AUDITING WITH PROCESS MINING

Summary

Auditors have a clearly defined process in which they carry out an audit. Process mining does not replace this traditional audit approach. However, it requires some changes and a conscious effort to fit process mining into the existing way of working. In some places, more work is needed. In other places, things get easier. In this article, we describe in detail how process mining fits into the different phases of the audit cycle based on a concrete project. We describe the changes that needed to be made to the audit process and the benefits and challenges.

City of Vienna Court of Audit

The City of Vienna Court of Audit is an autonomous and independent public audit institution. It audits the institutions and entities in the City of Vienna concerning their financial management and safety. And it supports those in positions of responsibility in politics and administration with audit reports and recommendations.

- **Process mining case study from the perspective of the audit discipline**
- **10 recommendations were given to the Wiener Stadtwerke as a result of this audit**
- **Key success factor was taking the time to truly understand the data**

In the framework of its audit work, the City of Vienna Court of Audit reviews the use of Vienna's public funds. It also monitors compliance with safety regulations to protect the citizens of Vienna and its visitors.

However, the resources to perform these tasks are limited. To set audit priorities, the City of Vienna Court of Audit follows a selection procedure for audits in the form of a risk analysis.

For the selected audits, the City of Vienna Court of Audit team then uses different audit procedures depending on the audit. Process mining has been used as an audit method in several audits since 2016. In 2020, the purchase-to-pay process of the Wiener Stadtwerke, one of Austria's largest infrastructure groups, was analyzed.

Using the example of the Wiener Stadtwerke, this article shows in detail how process mining was applied to perform a data-driven audit. The results of this audit are publicly available in the audit report [1] here [2]. The following article focuses on the method of process mining in the context of an audit. It describes in detail how process mining was leveraged in the different audit phases and the challenges and benefits we experienced.

Data driven audit approach with process mining

Internationally accepted audit standards guide the audit work of the City of Vienna Court of Audit. At the same time, It is the goal to further improve the existing standards in cooperation with national and international audit institutions while engaging in the exchange of experiences. Audits are carried out according to the standardized audit process (see Figure 1).

Figure 1: Overall audit process

Each step of the process depicted above contains several tasks. Figure 2 shows the tasks related to the step ‘Conducting the audit’: First, an audit concept is created. Then, the data is collected. This data is then used to perform a situational and deviation analysis, from which the audit results and recommendations are derived. In addition, the audit trail and evidence are documented, and the audit file is generated.

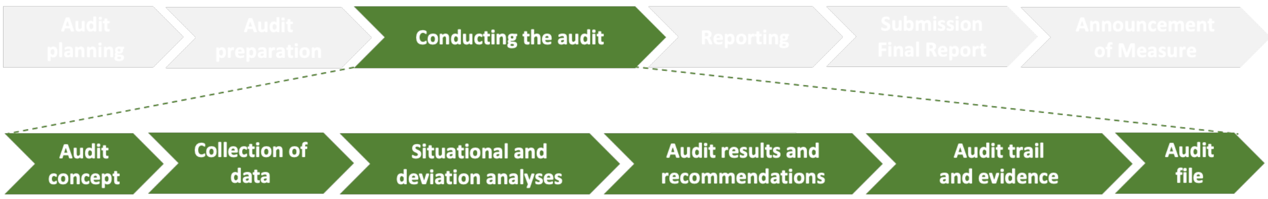


Figure 2: Conducting the audit

We must adapt our working method when we use process mining in our general audit approach. Especially the way of collecting and analyzing data changes within the audit process compared to other audit methods.

To include process mining into the ’Collection of data’ and ‘Situational and deviation analysis’ phases, we followed the nine steps shown in Figure 3.

Following this model helped us a lot to standardize our approach when using process mining in an audit. It summarizes the essential deliverables for collecting the data and performing the situational and deviation analysis. In the following sections, we explain each step and each deliverable of the model depicted in Figure 3 in more detail.

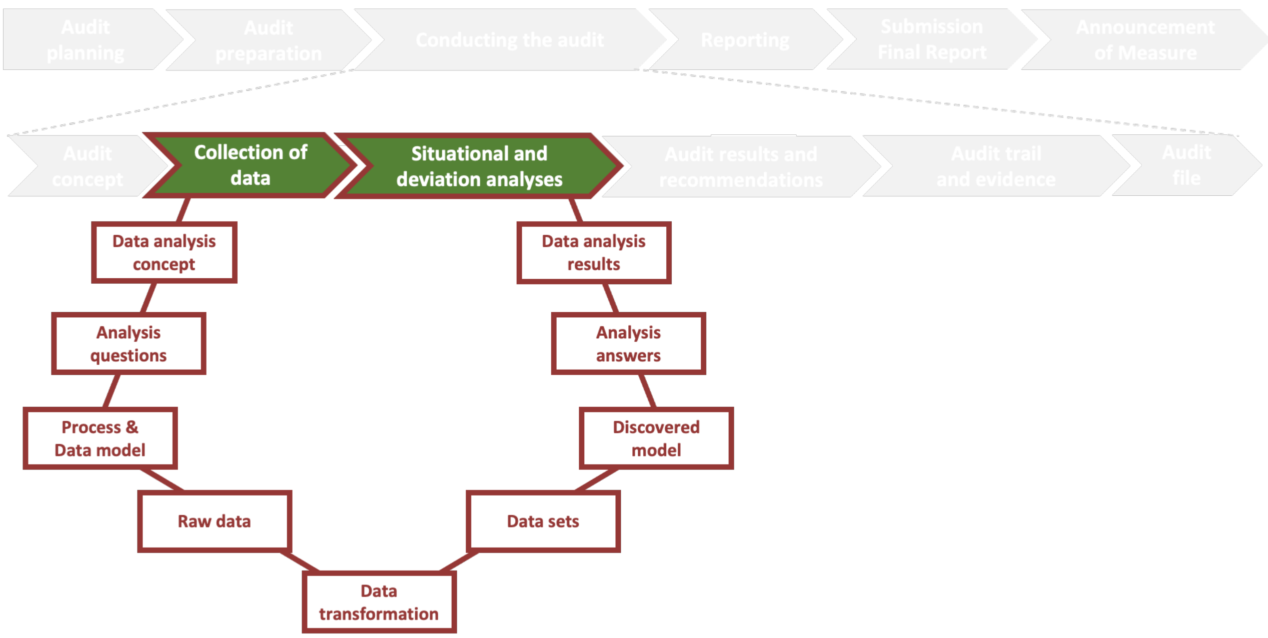
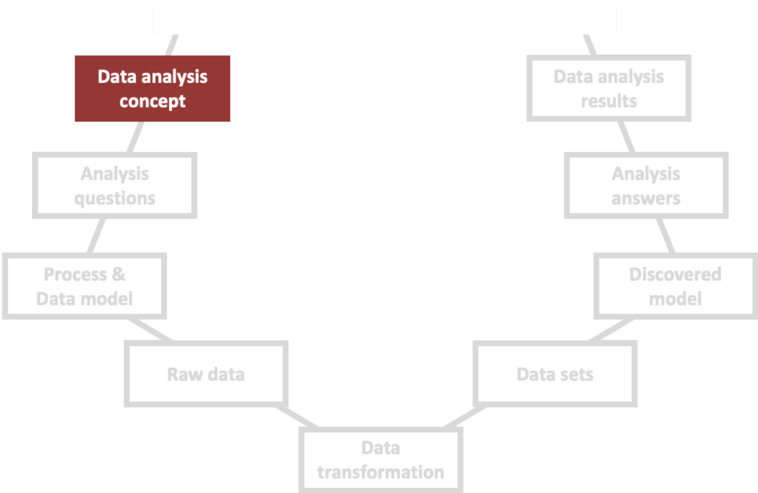


Figure 3: Data-driven audit approach with process mining

Step 1: Data analysis concept



The City of Vienna Court of Audit follows a risk-based audit selection procedure. Therefore, the audit was already defined within the annual audit planning. The 'Data analysis concept' determines the audit scope in further detail. It gives an overview of the audited party, the process of interest, the IT framework, and the main audit objective.

The audited party - Wiener Stadtwerke - is one of Austria's most significant infrastructure groups with about 15.000 employees. Its business activities can be categorized as follows:

- Energy (electricity, gas, heating, cooling)
  - Generation
  - Distribution
  - Grid operation
- Public transport (subway, tram, bus)

  - Traffic management and planning
  - Operation
  - Marketing
  - Distribution
- Funeral
  - Cemeteries
  - Cemetery nursery
  - Stonemasonry
- Car Parks

The total assets of Wiener Stadtwerke amounted to approximately Euro 13,900 million on 31 December 2020. The Wiener Stadtwerke Group was 100% owned by the City of Vienna.

We defined the process of interest in more detail by describing the general process scope. As we planned to audit the purchase-to-pay process, we delimited this process as depicted in Figure 4. The audited process scope comprised procurement and invoice processing, beginning with the demand report and ending with the payment of the corresponding invoice.

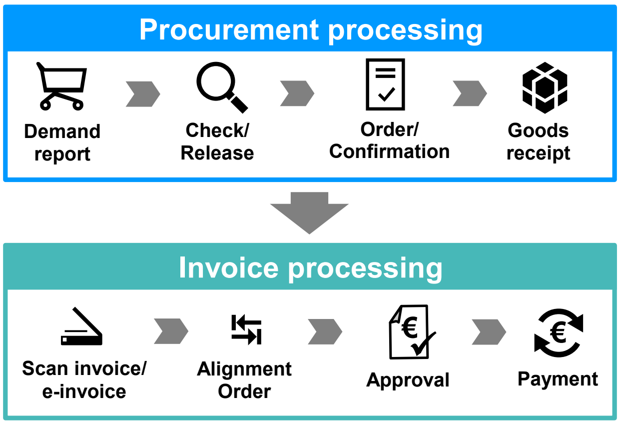


Figure 4: Purchase-to-pay process

This general process scope was used later as a reference point for further investigations and gave a first impression of the start and the end points of the process of interest.

The timeframe for the audit was determined to be the year 2019. More precisely, we considered all orders that were sent between 01 January 2019 and 31 December 2019.

We investigated the IT infrastructure of the City of Vienna Court of Audit and the audited party to get an idea about which type of data and which tools would be available to process and analyze this data.

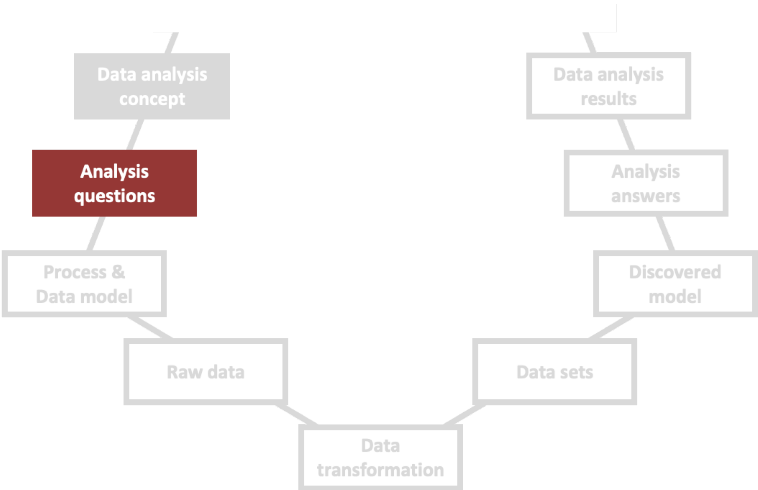
From our preliminary research, we knew that the audited party used SAP to administrate the purchase-to-pay process. We expected that we would need to transform the data after we exported it from SAP. As an ETL tool to do these transformations, we chose the KNIME Analytics Platform [3] because we already used this software for data transformations in earlier process mining projects and achieved good results.

The primary audit objective was to perform a compliance audit. We wanted to analyze the Wiener Stadtwerke's purchase-to-pay process concerning its regularity and compliance with organization-specific framework conditions.

Due to this main audit objective, we planned to address aspects like the completeness of the process, segregation of duties, adherence to the four-eyes-principle, and the effectiveness of the internal control system. In addition, we wanted to consider the lead time and the occurrence of bottlenecks from a performance perspective. User experience questions were outside the scope of this audit.

After defining the general framework and the primary audit objective, it was time to specify the focus areas of the audit in more detail. So, in the next step, we identified the concrete analysis questions we wanted to answer within our process mining analysis.

Step 2: Analysis questions



We defined the following 12 analysis questions (see Table 1). As shown in Table 1, most of the questions are related to compliance issues (our primary audit objective), only two are related to performance questions, and none are about user experience.

Besides formulating the general analysis questions, we also tried to define them as precisely as possible and make them measurable. Thus, we specified the metric, target value, process scope of interest, and influencing factors for each analysis question. Table 2 shows how we made analysis question No. 2 more concrete by defining these aspects.

When we want to answer the question “Are all orders released?” it seems straightforward initially, but it is a good idea to think further about how exactly we can measure the answer to this question. We expected that order releases would be registered within the information system. So, we chose the presence of the release activity as a metric. In addition, we assumed that all orders must be released without exception. So, we set the target value to 100%, which means that a release activity needs to be documented for each order.

Then, we defined the process scope to show which part of the purchase-to-pay process is relevant to find the information needed to answer the analysis question. For question No. 2, the process scope comprised all activities related to the release step in the information system.

Finally, we also collected the influencing factors we needed to consider while performing the data analysis and interpreting the results. Regarding the order release, we assumed that a four-eye principle might be

Nr.	Analysis questions	Performance	Compliance	Experience
1	Do the real processes fit the should-process?		x	
2	Are all orders released?		x	
3	Have all training orders been approved by the HR department?		x	
4	Is an invoice recorded for every order?		x	
5	Are all orders payed?		x	
6	Are all invoices payed?		x	
7	Have all invoices been checked before payment?		x	
8	Are orders payed although they were cancelled or deleted?		x	
9	Is the four-eyes principle maintained?		x	
10	Were all invoices that exceeded the order value released according to the four-eyes principle?		x	
11	Are the payment targets observed?	x		
12	Are there any bottlenecks within the process?	x		

Table 1: Analysis questions

Objective	**Performance**		**Compliance**	**x**	**Experience**	
Analysis question	Are all orders released?					
Metric	Release activity needs to be documented					
Target value	100 % of orders are released					
Process scope	PO – document released level 1 to n, order position released					
Influencing factors	- Value limits (four-eyes principle)					

Table 2: Analysis question No. 2 in more detail

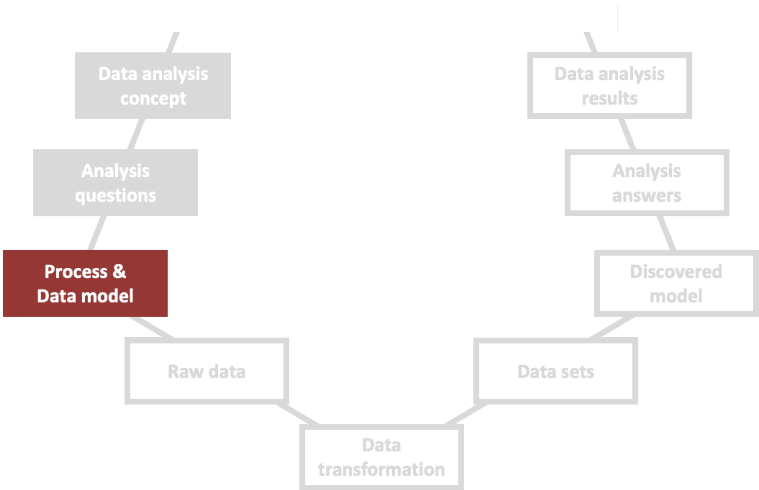
relevant above a specific value limit. Thus we defined the value limit as an influencing factor that needed to be kept in mind for later analysis.

We specified detailed definitions for all analysis questions in a similar manner as shown for question No. 2 above.

Because we defined the analysis questions at a very early stage of the process mining project, we needed to make certain assumptions, especially regarding the metric and the influencing factors. We adapted these analysis questions multiple times in later phases of the project because we kept gaining more insight into the process and the data.

Nevertheless, defining the analysis questions at this early stage of the project was very useful. It helped us to get a good overview of the data we needed. As a result, we could reduce the risk of forgetting certain aspects or data fields during data extraction.

Step 3: Process and data model



Looking at the purchase-to-pay process in more detail, it was clear that it was pretty complex. We had received a detailed process description from the audited party and decided to simplify the process and look at it from an aggregated perspective to handle the complexity. The high-level reference process we defined included only those steps that were essential to finding answers to the analysis questions described in the previous step.

We expected considerable amounts of data to be generated for this process in the information system. Based on the high-level reference process, we tried to identify the essential data fields populated while performing the process.

The purchase-to-pay process was mainly executed using SAP. For each process step, we looked for the corresponding database table and enriched the process model with this information. For example, the data for 'Create purchase request' could be located in the EBAN table in SAP (see Figure 5).

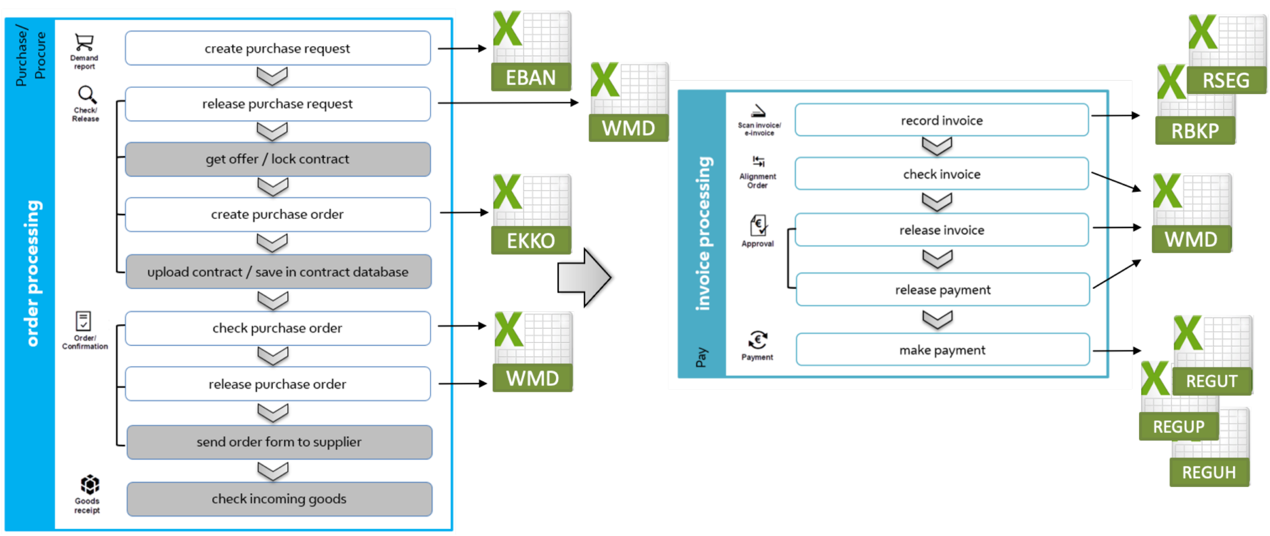


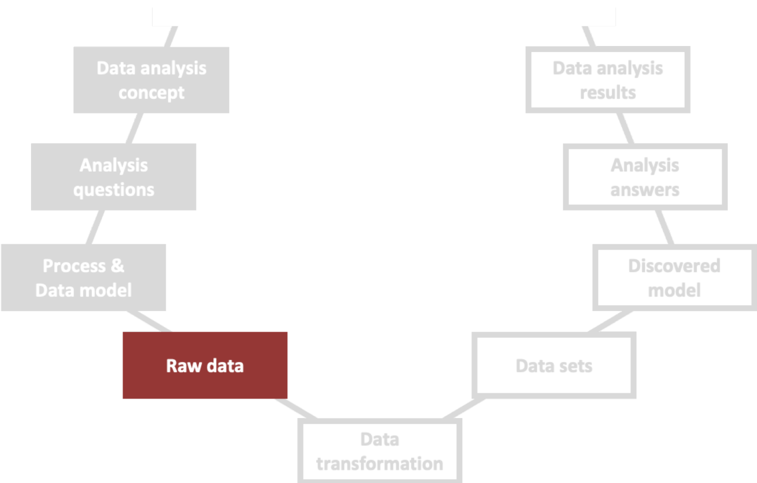
Figure 5: Process and data model

While creating the process and data model, we realized that not all the steps from the high-level reference process were done in SAP. For example, we detected that the approval and release workflows were performed with an SAP add-on called WMD xSuite feeder. As the approval and release steps were essential for the compliance audit, we included these data tables in the data model and the later data extraction.

Other steps like getting an offer, locking a contract, sending the order form to the supplier, and checking the incoming goods were neither performed in SAP nor with the WMD xSuite feeder. These steps (colored in grey in Figure 5) were performed manually or via E-Mail. Due to a lack of data availability, we excluded these steps from our process mining analysis.

After defining the process and data model, we had quite a good overview of the available data and where we could find the data. Thus in the next step, we extracted the raw data for further processing.

Step 4: Raw data



The data for our process mining analysis was stored in two different systems: SAP and the WMD xSuite feeder. We had identified the data tables that needed to be extracted from these systems when we specified the data model. We had no direct access to the Wiener Stadtwerke's information systems. Thus, the audited party extracted the data tables for us and provided the raw data in CSV files.

From the data model, we already knew in which tables the timestamps for each activity were located. For example, we knew that the timestamp for 'Create purchase request' could be found in the EBAN table. The timestamp for the 'Release purchase request' activity could be found in the WMD table, and so on. However, because the raw data was distributed across multiple CSV files, we also needed to find the connections between the individual data tables so that we could merge the files into one (see Figure 6 on the following page for the connections between the tables).

For each table, we identified which information could be used as a timestamp for an activity, resources, and other activity- or case-related attributes.

Based on the knowledge of the relevant data fields for the activity timestamps, attributes, and resources, and with this understanding of the connections between the raw data tables, we now had the basis for building our event log.

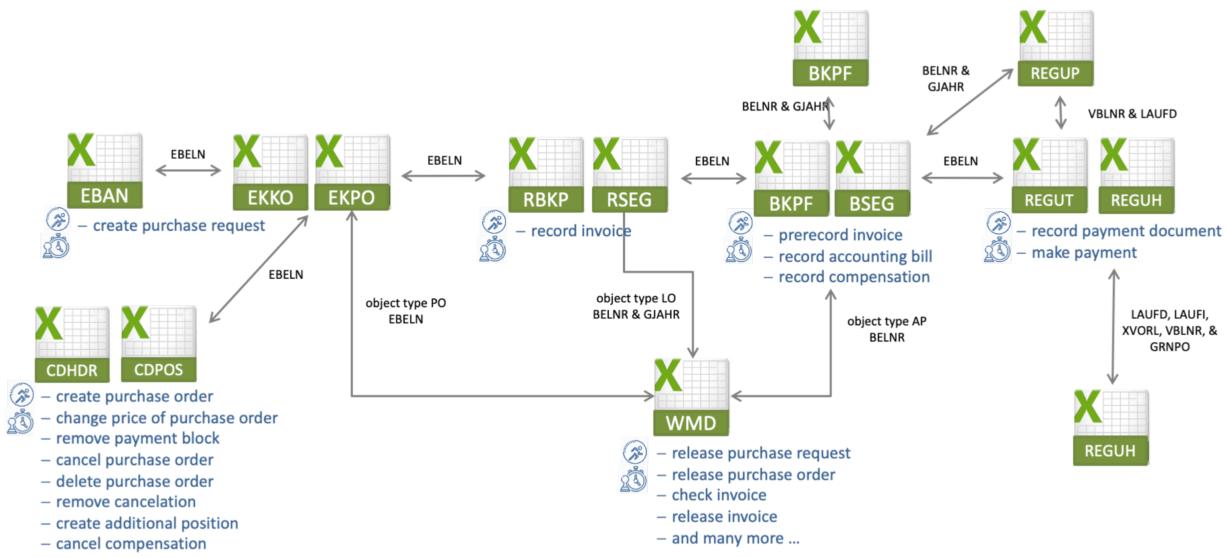
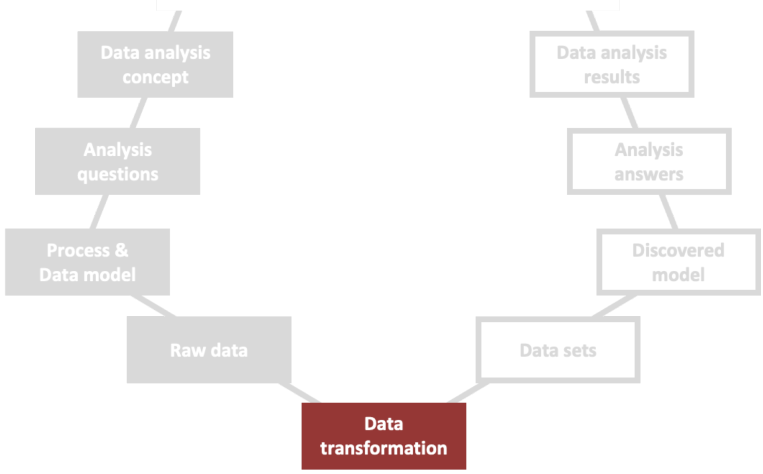


Figure 6: Raw data with connections between the tables

Step 5: Data transformation



The goal of the next step was to bring the raw data in a format that we could load into the process mining software. We filtered the relevant information from the raw data files and linked the data tables based on the prior defined connections. The output data was formatted as an event log, with a unique ID as case ID, activity names, timestamps, resources, and attributes for each event.

We performed the data transformation using the open-source software KNIME. To validate the transformed data, we performed crosschecking with the productive system whenever we implemented changes in the data transformation workflow. These validation steps showed quite some potential for improvement, and we adapted the workflow several times until the output data finally represented the data from the productive system (see Figure 7 on the next page).

The data transformation was the most time-consuming step within the process mining project. One of the factors was that we had no direct access to the productive system. Therefore, the audited party had to support the data validation process and help with crosschecking. This led to waiting times and delays within the project.

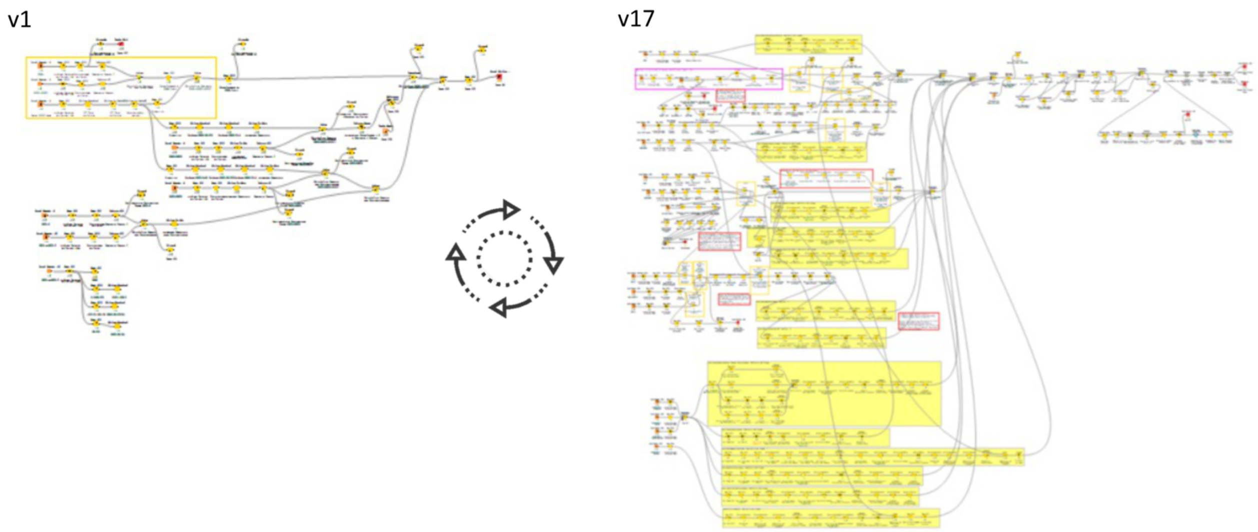
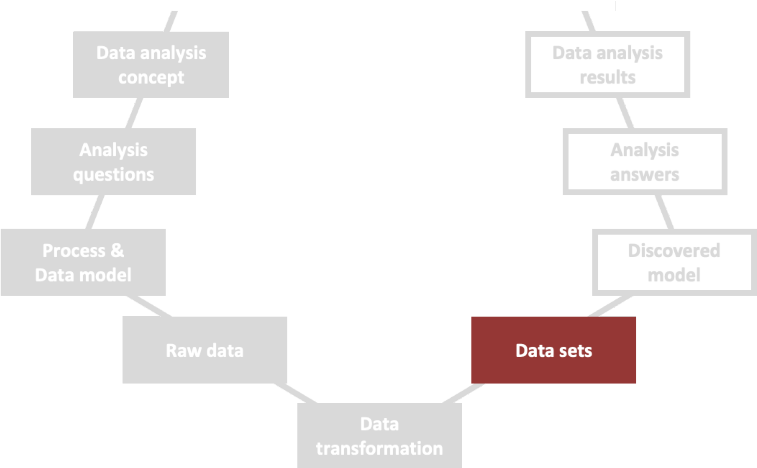


Figure 7: The first (left) and last (right) data transformation workflow version

Another factor was that we initially had not appropriately considered the 1:n and n:m relationships when tracing the case IDs. For example, one order can lead to several invoices and payments. Furthermore, one invoice can address multiple orders. One payment can cover more than one invoice, and so on. These many-to-many relationships [4] had to be adequately handled during data transformation.

After several adaptions to the transformation workflow, we passed all the validation steps and generated a data set we were confident working with.

Step 6: Data sets



The data transformation workflow generated a data set we could use for our process mining analysis. According to the data, between 01 January 2019 and 31 December 2019, a total of 2,550 orders with an order value of approximately 21 Mio. EUR were processed.

Initially, we had chosen the order number as our case ID. Therefore, all cases were analyzed from an order perspective (see Figure 8).

Case-ID	Timestamp	Activity	Invoice number
1030071289-10	10.01.2019 10:23:17	Create purchase order	
1030071289-10	11.01.2019 13:22:48	Release purchase order	
1030071289-10	23.01.2019 08:58:23	Record invoice	1230007
1030071289-10	26.01.2019 16:46:11	Record invoice	1230008
1030071289-10	25.02.2019 13:33:27	Check invoice	1230007
1030071289-10	25.02.2019 13:48:33	Release invoice	1230007
1030071289-10	01.03.2019 12:16:13	Check invoice	1230008
1030071289-10	01.03.2019 12:19:56	Release invoice	1230008
1030071289-10	14.03.2019 16:00:00	Make payment	1230007
1030071289-10	16.03.2019 14:16:30	Make payment	1230008

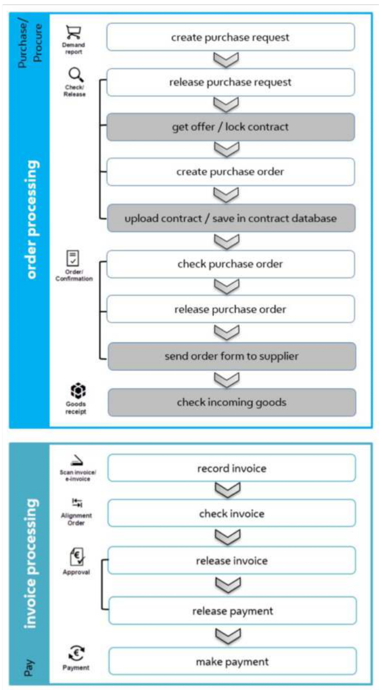


Figure 8: Order perspective (data set left, process view right)

However, during the analysis, it became clear that due to the 1:n relationship between orders and invoices, we could not answer all our analysis questions regarding invoice processing with this data set. For example, in Figure 8, one can see that two invoices (invoice 1230007 and invoice 1230008) are associated with order 1030071289-10. There are two events for activity “Check invoice” and “Make payment” (one for each invoice). This complicates answering questions such as analysis question No. 7 (“Have all invoices been checked before payment?”).

Therefore, we decided to generate a second data set focused on the invoice perspective. This was achieved by combining the order and invoice numbers into a new case ID. The scope of this second data set is smaller (invoicing and payment only). The benefit is that the activities related to invoice 1230007 and the activities related to invoice 1230008 now appear in their own case and can be analyzed separately (see Figure 9).

Case-ID	Timestamp	Activity
1030071289-10-1230007	23.01.2019 08:58:23	Record invoice
1030071289-10-1230007	25.02.2019 13:33:27	Check invoice
1030071289-10-1230007	25.02.2019 13:48:33	Release invoice
1030071289-10-1230007	14.03.2019 16:00:00	Make payment

Case-ID	Timestamp	Activity
1030071289-10-1230008	26.01.2019 16:46:11	Record invoice
1030071289-10-1230008	01.03.2019 12:16:13	Check invoice
1030071289-10-1230008	01.03.2019 12:19:56	Release invoice
1030071289-10-1230008	16.03.2019 14:16:30	Make payment

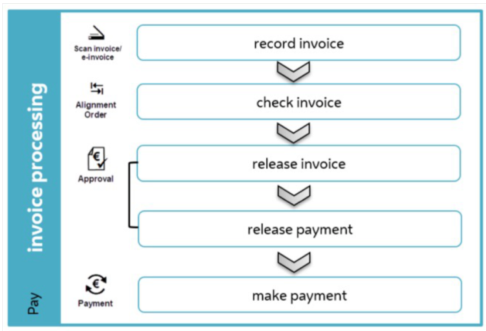
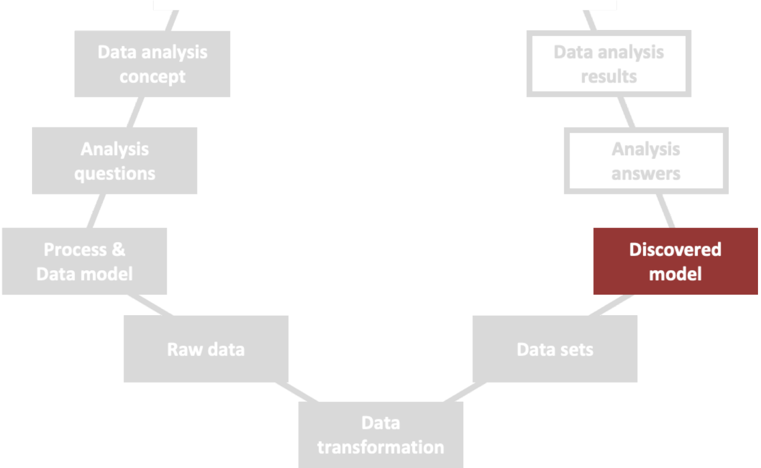


Figure 9: Invoice perspective (data set left, process view right)

Based on these two data sets – one from an order perspective and one from an invoicing perspective – we could now start answering our analysis questions.

Step 7: Discovered model



Once we had access to our transformed data sets, we loaded the data into the process mining software Disco [5] and got a first impression of the complexity of the process.

Although we had worked with simplification methods from the beginning and focused on the activities from the high-level reference process depicted in Figure 5 to identify relevant data tables, the process map was still very complex. Figure 10 shows the discovered process model from an order perspective.

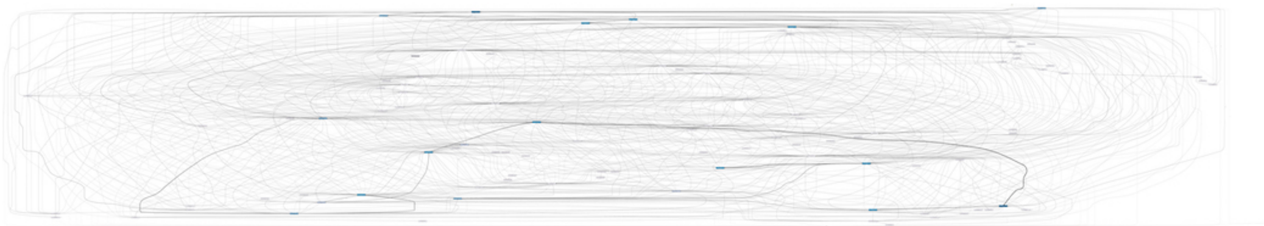


Figure 10: Discovered process model

Due to the high complexity, we applied further simplification strategies to enable an explorative analysis and a should-be comparison of the real process paths and the reference process.

Firstly, by including most of the timestamp fields that we could find, we had derived a high number of activities from the raw data files. Among these activities were administrative process steps that were outside our reference process. We reduced the number of activities by only keeping those process steps that we could directly map to the high-level reference process (Milestone simplification method [6]). This reduced the number of activities from more than 100 to approximately 50. Note that the data in the IT system was still more detailed than the high-level process. For example, a purchase order could be checked, rejected, and released on different levels (see Figure 11).

Secondly, there was still a high variation regarding the process paths. Therefore, we decided to cluster the data into four groups (Semantic variant simplification method [7]). These four groups were:
(1) canceled cases,
(2) cases without an invoice,
(3) cases with one invoice, and
(4) cases with multiple invoices.
By looking at each data segment separately, the number of process variants was further reduced.

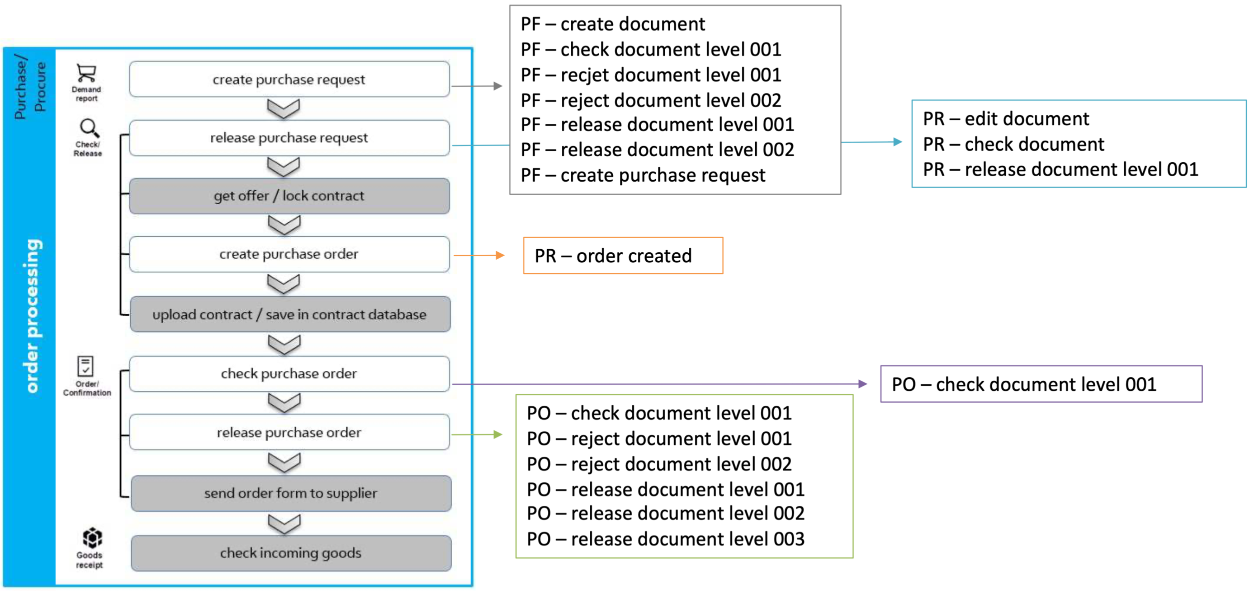


Figure 11: Mapping data to the high-level process

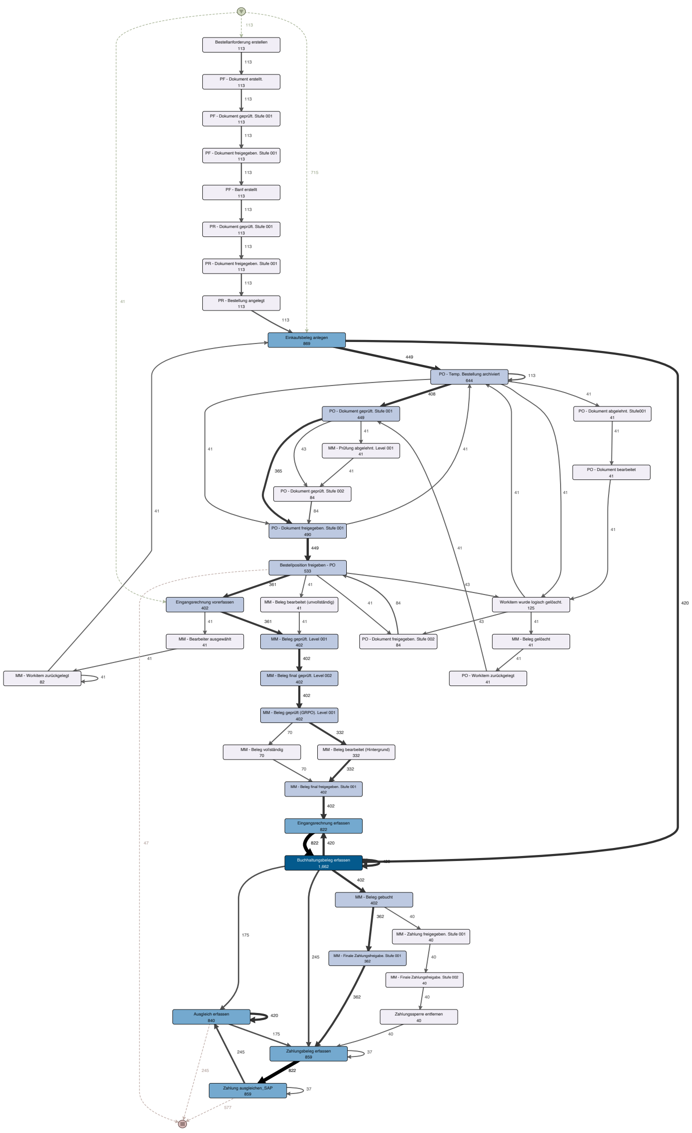
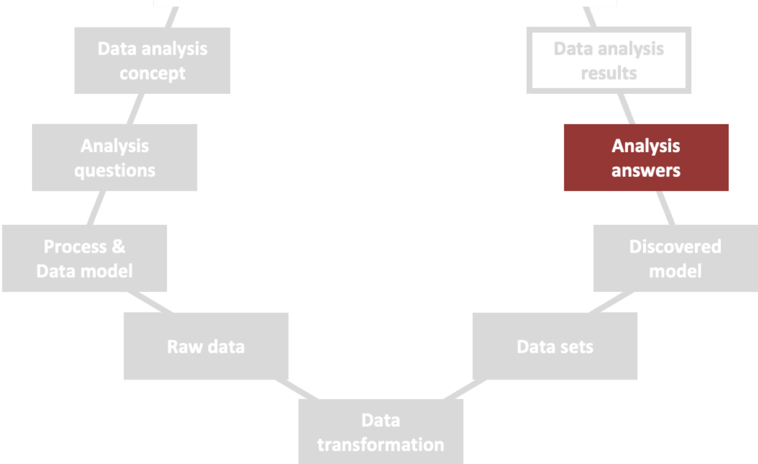


Figure 12: Discovered model after simplification

Finally, we also decided to focus on the most common process paths to get an overview of the mainstream behavior (Variant simplification method [8]). Figure 12 shows the discovered model based on only the ten most frequent process variants. This helped us to get an overview of the main process before going into detail and analyzing the less frequent paths and how they deviate from mainstream behavior.

Due to the complexity reduction, we could now perform an explorative analysis, searching for inconsistencies and analyzing unexpected process paths in more detail.

Step 8: Analysis answers



To answer our analysis questions from Table 1, we used both an explorative and targeted analysis.

During the explorative analysis, we investigated the discovered models looking for unexpected or strange process paths, long waiting times, and other abnormalities in a broader way. With this type of analysis, we were able to address questions No. 1 (“Do the real processes fit the should-process?”) and No. 12 (“Are there any bottlenecks within the process?”).

Questions No. 2 to 11 were answered through a targeted analysis. We translated each analysis question into a customized set of filters based on the definitions we had created in the ‘Analysis questions’ step. Although we had already defined the analysis questions in some detail, we further refined these specifications to ensure we only detected cases that violated the process requirements.

For example, the goal of question No. 2 (“Are all orders released?”) was to find out whether there were cases with a missing order release. The target value for this analysis was 100%. However, there could be legitimate reasons why an order release could be missing. We concluded that only orders that were executed needed a release activity. Furthermore, there might be orders that were never carried out. Thus, we decided to exclude canceled cases from the data basis for this question.

So, we used a combination of three different filters to answer question No. 2 (“Are all orders released?”). First, we excluded all orders that were canceled, which in our data set meant removing cases with the attribute value L or X in the attribute “Löschkennzeichen” (see Figure 13 on the next page).

Second, finding out whether an order was really executed was not easy. We had no data that indicated when an order form was sent to the supplier. There was also no data that showed when the contract was

Objective	Performance		Compliance	x	Experience	
Analysis question	Are all orders released?					
Metric	Release activity needs to be documented					
Target value	100 % of orders are released					
Process scope	PO – document released level 1 to n, order position released					
Data set used	Order perspective					
Influencing factors	- Value limits (four-eyes principle) - Cancellations					

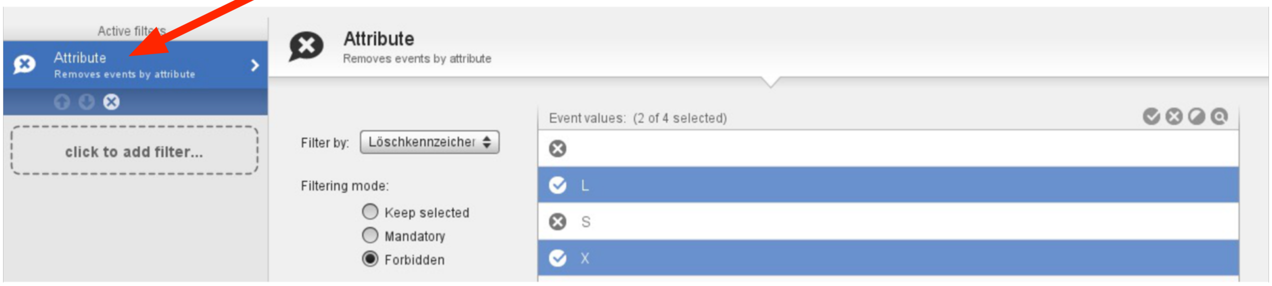


Figure 13: Excluding canceled orders

locked or when the incoming goods were checked. Therefore, we chose the invoice as a means to measure if an order was actually performed. Figure 14 shows the second filter that includes only orders with an invoice (“Eingangsrechnung erfassen” in German).

Objective	Performance		Compliance	x	Experience	
Analysis question	Are all orders released?					
Metric	Release activity needs to be documented					
Target value	100 % of orders are released					
Process scope	PO – document released level 1 to n, order position released					
Data set used	Order perspective					
Influencing factors	- Value limits (four-eyes principle) - Cancellations - Order was really placed					

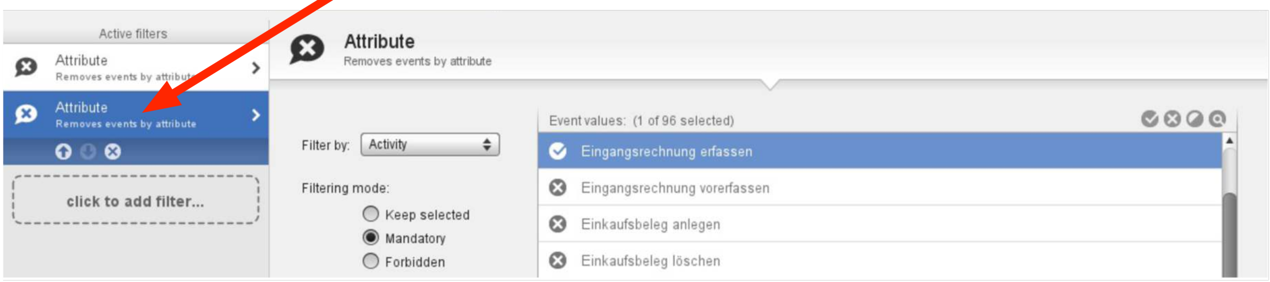


Figure 14: Including only orders with an invoice

Third, because we wanted to find violations of question No. 2, we excluded all orders with an order release activity. So, we searched for the opposite of the required behavior in the process. Figure 15 shows this third filter that excludes all cases that contain any of the selected order release activities (“PO - Dokument freigegeben. Stufe 001” - “… 003”).

Objective	Performance		Compliance	x	Experience	
Analysis question	Are all orders released?					
Metric	Release activity needs to be documented					
Target value	100 % of orders are released					
Process scope	PO – document released level 1 to n, order position released					
Data set used	Order perspective					
Influencing factors	- Value limits (four-eyes principle) - Cancellations - Order was really placed					

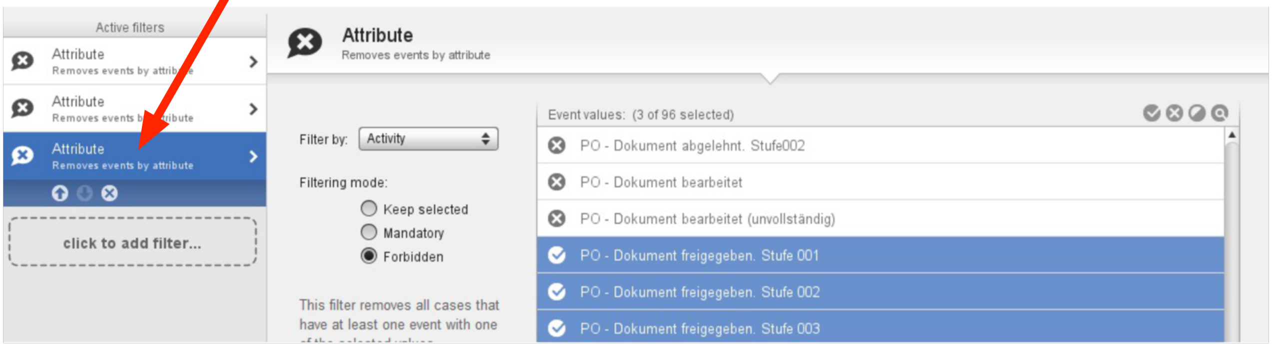
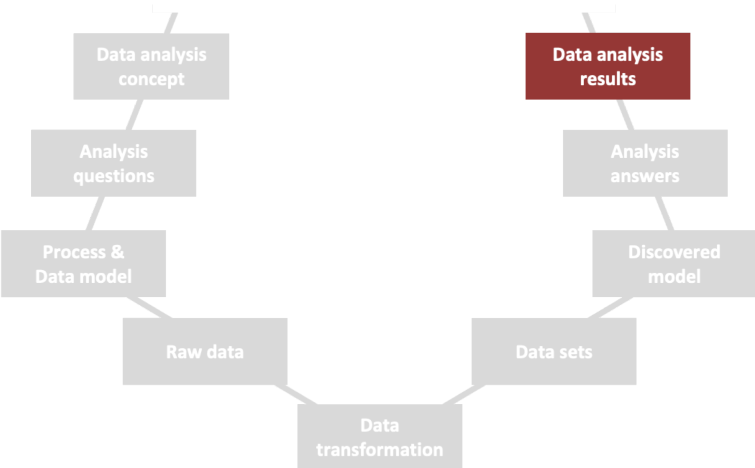


Figure 15: Excluding orders with an order release activity

As a result of applying these three filters, we found that 17% of all non-canceled cases lacked an order release, although an invoice was recorded. Looking at those cases in more detail, we discovered that all orders referred to the goods group of office supply. The audited party explained that office supplies did not need to go through the regular order release workflow and could be ordered without a prior order release. Thus, these orders also complied with the procurement guidelines despite the irregularities we discovered.

Similar to the example above, we answered each of the remaining analysis questions by translating them into customized filter settings. In addition, we discussed all the discovered irregularities with the audited party to determine whether there was a genuine reason for the deviation or whether we had found a compliance violation within the process.

Step 9: Data analysis results



The data analysis results summarized the answers to the analysis questions and were the basis for our audit report, including the findings and recommendations.

Figure 16 shows that we could answer all our questions. Only questions No. 1 and No. 12 could not completely be answered because on the one hand we had some blind spots in the process due to a lack of data and on the other hand we had to simplify the process considerably to deal with its complexity. Therefore, we set their status to yellow.

Nr.	Analysis questions	Performance	Compliance	Could be answered
1	Do the real processes fit the should-process?		x	⚫🟡⚫
2	Are all orders released?		x	⚫⚫🟢
3	Have all training orders been approved by the HR department?		x	⚫⚫🟢
4	Is an invoice recorded for every order?		x	⚫⚫🟢
5	Are all orders payed?		x	⚫⚫🟢
6	Are all invoices payed?		x	⚫⚫🟢
7	Have all invoices been checked before payment?		x	⚫⚫🟢
8	Are orders payed although they were cancelled or deleted?		x	⚫⚫🟢
9	Is the four-eyes principle maintained?		x	⚫⚫🟢
10	Were all invoices that exceeded the order value released according to the four-eyes principle?		x	⚫⚫🟢
11	Are the payment targets observed?	x		⚫⚫🟢
12	Are there any bottlenecks within the process?	x		⚫🟡⚫

Figure 16: All our analysis questions could be answered

Within the explorative analysis, we discovered that in 2019 approximately 1% of all orders were canceled. In addition, for 1.4% of all cases, an order was placed in the information system, but no further activities were recorded. These cases appeared to be canceled as well, but the cancellation was not documented in the information system. While the total percentage of canceled orders was in a normal range, we identified a potential for improvement regarding the documentation of canceled cases.

Furthermore, we found that ca. 7% of all cases started with an invoice recording (the corresponding order was only placed later). These cases indicated that the formal procurement process was not always observed. Instead, some employees made purchases without prior authorization, an undesirable behavior called ‘Maverick Buying.’

Within the targeted analysis, we had realized that 17% of all invoiced orders were not released. However, this irregularity was tolerable because they all belonged to the product group office supplies, which did not require an order release activity.

The procurement guidelines of the audited party also required that all orders with an order value of more than 20,000 EUR had to be released following the four-eyes-principle. Our analysis showed that, except for one case, the four-eyes-principle was observed whenever the order value of 20,000 EUR was exceeded. In this one case, a two-staged release was documented in the information system, but one person was not entitled to release an order of this value. This deviation from the defined release process indicated that there might be a weakness in the internal control system. Therefore, we recommended that the release permissions should be evaluated on a regular basis.

Regarding invoice processing, the procurement guidelines demanded that all invoices be checked before payment. Our process mining analysis showed that these checks were not consistently documented in the information system. Thus, it was not transparent if the necessary checks were performed before payment. In addition, a four-eyes-principle was mandatory during invoice release whenever the invoiced sum deviated more than 10% or 1,000 EUR from the order value. In individual cases, this four-eyes-principle was not followed. Therefore, we recommended that further control measures needed to be implemented to ensure that all invoices were checked properly before they were released for the final payment. Furthermore, our process mining analysis showed that approximately half of all invoices were immediately due for payment, although the general payment target was set to 30 days after receiving an invoice. Thus, there was potential improvement regarding the documentation of payment targets.

Finally, the payment target was not observed in approximately 6% of all cases. The audited party reasoned that, in some cases, there was still a need for clarification regarding the incoming goods or the conformance of performance when an invoice was due for payment. Further investigations also showed that when the payment targets were not met, there often had been a delay in releasing the invoice for the final payment. We recommended monitoring the observance of payment targets and implementing measures to reduce waiting times if needed.

Conclusion

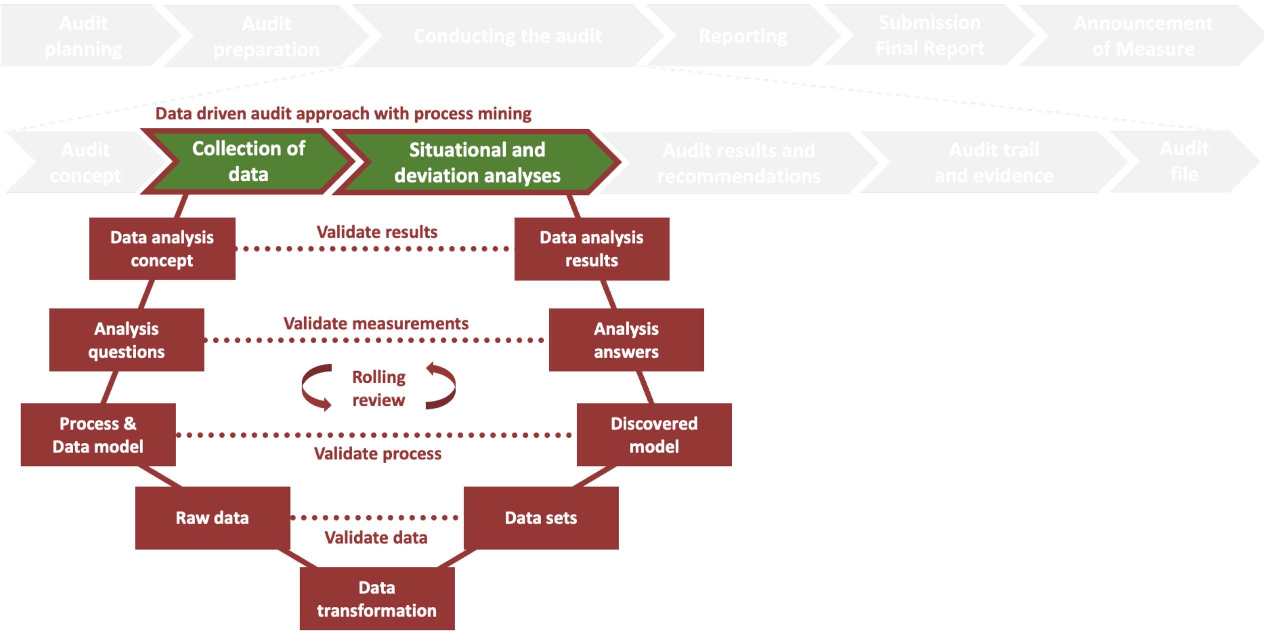


Figure 17: Rolling review with validation steps on multiple levels

Within this case study, we followed the 9-step model from Figure 3 to apply process mining in our audit. Throughout our journey, we experienced that the nine steps were not in a strict sequence. We frequently could use the things we learned in later phases of the project to improve deliverables from earlier steps. For example, we reworked our analysis questions multiple times as we gained new information regarding data availability and quality in later phases of the project.

Thus, the model has to be seen as an iterative approach where a rolling review contains validation steps on multiple levels. Figure 17 shows the most important validation steps.

First, we evaluated whether the data sets used as the input for the process mining analysis matched the raw data and the data from the productive system. Within this validation step, we ensured that we based our data analysis on a reliable data source.

Then, we validated the process itself by checking if the discovered model represents the process model defined in an earlier phase of the project. Within this step, we examined if we used the correct data or if there was any need to adapt the data model.

Next, we checked whether the analysis answers covered all the analysis questions. This way, we could check for any analysis question we had not answered yet, whether that was due to a lack of data or simply forgetting it.

Finally, we validated the results of the data analysis by checking if they met the requirements of the data analysis concept and if we had considered the primary audit objective sufficiently. With this evaluation, we could determine if the final analysis covered what we planned to audit.

All these validation steps helped us to get reliable results on a certain level of assurance and quality and improve the deliverables made throughout the process mining project.

Challenges and Limitations

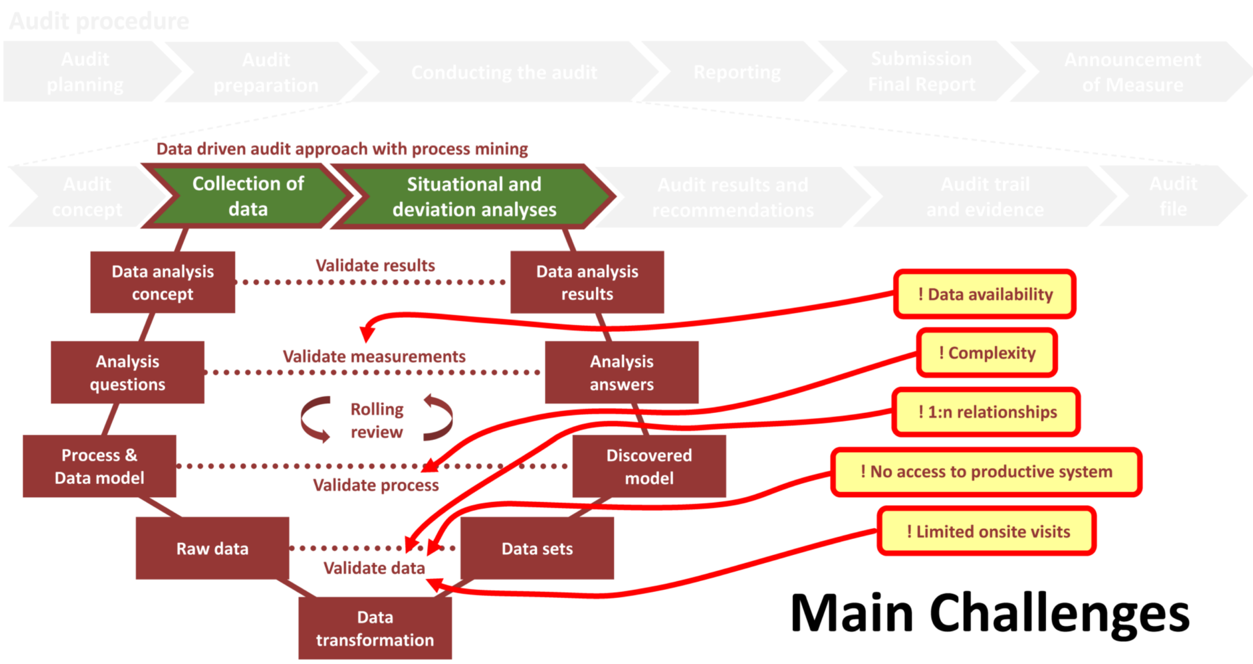


Figure 18: Challenges and limitations

We encountered a number of challenges during our process mining project (see Figure 18).

Data preparation was one of the main challenges. As previously discussed, only for some process steps we wanted to consider was data available. Thus, we could not answer all the analysis questions we had in mind in the first place.

The available data was spread over different tables, which had to be linked to each other. 1:n and n:m relationships made data preparation more complicated, and we had to implement multiple versions of the data transformation workflow before it provided reliable and validated data we were confident in.

The lack of direct access to the productive systems made data validation even more time-consuming because there was a dependency on the audited party to provide the data for cross-checking. Furthermore, as we performed the audit during the covid-19-pandemic, only limited on-site visits were possible due to contact restrictions.

The high complexity of the analyzed process and the vast number of events in the data sets made the explorative analysis quite challenging. As a result, analysis questions No. 1 (“Do the real processes fit the should-process?“) and No. 12 („Are there any bottlenecks within the process?”) were hard to analyze. We had to simplify the data to reduce complexity to a level that made the process analyzable. Ultimately, it was impossible to state a definite percentage of how many cases did or did not fit the should-be process.

The 1:n and n:m relationships made it necessary to work with different data sets for different analysis questions. As a result, we made our process mining analyses from the order perspective as well as from the invoice perspective.

Despite the challenges and limitations listed above, the process mining analysis gave us an excellent insight into how the purchase-to-pay process was performed in reality. Thus, the benefits of using process mining in the audit (see Figure 19) did exceed the challenges we needed to overcome.

Benefits of using process mining in an audit

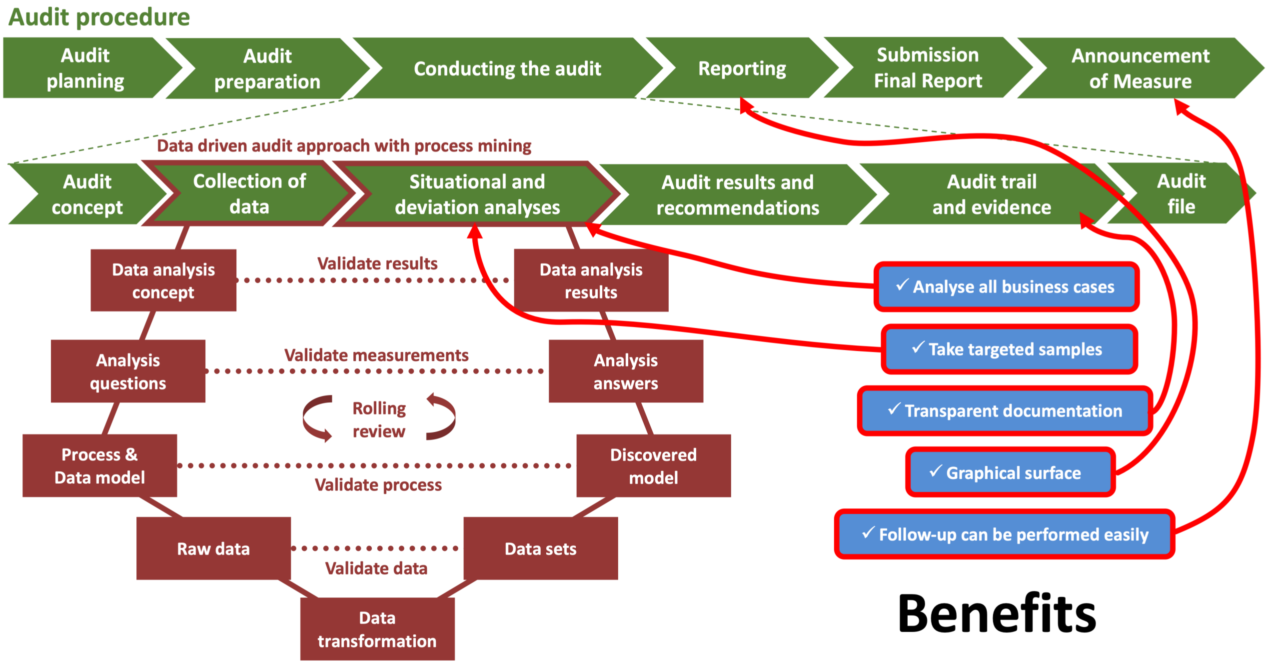


Figure 19: Benefits of using process mining in our audit

Within our audit work, we are often confronted with massive amounts of data and a high number of cases. As our resources are limited and we need to finish our audits within a specific timeframe, traditionally, we choose a sample of cases from the relevant process data and look at those cases in more detail.

Using process mining, we do not need to pick a sample anymore. Instead, we can analyze all cases regardless of the total number. For example, in the audit of this study, we could perform a complete examination of all 2,550 cases and give statements about order and invoice releases for all purchase orders from the year 2019.

We still used sampling techniques for a targeted investigation of those cases where we detected irregularities. This way, we allocate our resources more efficiently. From our experience, this led to a higher quality of the audit results and a higher assurance of detecting potential weaknesses in the internal control systems.

Furthermore, working with enormous amounts of data, it is often also hard to present the analysis results in a way everyone can follow. The graphical interface of the process mining software is very beneficial regarding this aspect. We could easily perform the analysis steps in attendance of the audited party to make transparent what we had done to come to the specific result. Of course, this benefit is even bigger when the audited party also has experience with process mining. In this case, they can retrace the analysis and measure if the changes they made due to the auditors' recommendations have the expected impact.

Finally, using process mining to evaluate the effect of changes made to the process can also be very beneficial from the auditor's perspective. If a follow-up audit is performed, process mining can be used once again to fully examine all cases and verify whether the implemented changes have improved the quality of the process.

As shown in this article, using process mining in an audit can be very beneficial and allows a deeper insight into the process of interest. Getting started with process mining in audit work is undoubtedly challenging, but it gets easier with more experience. We started using process mining in our audits in 2016 and have worked on improving our practice ever since. Every process mining project has been a new chance to improve our approach and make the audit trail more transparent.

With this article, we want to encourage other auditors to learn more about process mining and incorporate it into their audit method. We value exchanges with our peers and invite you to contact us to discuss your experiences with us via the contact details below.

Authors

Jasmine Handler, MA MSc - jasmine.handler@wien.gv.at (corresponding author)
Ing. Dipl.Ing.(FH) Andreas Preslmayr, MSc - andreas.preslmayr@wien.gv.at

References

[1] Stadtrechnungshofausschuss Periode 2021. Sitzung vom 02. Dezember 2021
URL: https://www.stadtrechnungshof.wien.at/ausschuss/05/inhaltsverzeichnis.htm

[2] StRH I - 8/20. WIENER STADTWERKE GmbH, Prüfung von SAP Beschaffungsprozessen. URL: https://www.stadtrechnungshof.wien.at/berichte/2021/lang/05-11-StRH-I-8-20.pdf

[3] KNIME Analytics Platform. URL: https://www.knime.com/knime-analytics-platform

[4] Process Mining in Practice. Many-to-many relationships in process mining: https://fluxicon.com/book/read/datasuitability/#many-to-many-relationships-between-different-case-ids

[5] Process mining software Disco. URL: https://fluxicon.com/disco/

[6] Process Mining in Practice. Milestone simplification method: https://fluxicon.com/book/read/simplification/#strategy-9-focusing-on-milestone-activities

[7] Process Mining in Practice. Semantic variant simplification method: https://fluxicon.com/book/read/simplification/#strategy-5-semantic-process-variants

[8] Process Mining in Practice. Variant simplification method: https://fluxicon.com/book/read/simplification/#strategy-2-focusing-on-the-main-variants