insightful diagnostics on the deviation of the actual business process from the desired process;
- they learned about the actual and potential risks in their business processes and their negative consequences;
- they received concrete action points and well-selected improvement opportunities to reduce costs and improve the quality of their business processes.

The company successfully implemented proposed modifications in the processes and the underlying information systems.

Case: Analysis of a change management process as part of general IT controls (GITCs)

This use case describes the use of process mining as part of a financial statement audit of an international corporate client. For this audit, the auditor verified the system transportation process of one of the SAP systems used. Usually, when a change in a system is requested, a change ticket is issued. If the change request is approved, it triggers a transportation process. During the transportation process, a change to the system will be prepared in the development environment, then the transport will be imported to the quality environment. Finally, if the transport passes all required checks, it will be promoted to the production system. One of the change logs that the auditor analyzed contained about 577 transports. Several activities have been performed for each of these transports. Figure 3 shows the difference between the number of activities that were executed for different transports. Some transports were processed with only a few activities, and for some more than 3000 events were logged (each event represents one execution of an activity). The question that was raised was why certain transports include so many events while some do not. The auditor further analyzed these occurrences and noted for example that several changes have been transported to multiple target systems and consequently some events are duplicated for those transports.

The figure below shows the complete process model discovered for these 577 transports. As is shown in Figure 4, the discovered process model is rather complex and
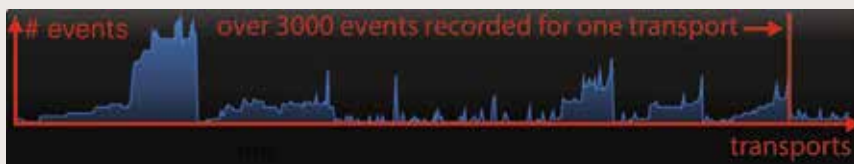

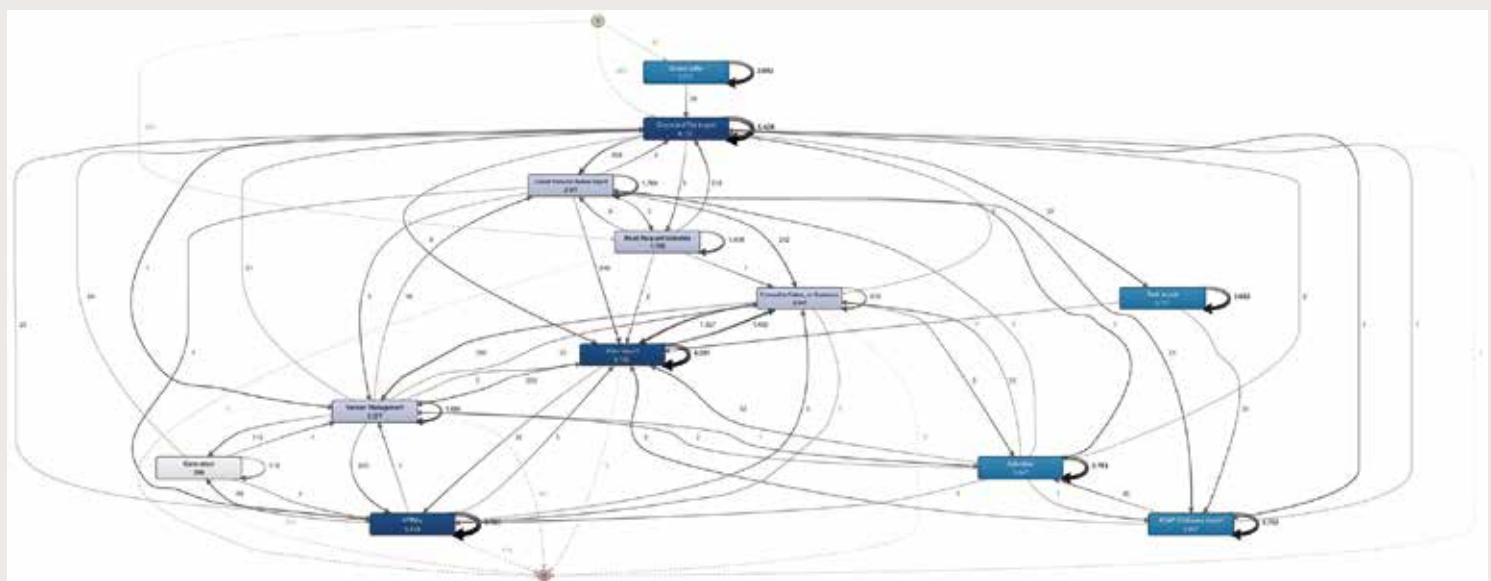Figure 3. **Number of events recorded for each transport.**

Figure 4. **The complete process model discovered from the process followed by 577 transports.**
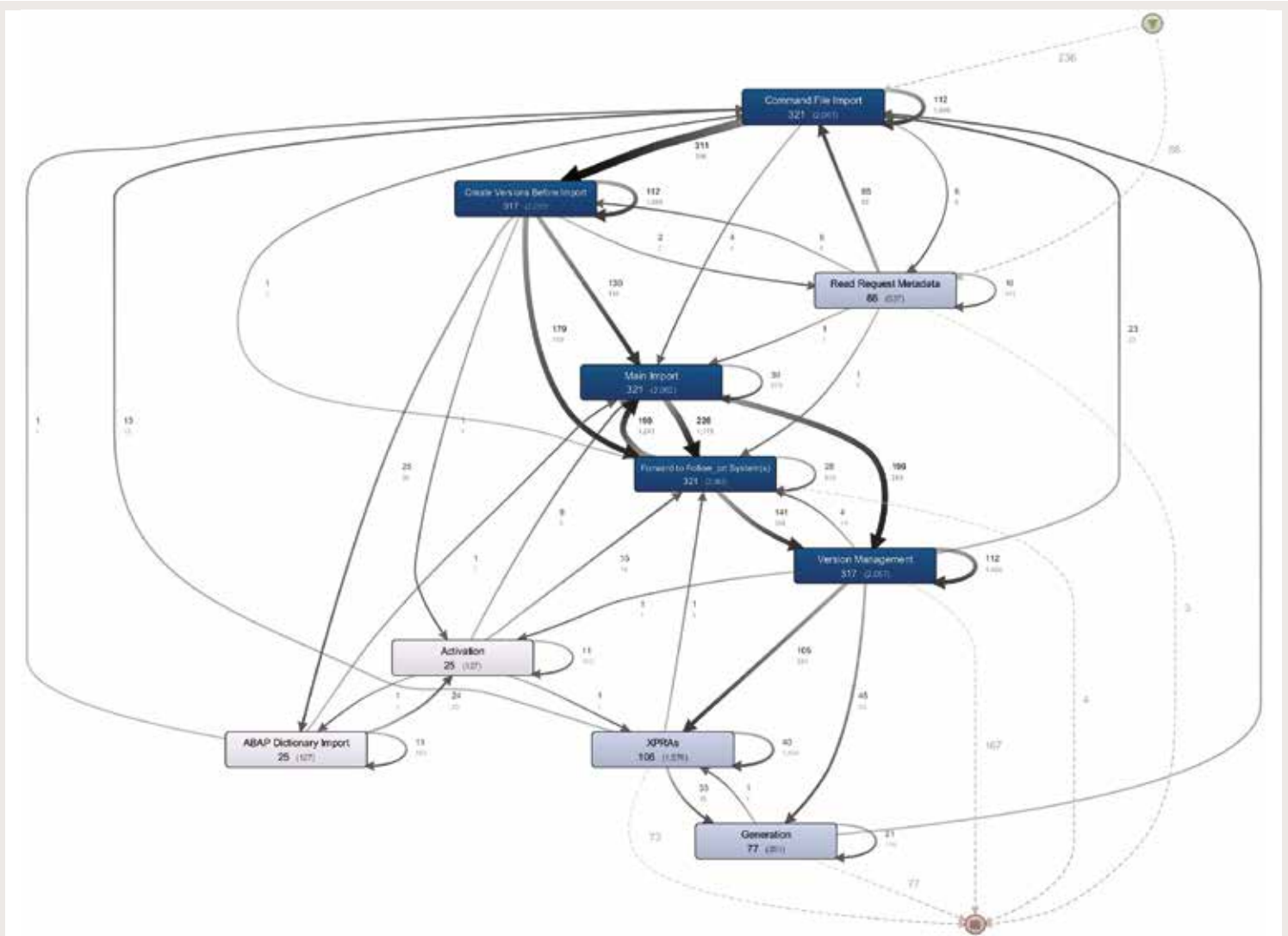
Figure 5. **The transportation process discovered for transports that originate from system 1.**

reveals a so called 'spaghetti model' which is difficult to understand. The auditor noted that various process flows have been followed, based on the source system of the transports. In this use case we highlight the process flows of two of these source systems.

Figure 5 shows the process model discovered for the transports that originate from system 1 and Figure 6 shows the process model discovered from the transports that originate from system 2. As can been seen, the flow for transports in system 2 is simpler and more straightforward compared to the process flow related to transports originating from system 1.

As part of this analysis, we analyzed the quality of the transportation process. As explained earlier the transportation process starts in the development environment. Several activities may be executed in the development environment, it should then move to the quality environment and finally to the production environment. We found 36 cases that did not follow the process path as it was designed. Figures 7 and 8 show examples of transports that did not adhere to the desired flow of activities in development environment followed by testing activities in quality environment followed by activities in production environment.
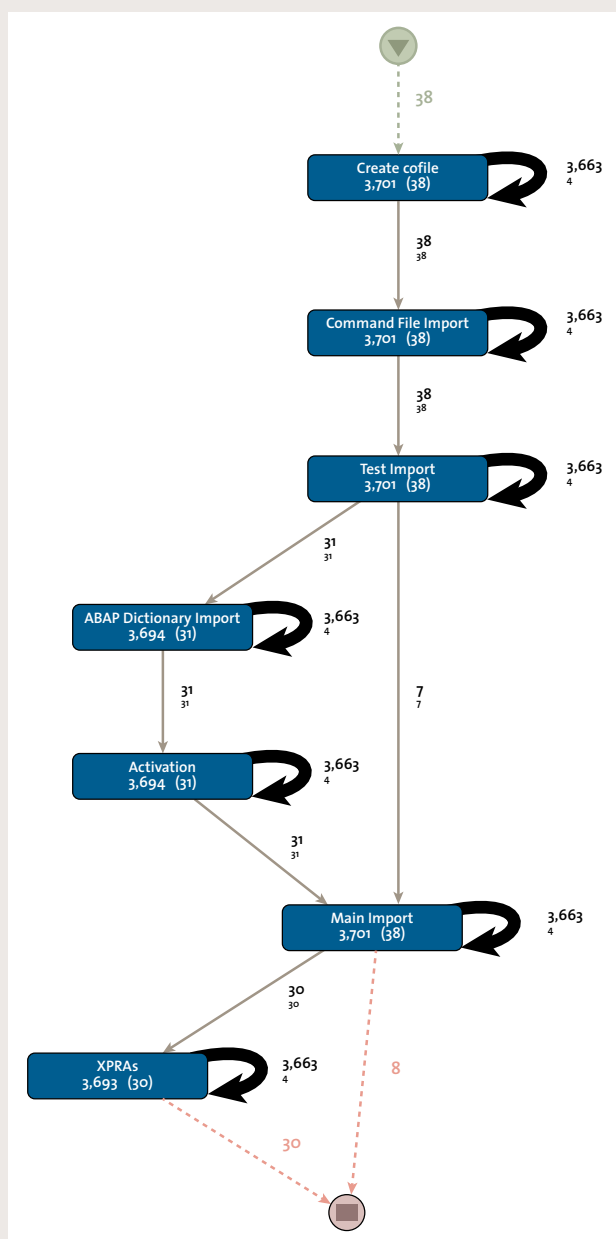
Figure 6. **The transportation process discovered for transports that originate from system 2.**
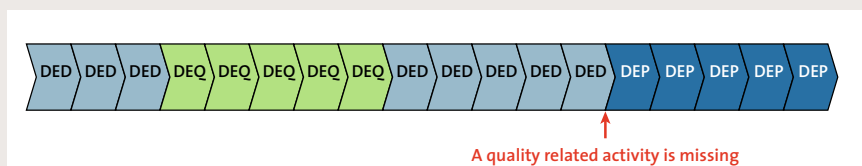

Figure 7. **An example of a violating transport that was imported to production environment directly from the development environment.**
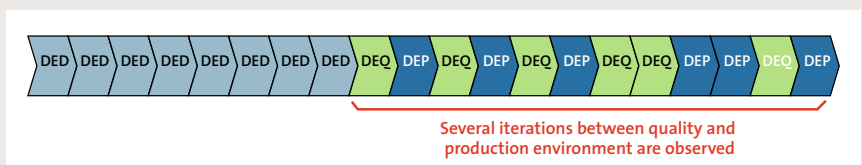

Figure 8. **An example from a transport that was imported from the production back to the quality environment multiple times.**

As is shown in Figure 7, the transport was directly imported from the development environment (DED) to the production environment (DEP) without passing through the quality environment (DEQ).

In another example shown in Figure 8, the auditor observed several iterations between the quality (DEQ) and the production environment (DEP). This observation reveals that the transport needed to return to the quality environment several times. Therefore, there is a high probability that the transport did not have the expected quality and should not have been imported to the production environment in the first place.

Using process mining, the auditor was able to do a precise and detailed analysis on the transportation management process. The auditor used the entire population of data to analyze the quality of the transportation process and provided detailed and precise diagnostics about the violations.

Process mining methodology

Conducting a successful process mining project whether in the context of an audit or a process analysis and improvement project, is not trivial. It is recommended to follow a systematic approach. Below, we present a generic methodology that can be adopted according to the requirements of a specific project.

A process mining project can have a very specific target, e.g. increasing efficiency and reducing costs by 10% for a given process. The goal may also be more abstract, e.g. obtaining valuable insights regarding the compliance state of several processes. Either way, it is necessary to translate these goals into specific analysis questions. Using process mining methodology, these questions are iteratively investigated, refined and answered. The results of such analysis are the basis of improvement initiatives and concrete action points for the selected process.

An overview of the process mining methodology is shown in Figure 9. The methodology consists of six stages that relate to several different input and output objects.

The first two stages of the methodology are **(1) planning** and **(2) extraction**, during which initial analysis questions are defined and event data is extracted. The processes in scope, the period of the analysis, the business questions to

Usually process owners and middle-level managers are the group which appreciate the true value and insight that process mining technology can add to a business
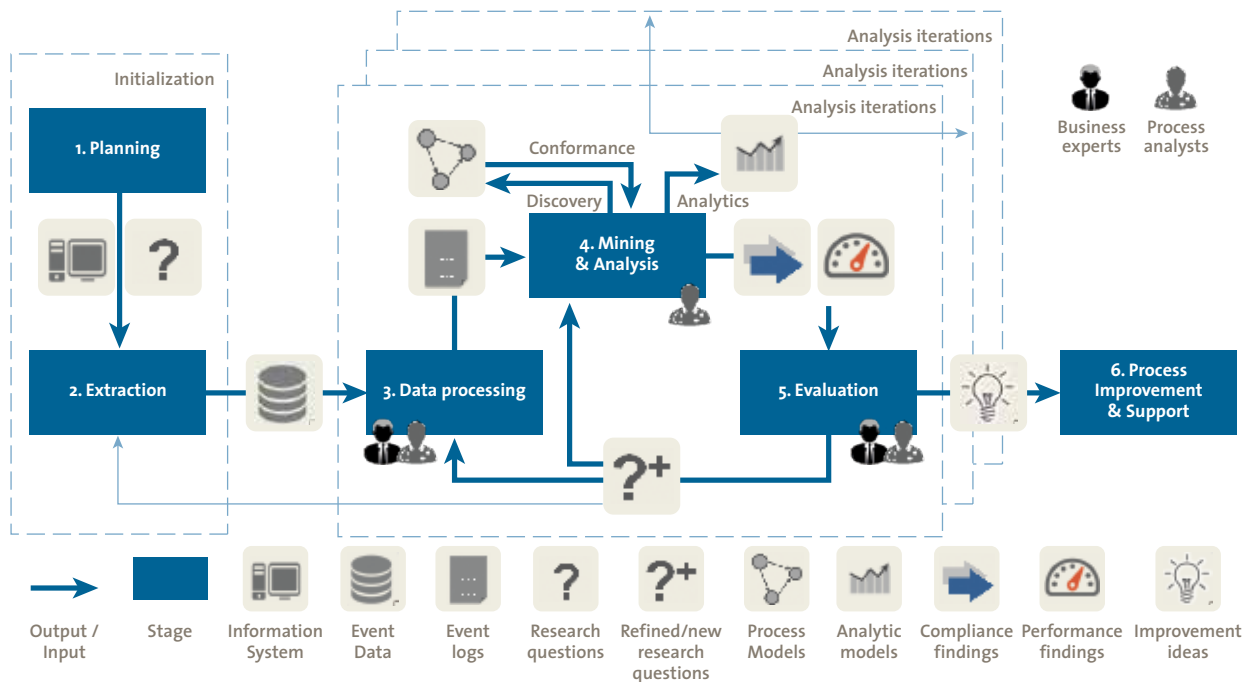

Figure 9. **Process mining methodology ([AALS15]).**

be answered, the team composition and the analysis timeline are defined during the planning. During data extraction, data requirements are defined, the scope of the data extraction is set and the data is retrieved. The systems and tables that need to be retrieved, the attributes (data fields), granularity of the data, and the logic with which the data should be collected and connected must be defined at this stage.

The next stage is **(3) data processing**. In many situations, the retrieved data cannot be directly used for process mining, rather some preparation and transformation steps are required. Depending on different analysis questions, data processing may be executed multiple times to enable a specific analysis. In the **mining & analysis stage (4)**, we apply different process mining techniques to answer the analysis questions. As we discussed previously these techniques include: *process discovery, conformance checking, process enhancement* and *process analytics.*

Established techniques can be used for standard analysis questions such as 'what does a specific process look like?'. However, if the analysis questions are more abstract, more explorative analysis is required and sometimes a combination of several techniques must be used.

The object of the fifth stage in the process mining methodology, **(5) evaluation**, is to relate the analysis findings to improvement ideas that achieve the project goals. This includes the correct interpretation of the results. Note that this interpretation need to be validated and verified by domain experts.

Finally during the **(6) process improvement & support** stage, the insights obtained from the previous stages are used to modify the actual process execution. Note that at this stage, process re-engineering techniques and Six Sigma can be leveraged. Finally process mining can be used to continuously monitor the processes and support a sustainable change in the operation.

Process mining tool market

Although process mining is a relatively new discipline in data science, many commercial tools have been introduced to the market during the last few years. There are various ways to characterize process mining software. Process mining software can be dedicated to purely process mining or it can be embedded in a larger suite such as a larger BPM, BI or data mining suite. Some criteria that can be used to characterize process mining tools are listed below. As such it is important to determine which functionalities are important, and look for a process mining tool which best fits the requirements.

Data import. As discussed before, the input for process mining is event data. Process mining software may have different mechanisms to import event data: *File* (event data stored as XES, MXML, CSV, or Excel file), *Database* (event data loaded from a database system), *Adapter* (event data from a particular application (e.g. SAP) through a dedicated piece of software), or *Streaming* (stream of events read through an event bus or a web service).

Hosting. Process mining software may run locally or remotely as: Stand-alone (the software runs locally, e.g., on the laptop used for analysis), On premise (the software runs on a server inside the organization) and Cloud (the software runs on cloud).

Supported process mining techniques. We can categorize process mining tools with respect to the process mining techniques they support: Discovery, conformance checking, enhancement and process analytics.

Open source versus closed source (commercial). Process mining software can be open source or commercial. ProM is the leading open-source process mining tool with over 1500 plugins that cover the state-of-the-art in process mining techniques. There are some other non-

commercial tools like PMLAB or CoBeFra, however the majority of research in process mining is implemented in ProM.

Table 2 shows an overview of the different commercial process mining tools in alphabetical order. However this list shows only some of a selection of process mining tools.

As mentioned earlier, several commercial tools have emerged on the market in recent years. Compared to ProM, these tools are easier to use, but provide less functionality compared to academic tools.

Software name	Vendor
Celonis	Celonis GmbH
Disco	Fluxicon
EDS	StreoLOGIC Ltd
Fujitsu	Fugitsu Ltd
LANA	Lana Labs GmbH
Minit	Gradient ECM
MyInvenio	Cognitive Technology
Process Mining Factory	icris B.V.
Perceptive	Lexmark
QPR	QPR
Rialto	Exeura
Signavio Process Intelligence	Signavio GmbH
SNP	SNP Schneider-Neureither & Partner AG
PPM	Software AG

Table 2. **Examples of commercial process mining tools.**

Most of the commercial process mining tools support process discovery but only a few support conformance checking. The functionalities of the commercial tools are rather limited compared to academic tools. However these tools aim at supporting less experienced users. Most of the commercial tools can work efficiently with large sized event data.

How to Conduct a Successful Process Mining Project

Usually process owners and middle-level managers are the group which appreciate the true value and insight process mining technology can add to a business. Hence, process mining is not usually a top-down initiative in companies. Nevertheless, support and commitment of high-level management is always required to overcome the difficulties on the way.

Similar to any project that introduces and establishes a new technology in an organization, process mining projects may face the typical challenges. Along with these challenges one should not have an unrealistic picture of data availability and data preparation. Although in the big data era, omnipresence of data has stimulated several data analysis projects in companies, we should know that obtaining meaningful data is not always easy. As described earlier, data has to comply with minimal requirements when using process mining techniques. In addition, like many other data analysis projects, data quality plays an important role in gaining useful results. For example, in many situations the timestamp of events are missing or they are only recorded at a day level whereas several events occurred during a day. Consequently, it is not possible to know the exact sequence of these events. Note that, many of these problems have already been researched and for many problems solutions exist to overcome them or minimize their negative impact on the analysis. Nevertheless, we would like to emphasize the importance of data preparation before starting a process mining project. In the following, we list some criteria that help to increase success in a process mining project.

Understand the business problem and prove value
First of all, it is important to understand and know in detail the business questions that should be answered using a process mining analysis. Process mining technology has an explorative nature and offers a lot of possibilities to analyze a dataset. However, one can get bogged down in the analysis results. It is therefore essential to specify exactly what business questions should be answered and iteratively relate the findings to these questions. That is, one needs to focus on the value that the results of the analysis should provide instead of getting excited about all the different analysis possibilities.

Choose your first project wisely
The event data obtained from workflow driven systems are usually the best candidates to start with process mining. It is better to start with datasets that require less prepara-

tion. It is also important to start with a process that you know and has a clear structure. It will then be easier to interpret the results and obtain insights. In addition, it will be easier to convince higher-level management about the value that can be added to the business using process mining. Special care is also required during the preparation of data. Sloppy data preparation can lead to inaccurate results. In this case, you may lose the interest of higher-management ([Rozi17]).

Involve people and communicate openly
Input for process mining projects are event data. However, in many situations the context of a business process is not captured in data. It is therefore important to communicate the results with relevant people as observations rather than presenting subjective conclusions. Involve domain experts and speak openly and transparently about the data that you use and about the facts that come out of this analysis.

References
[AALS15] W.M.P. Aalst, *PM 2: A Process Mining Methodology*, Advanced Information Systems Engineering - 27th International Conference, Stockholm, Sweden: Springer, 2015, pp. 297-313.

[AALS16] W.M.P. Aalst, *Process Mining*, Data Science in Action, Second Edition, Springer, 2016.

[ROZI17] A. Rozinat, *Flux Capacitor*, Fluxicon, 2017, https://www.fluxicon.com.

About the authors
Elham Ramezani Taghiabadi Ph.D. is a senior consultant at KPMG Risk Consulting in the Data & Analytics team. She did an innovative research in process mining from theory to application and developed methodology and techniques for a comprehensive data-driven analysis of business processes. Elham leveraged process mining to analyze and improve processes in different business sectors including the public sector, health care and banking.

Ir. Marijn Nagelkerke RE is a manager at KPMG Risk Consulting and is involved in many audit and advisory engagements with national and international clients. He is part of the Data & Analytics team and has extensive experience in performing data analytics as part of audits.

Drs. Paul Kromhout RE CISA is a director at KPMG Risk Consulting Financial Services. He has more than fifteen years of experience in the financial services industry and started using process mining four years ago to support the financial audit and as part of a broader data analytics proposition in the financial services sector (Banking, Leasing, Insurance and Pensions).