Process Mining for optimization of a P2P process of a company in the coatings and paints industry

Florian Guschl[1, 2], Christopher Tzakov[1, 3], Peter Lingner[1, 4], Albert Fichtenau[1, 5], Kira-Marie Liebig[1, 6]

[1] Leopold-Franzens-Universität Innsbruck, Innsbruck 6020, Austria
[2] Florian.Guschl@student.uibk.ac.at
[3] Christopher.Tzakov@student.uibk.ac.at
[4] Peter.Lingner@student.uibk.ac.at
[5] Albert.Fichtenau@student.uibk.ac.at
[6] Kira-Marie.Liebig@student.uibk.ac.at

Abstract. Business operations and the processes behind them usually do not run smoothly. Why this is the case is often difficult to determine in large companies. Process Mining is a method to solve tasks of this range. The Business Process Intelligence Challenge offers the possibility to practice process mining techniques with real data. Hence, this paper analyzes a real-life event log of a Dutch organization, operating in the field of coatings and paints. First, all processes are modelled according to the BPMN 2.0 standards. For the subsequent challenges the purchase-to-pay process, particularly focusing on 3-way-matching, invoice before goods receipt is analyzed. The data is additionally restricted to the top ten vendors in the area of packaging for the year 2018. This paper aims to identify optimization possibilities in the area of compliance and throughput time. The tool Celonis is used throughout the analysis. The most relevant bottleneck for all of the top ten vendors is the direct process flow from the activity Record Invoice Receipt to Clear Invoice, which makes up a major part of the throughput time. The majority of the top ten vendors has compliance issues, hence recommendations for areas of improvement have been formulated.

Keywords: Process Mining · BPIC 2019 · Purchase to Pay · Compliance · Celonis.

1 Introduction

Process mining techniques allow for extracting information from business data and can be used to discover models describing processes, organizations, and products [1]. Relevant information of companies is often stored in so called event logs and enable the possibility of process mining [2]. Event logs contain information like Case ID, the timestamps of the start and end times, and other attributes of the event recorded by an IT system. Thus, an event log maps one or more cases of a business process but can also be a documentation of several, related business processes.

Process mining has experienced some major development over the last decade because companies need to learn how their processes perform in the real world [3], and the amount of data collected via information systems has increased substantially [4]. This discipline is the focus of this year's Business Process Intelligence Challenge (BPIC) organized by the Process Mining Conference 2019. For this challenge participants are provided with a real-life event log from a large multinational company operating from the Netherlands in the area of coatings and paints. The purchase order handling process for some of its 60 subsidiaries should be investigated, with special attention towards compliance questions.

Our approach to the challenge is structured as follows. This chapter explains the structure and research questions of this paper. The following chapter describes our used tools, gives an overview about the provided data set and clarifies the research objective and the used methods. The third part is about the three main challenges stated by the company. To solve the first challenge, the following question is answered:

1. How can the underlying processes be modeled according to BPMN 2.0 standards?

This helps to get an overview and first understanding of the process. Since we participate in the BPI Challenge as a team of the student category, we focused on specific cases for the second and third question. Those are:

2. How is the invoicing process performance in terms of throughput time affected by different process characteristics?

3. Can particular purchase documents of the invoicing process be successfully identified as potentially improvable irregularities? Which vendors are mainly responsible for deviations, regarding compliance?

The process to restrict the data on certain factors is described in detail in the next chapter. In order to generate meaningful insights, we further focus on issues regarding the invoicing process, with the major activity’s vendor creates invoice, record invoice receipt and clear invoice. To structure possible findings in a meaningful way, we formulated the following research question that will be answered in the chapter of the respective challenge: Are there any optimization possibilities in the area of compliance and throughput time regarding the top 10 vendors?

2 Description of the data and methods

The provided and in this paper analyzed dataset is a fully IEEE-XES compliant event log and contains 76,349 purchase documents containing in total 251,734 items. Hence,

there are 251,734 cases. In these cases, there are 1,595,923 events relating to 42 activities performed by 627 users (607 human users and 20 batch users). The vast majority of events took place in 2018. The dataset contains around 1.5 million events for purchase orders (PO) of the dutch company submitted in 2018. The data shows the purchase-to-pay process (P2P) without the approval workflow of the PO's and the invoices. The data refers to many different categories of goods and services and include many different types of vendors.

The assignment in the student category is to select a specific aspect of interest. We found an area playing an important role for the company. This is the spend area of ordered goods related to packaging. As seen in figure 1, we selected the top 10 vendors in this area by the number of cases. They are involved in around 67,000 cases that correspond to 26% of the total cases in 2018.

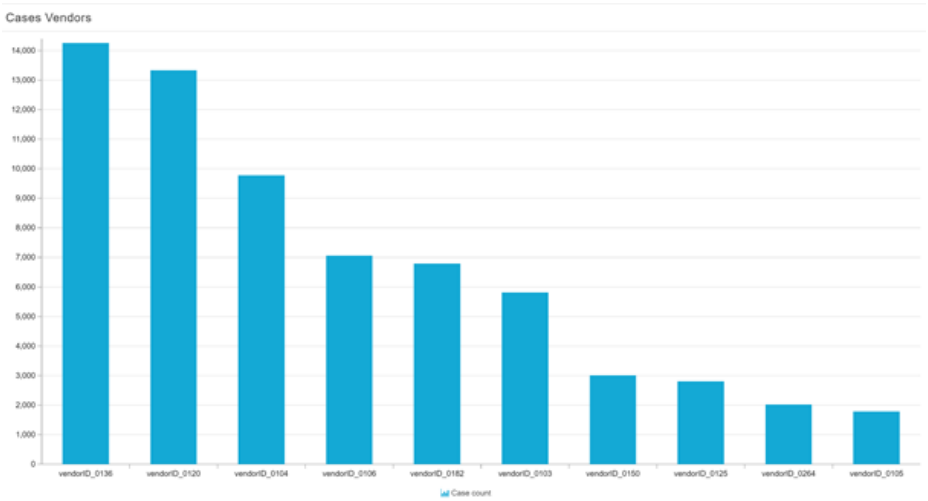

Fig. 1. Top 10 vendors in the spend area packaging (Own Depiction).

The main tool used for the analysis is Celonis. It is a suitable tool to investigate cases and events in detail but at the same time not losing the overview of the processes. Additionally, Microsoft Access and Microsoft Excel are used to conduct special investigations for a small number of cases but where more specific calculations are required.

3 Analysis of processes regarding the stated three challenges

3.1 Modeling of the main process flows

In this section an overview of the main processes is provided establishing a foundation for the subsequent analysis. The process flows are modelled according to the BPMN 2.0 standards and were simplified, so the most relevant path is depicted. The vendor is modeled as an empty lane as the majority of the vendor process model would be based

on presumptions. Thus, the models, depicting the happy path, contain almost no message flows between the company and the vendors. Only "Vendor creates invoice" (VCI) was associated with the vendor as an activity for the purpose of illustration.

3-way matching, invoice after goods receipt

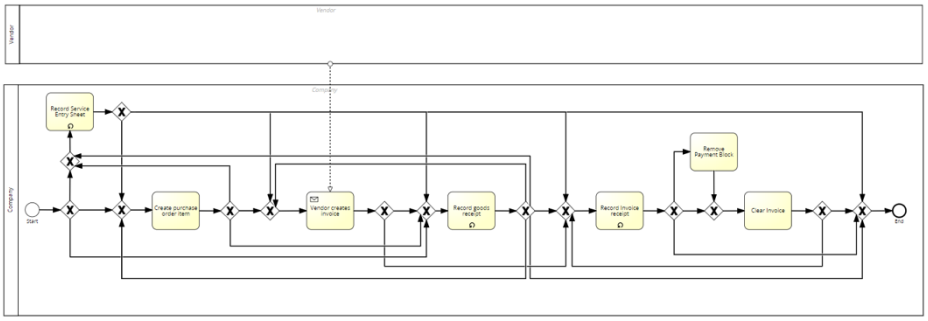

Fig. 2. BPMN 2.0 model of "3-way matching, invoice after goods receipt" (Own Depiction).

The "3-way matching, invoice after goods receipt" process flow consists of 15,182 cases (make up for 6% of total cases). The main focus in this process is that the "Record Invoice Receipt" (RIR) happens after "Record Goods Receipt" (RGR). For these items, the value of the goods receipt message should be matched against the value of an invoice receipt message and the value put during creation of the item. In figure 2 the process is modelled with 96.2% of all activities and 91.6% of all connections. The process usually starts with the creation of a purchase order item, followed by the creation of the invoice by the vendor. Then, the goods' receipt is recorded, followed by the record of the invoice receipt. In the end, the invoice is cleared. This most common path is also called the happy path and happens in 15.3% of the cases.

3-way matching, invoice before goods receipt

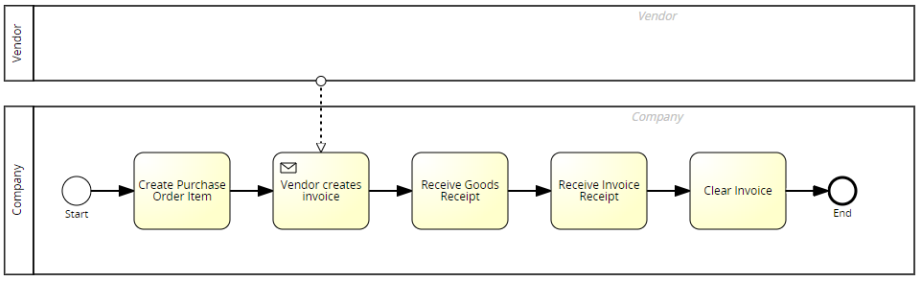

Fig. 3. BPMN 2.0 model of "Happy Path: Three-way matching: invoice before goods received" (Own Depiction).

The "3-way matching, invoice before goods receipt" process flow consists of 221,010 cases (make up for 87.8% of the cases). The items require a goods receipt message,

while they do not require Goods Receipt-based invoicing. Invoices can be entered before the goods are receipt, but they are blocked until goods are received. This unblocking can be done by a user, or by a batch process at regular intervals. Invoices should only be cleared if goods are received and the value matches with the invoice and the value at creation of the item. The happy path (figure 3) for this process flow starts out with the purchase item being created and presumably send to the vendor. Upon receiving the order, the vendor creates the invoice, which is likely documented through a confirmation being send to the company. Next, the company receives the order, which is documented through a goods receipt, followed by the invoice receipt from the vendor. The last step in the happy path is the clearing of invoice. This happy path accounts for 47,957 of the 221,010 cases falling under the selected category. This means that roughly 21.7% of the cases are represented. The reason for such a low percentage is, that the process contains a vast collection of activities and connections which occur only infrequently, but which cause the majority of the cases to deviate from the happy path in terms of extra activities.

Interestingly, the goods receipt actually arrives before the invoice receipt in the happy path, which is counterintuitive given the name of this category. Filtering for cases in which the invoice does in fact arrive before the goods received leads to an entirely new happy path. The happy path also starts with the purchase order item being created, moving on to the vendor creating an invoice, which again we presumably know due to a confirmation message, which is followed by the invoice receipt, which in turn leads to the goods receipt which causes the payment block to be removed leading to the payment of the invoice. This new path is only present in 14,830 cases and the following happy path represents 59.6% of these (8,836), as such not being very representable for the 221,010 cases categorized as invoice before goods receipt.

Comparing the probabilities or rather frequency in which the two paths occur leads to the conclusion that while the invoice may be created early on this is probably only derived from an order confirmation message or later from the invoice date and it does not actually suffice for proceeding to the payment stage because in most cases the company only clears the invoice after another process step called: record invoice receipt – and even when both events occurs before goods receipt in only 685 of the 221,010 cases the invoice is actually cleared before the goods receipt is booked, which might indicate that these cases represent exceptions.

2-way matching (no goods receipt needed)

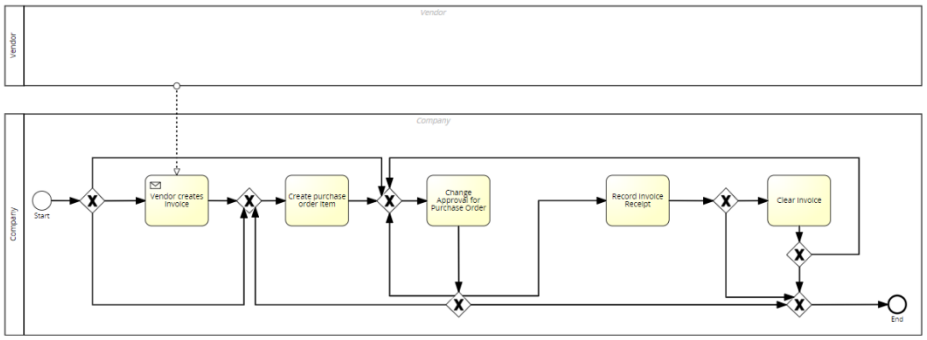

Fig. 4. BPMN 2.0 model of "2-way matching (no goods receipt needed)" (Own Depiction).

The "2-way matching (no goods receipt needed)" process is the process with the least amount of cases. It consists of 1,044 cases (makes up for 0.4% of total cases). For these items, the value of the invoice has to match the value at creation, but there is no separate goods receipt message required. In figure 4 the process is modelled with 99.1% of all activities and 94.7% of all connections. Due to the relatively small amount of cases it is not purposeful to model the process in even more detail. In contrast to the two 3-way matching processes, 2-way matching does not require a goods receipt and it inherits the activity "Change Approval for Purchase Order" (CAPO). Usually the process starts by either a vendor creating an invoice, creating a purchase order item or change approval for purchase order. The last two options mentioned, happen in every single process. If a vendor creates an invoice, then an invoice receipt is recorded in the end.

Consignment

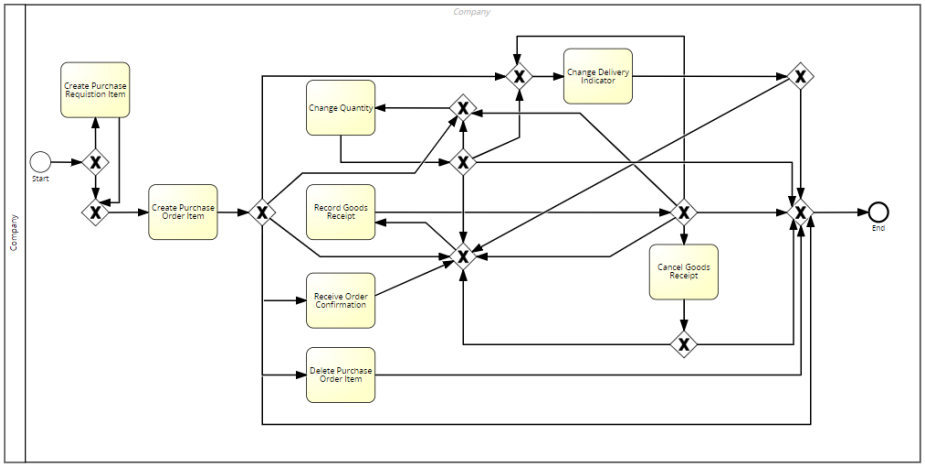

Fig. 5. BPMN 2.0 model of "Consignment" (Own Depiction).

The consignment process consists of a similar amount of cases as the 3-way matching, invoice after goods receipt process (14,498). The major difference to all other processes is that no invoicing is required. Instead the goods are already owned by investigated company and only shipped from the vendor. Vendor_0188 makes the most significant contribution with almost a third of all cases. Due to the lack of invoicing required the happy path is simply the creation of purchase order items followed by recording goods receipt. 60.5% of all cases follow the happy path. In figure 5 the consignment process is modelled with 99.7% of all activities and 99% of all connections. The next more noteworthy activity is the creation of a purchase requisition item before the actual creation of the purchase order item. After the creation of purchase order items, a change in quantity is required before the goods receipt activity which sometimes re-occurs before the process ends. All other parts of the process refer to less than 1,000 cases.

3.2 Investigation of the invoicing process

Throughput Time

To answer the second question as formulated in the introduction, the same selection process is chosen as already explained in chapter one. To investigate the average throughput time of cases inheriting specific activities, the data set is modified to only depict cases including the activities “Record Goods Receipt” (RGR) / “Record Invoice Receipt” (RIR) and “Clear Invoice” (CI). This selection matches 20% of all the cases analyzed.

The average throughput time for the cases involving the activities RGR / RIR and CI is 100 days, as presented in figure 6. The Happy path for the selected cases is a process consisting of five activities: “Create Product Order Item” (CPO) → RGR → VCI → RIR → CI.

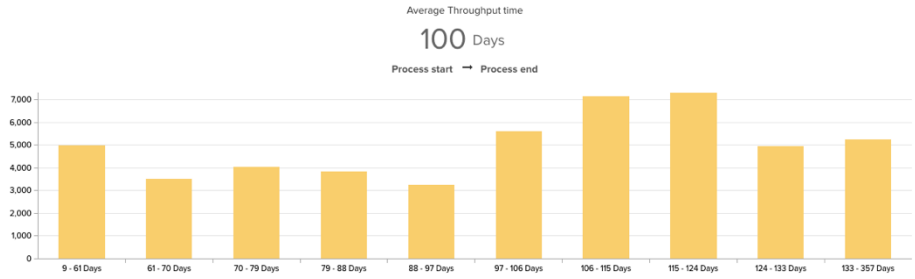

Fig. 6. Illustration of Average Throughput Time for cases involving the activities record goods receipt/ record invoice receipt and clear invoices (Own Depiction).

The average throughput time for the last two activities (RIR → CI) is 75 days affecting 67% of the cases. These two directly following activities constitute the slowest bottleneck. How this can be interpreted depends on the company’s internal defaults. However, the investigated company could reflect upon why this process takes such a

relatively long time. When analyzing whether there is a specific user who does not pay or does not pay fast, we can see that more than 50% of the cases are inherited by an unknown user (user_NONE), where the activity CI does not flow through the process. User_NONE indicates no user was recorded in the source system making the finding difficult to investigate further. Therefore, we consider the two directly following largest user, user_235 and user_029 as shown in figure 7.

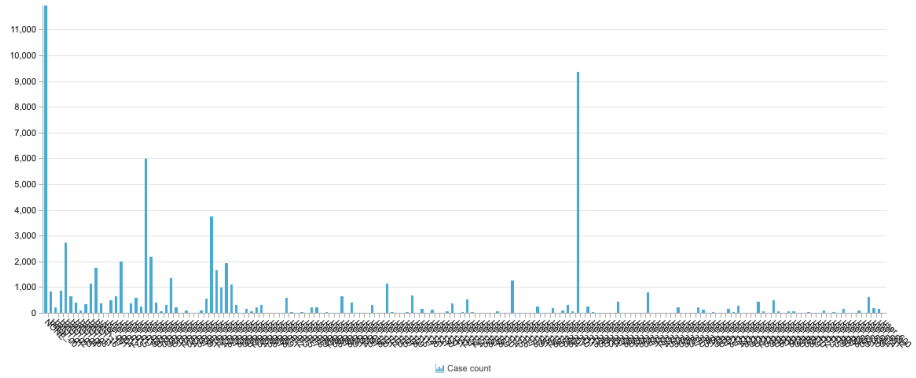

Fig. 7. Number of Cases depending on Users for cases that have not been paid (Own Depiction).

For user_235 we found that most cases which are not paid, start with the activity "Create Product Receipt" (CPR) while the cases which do contain CI, start with CPO.

For the vendors we found, that vendor_0136 and vendor_0120 are mostly involved in cases which do not flow through the activity clear invoice. But as depicted in figure 8 there is no great distance between most of the vendors.

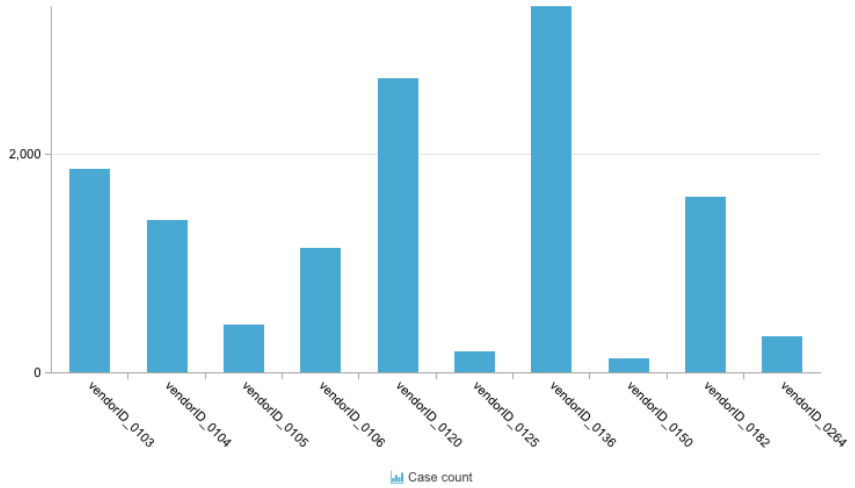

Fig. 8. Number of Cases depending on Vendors for cases that not have been paid (Own Depiction).

To answer the question, whether there are users which do pay relatively fast, we analyzed the throughput time depending on the users. We found no big difference between most of the users, while batch_01 sticks out as their throughput time is twice as long, shown in figure 9. This could be a result of the batch being many users, making coordination relatively hard, or this could also be a default instruction to hold back payments by the company investigated.

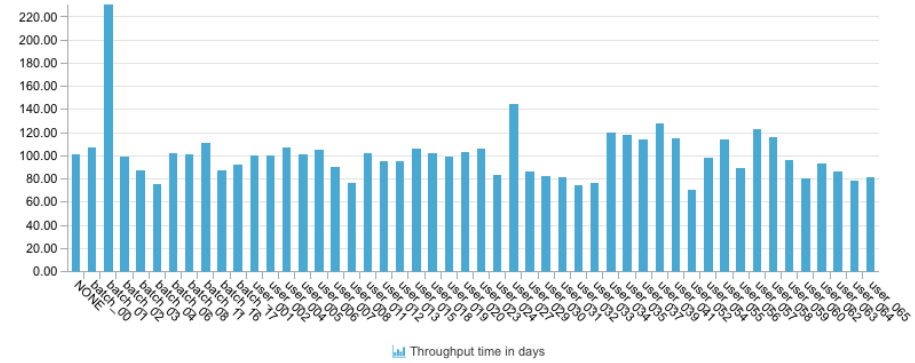

Fig. 9. Throughput time in days depending on the users for cases that flow through the activities Record Goods Receipt/ Record Invoice Receipt and Clear Invoices (Own depiction).

Furthermore, we investigated whether there is a specific user or vendor receiving or creating the activity, “Vendor Creates Debit Memo” (VCDM). We found that user_002 is involved in the biggest share of VCDM activity, as shown in figure 10. While this finding could also result from the fact of user_002 being in charge of all payments or a related company structure reason it might be worthwhile to focus on this user in future research.

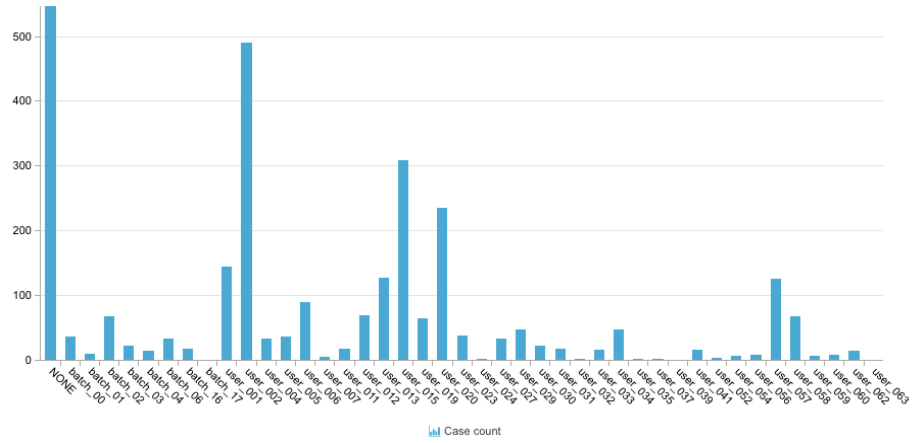

Fig. 10. Number of cases depending on users for cases involving the activity vendor creates debit meme (Own Depiction).

More interesting is the finding of the vendor_0136 creating with great distinct the most debit memos as shown in figure 11. Here one could interpret the vendor_0136 may have different regulation, for after which time limit the debit memo is sent out. To optimize this process, the investigated company could get in contact with the vendor and find a solution on how to minimize the debit memos. This could be of advantage, as the process gets more complex and employees have to file the memo within the internal database.

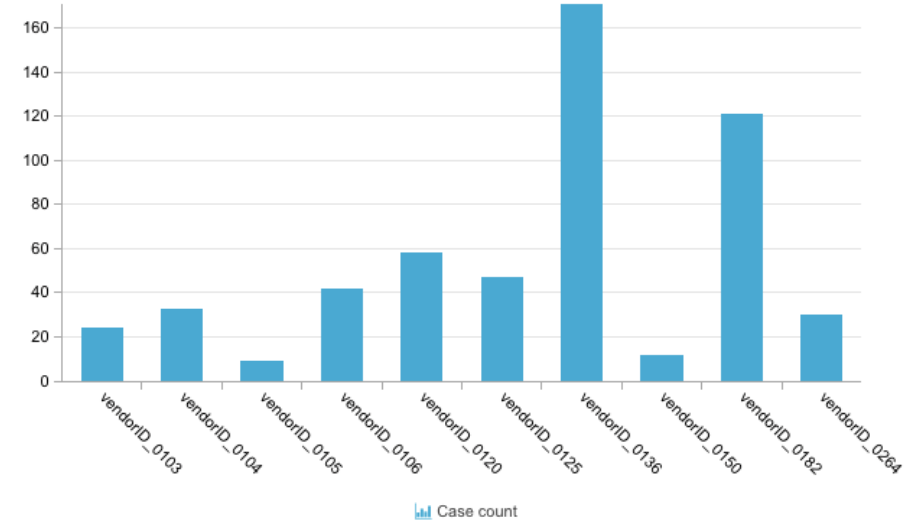

Fig. 11. Number of cases per vendor for cases involving the activity vendor creates debit meme (Own Depiction).

Examination of the top ten vendors

With our research question being “Are there any optimization possibilities in the area of compliance & throughput time regarding the top 10 vendors? “, this chapter focuses on the analyzes of the top ten vendors.

Therefore, we investigated the conformance for each vendor. Hence, we identified the respective happy path and compared it to the as-is process. We chose the happy path as it is the path that is executed most often, assuming the process to be the desired to-be process. The following chapter is sorted from the vendor inheriting the most cases to the vendor inheriting the least cases. We only investigate the invoicing process; therefore, we select only cases where the activities RGR / RIR and CI are included.

To generalize we investigated 50% of the top ten vendors mainly follow the happy path: CPO - RGR - VCI - RIR - CI (Vendor_120, Vendor_182, Vendor_150, Vendor_125, Vendor_264). Additionally, we find that 40% follow almost the same happy path only inserting the activity ROC as the second activity: CPO – “Receive Order Confirmation” (ROC) - RGR - VCI - RIR - CI (Vendor_104, Vendor_106, Vendor_103, Vendor_105). Only the Vendor_136 has a different happy path with the activity VCI following directly the first activity CPO: CPO - VCI - RGR - RIR - CI.

Looking at the top 10 vendors in the packaging spend area, it can be said that all vendors have issues with regard to the invoicing process but the amount is manageable. Vendor_136, Vendor_104 and Vendor_106 follow the happy path to a comparatively low degree. The process after the invoice is received until the invoice is paid takes a long time in almost all cases. However, as suggested in chapter 3.2 this could be due to internal regulations until invoices should be paid. The ten vendors are now briefly described individually with the individual most important findings regarding the payment process.

Investigation Vendor 136
The average throughput time for this vendor is 126 days. 20.99% of the process flows follow the happy path. This is the only vendor where the activity VCI follows directly CPO. The two largest bottlenecks are two activities that occur before CI: RIR (46% of cases) and RPB (52% of cases), which extend the process by 97 and 89 days respectively. 54% of the vendor's cases are not compliant with the Happy Path, as Remove Payment Block (RPB) occurs. This extends the process by three days. A root cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

Investigation Vendor 120
The bottleneck increasing the throughput time most considerable is the activity RIR → CI. This process takes 89 days in 76% of the cases. For this process no user can be identified that takes the longest with great distance, while for the whole process user_073 takes the longest with great distance. Opening up the opportunity of an internal investigation for the user_073.

Investigation Vendor 104
The happy path is represented by 30.16 % of the cases. Performing a conformance analysis based on deviations to the category happy path leads to the detection of several activity sets.

For 90% of the cases the step ROC occurs, however while adding on average 1.9 additional steps to the process the throughput time is reduced by 3.9 days on avg. and thus this violation is whitelisted.

For 65% of the cases the process flow deviation of RGR following CPO is detected. This not only increases the process complexity by 1 step on average, but also causes the duration to be 4.3 days longer. Performing a root cause analysis shows, that there are three users which together participate in close to 50% of the deviations. These users are user_065, user_056 and user_062. Because user_065 alone participates in close to 25% of the deviations the corresponding process is analyzed. A time-based process visualization shows that for no apparent reason there is a 62-day delay between receiving the invoice and clearing the invoice when user_065 participates. This affects 532 cases. Opening up the opportunity for an internal investigation.

This observation also holds true for the other two users. A delay of 64 days for 324 cases of user_056 between the steps RIR and CI is found.

The third significant deviation, occurring for 39% of the cases, is the occurrence of the undesirable process step RPB. In most of the cases goods are received before the vendor created the invoice and this invoice is received. Naturally payment can't occur without an invoice. However, looking closer shows that most of the time delay is caused by the 1516 cases that connect RIR and RPB. This step takes 38 days on average. Overall this deviation causes the process to take 7.9 days longer on average.

Investigation Vendor 106

The average throughput time for vendor 106 is 73 days and only 28.56% of all 6,310 cases flow through the happy path. Vendor 106 is the main vendor for metal containers and lids below 30L. In the analysis receiving the order confirmation is identified as part of the process which is mostly executed by user_29. In this process this is the only activity he performs. It decreases the throughput time on average but increases the number of steps by 1.6 per case. Cases without an order confirmation increase the duration of the process by an average of 27 days. Most likely the order handling on the vendors side is not consistent confirming the order. When the order is confirmed the goods are usually received faster. Additionally, it has been noticed that the invoice receipt for different orders is recorded more than once at the same time, sometimes by the same and sometimes by different users. In many cases, when the order is not confirmed by the vendor, first the goods and then the invoice is received. This deviates from the happy path where usually first the goods are received before the vendor creates and sends the invoice.

Investigation Vendor 182

The happy path for vendor 182 is represented by 50.31% of the cases, with an average throughput time of 111 days. The process step 'RIR to CI increases the throughput time most considerably by 81 days in 79% of the cases. An interesting finding is that the step RPB is missing here opening up the opportunity for further investigation why this step is missing.

Investigation Vendor 103

The average throughput time for vendor 103 is 108 days. The happy path is followed by 65.13% of the 3763 cases. As with vendor 182, the bottleneck is the process step RIR to CI but for this vendor increasing the throughput time even more by 99 days in 89% of the cases. Here the step RPB is missing. Another step increasing the throughput time considerably by 93 days for 10% of the cases is the process step RPB to CI. The analyzed violations for this process are, e.g. the undesired activity 'receive order confirmation'. A deeper analysis disclosed many possible root cause violations. User_029 is also accountable for 3,000 violations and therefore could be further investigated.

It appears for 92% of cases increasing the path by 6.3 steps and 108 days. Another relevant violation is that RIR is followed by CI adding 6.3 steps per case and 109 days to the process in 89% of the cases.

Investigation Vendor 150

Vendor 150 mostly sells labels and appears to be a relatively efficient vendor with almost 78 % following the happy path. The average throughput time is 61 days and most of the time is taken by the step from RIR to CI with 40 days. The step slowing down the process by 12 days is RPB 12% of all cases of the vendor. User_15 and user_6 are mostly involved in these cases. In these cases, it could be worth investigating why the payment block is instantiated and kept for several days. Additionally, there are several cases where either user_13 or user_15 record the receipt of invoice again after a long time, after it has already been done by batch_01. This indicates some error in the processing of the order since in almost all cases of these cases the next step is to remove the payment block. They appear to get lost along the process and are "reactivated" later. Vendor_150 waits a long time until reminding for payments since the next invoice is only sent after an average of 103 days and it took the vendor itself a long time until it sent the initial invoice.

Investigation Vendor 125

The happy path is represented by 52.46% of the 2,377 cases. The most frequent violation compared to the categories happy path is that CPO is followed by RGR instead of VCI. This raises the duration on average by 8.6 days and increases the complexity by only 0.6 events. The violation affects 82% of the vendors activities and a root cause analysis shows that over a third, are events in which the user_058 participates. Interestingly while the goods are received before the invoice for most of the violations with this user, barely any payment blocks occur. In general, for this violation the frequency of process steps which take longer than the average seems to be low.

Another violation, similar to the violation of vendor_104 is the occurrence of the undesirable event RPB. The observations regarding the process structure also do not deviate from the observations concerning vendor_104's violations. This causes the process to take 8.3 days longer on average, increasing the complexity by 1.6 events on average, compared to non-violating events, for 26% of the cases.

Investigation Vendor 264

For vendor_0264 65.09% of the cases flow via the happy path. Here, we want to point out the activity RGR being in front of VCI for most of the cases. The average throughput time is 98 days, and just as for almost any other vendor, the most significant bottleneck is the RIR → CI taking 51 days. No relevant violations were found. The most products are labels and Metal Containers & Lids < 30L and packaging other.

Investigation Vendor 105

The average throughput time for this vendor is 64 days, which is about half of the time of the biggest vendor. 76.85% of the cases follow the happy path. This happy path is similar to the one of e.g. Vendor_0136, however, the ROC is missing for Vendor_0136. The two largest bottlenecks are two activities that occur before clear invoice: RIR directly before CI (93% of cases) and RGR before CI (97% of cases), which extend the process by 37 and 24 days respectively. 54% of the vendor's cases that are not compliant with the Happy Path, as RPB occurs. This extends the process by three days. A root

cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

3.3 Invoice, compliance and deviations investigation

As a first step in the compliance investigation a definition of a compliance deviation is required. In this paper every transaction in which the net value of a case changes for inexplicable reasons is considered a compliance irregularity. Thus, in order to identify problematic invoices, irregularities were identified in three different analysis sections.

For this analysis the otherwise established scope was disregarded. This decision has been done, as the analysis of the entire dataset is not significantly more complex with the methods chosen, than the analysis of a subsection. Consequently, the workload-oriented reasoning for a reduced scope does not apply to the dealt with sub-task.

In the first compliance analysis, we calculated the average of the cumulative net worth per case and compared it to the median worth of the case in Celonis. In case the two values would not be the same it indicates that the cumulative net worth is not the same in all events of the case.

As seen in figure 12, in total these cases account for less than two percent of the total number of cases, while they accumulate to almost 18% of the total number of events. This finding represents that these cases generate a significant amount of extra work.

The vendor with the most activities is vendor_234 with almost 50,000 activities distributed on 207 cases. Looking more closely at vendor_234, no invoice receipts have been recorded. The vendor with the most cases is vendor_233 with more than 300 cases and around 4,000 activities.

However, with regards to compliance, not all of the 4,632 cases do necessarily have a compliance issue, since several goods receipt messages or several invoice clearings could match together for a compliant process. This concludes the first analysis.

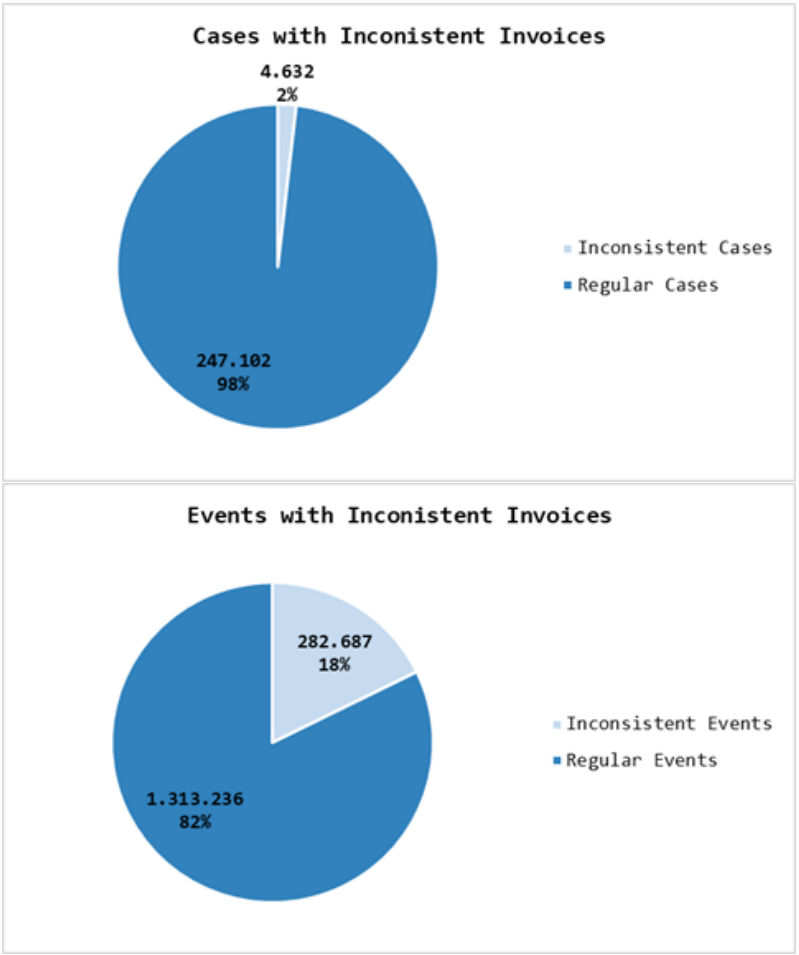

Fig. 12. Comparison of the number of events and number of cases with and without invoice issues (Own Depiction).

The second compliance investigation is conducted with Microsoft Excel and Microsoft Access of the Microsoft Office 365 suite. Before the dataset is analyzed in Excel, Access is used to reduce and compartmentalize the data into excel compatible subsets. Next these subsets are imported into Excel and potential compliance issues are identified with a rough screening of whether all individual events, of a case match in terms of net value. Through this selection step, the initial data volume is reduced by 85% to a set of 242,020 unique events or 4,632 cases.

This screening and subsequent volume reduction enabled the detailed analysis with the available computational power. In the next step of the screening process the minimal net value is identified for each case and then the sum of net values is computed for each case, followed by the calculation of the remainder of the division of the sum of net values by the minimal net value of each case. This step allows the immediate exclusion

of all those cases in which e.g. the items are processed separately and then merged together causing the net value of the subsequent event to spike (deviate), even though there is no compliance issue. This step reduces the amount of potential compliance irregularity cases to 4,019, excluding 613 of the candidates. Next the remaining value of the division is divided by the count of events per case. If the result of this is less than 1, the deviation is likely caused by rounding differences and thus the candidate is excluded. This leaves 1,425 cases consisting of 39,199 events for further analysis.

As the next and final step in the selection process the remaining cases are screened for unique combinations of net value and case ID, followed by a calculation between the different events of the cases, sorted in a date based ascending order. Computing this change for cases allows for the evaluation how much deviations occurred between the different values for vendors.

The result is the identification of 28 vendors who are responsible for the possible compliance deviations. Some of these vendors are responsible for comparably many events but few cases, some for few cases but these contain comparably high net value changes. Based on this analysis a definitive categorization of compliance violating vendors is not possible, but it identifies the vendors which need to be analyzed further. A selection of the mentioned data points can be seen in table 1.

It can also be observed, that the identified, potentially violating cases are to more than 99% of the case item category “3-way match, invoice after GR” and to more than 99% concerning the case Spend area text “Logistics”. This might mean that instead of identifying potential compliance violations, the applied selection filters identified characteristics of these particular categories.

As an alternative investigation the cases containing RGR or CI are analyzed.

Table 1: Main vendors responsible for compliance deviations

VENDOR	CUMMULATIVE CHANGE (>0)	CUMMULATIVE CHANGE (<0)	AVG CUMULATIVE NV	COUNT EVENTS
vendorID_0201	2,6788E+12	0	8,92936E+11	3
vendorID_0230	2,89948E+11	-2,89946E+11	1685828475	172
vendorID_0231	410151	-1587588	23522,69712	208
vendorID_0232	5563066	-5335410	99492,82051	78
vendorID_0233	22671608	-60770369	195867,9716	352
vendorID_0234	906441	-742950	198463,2222	18
vendorID_0235	4165972	-3788390	11082,33686	1514
vendorID_0244	523375	0	66064,5	10
vendorID_0330	3,90576E+12	-4,26083E+12	4,84188E+11	11
vendorID_0363	9986920	-6219675	27912,65181	718
vendorID_0364	13173178	-7411052	34740,47342	790
vendorID_0365	7401349	-3344944	17836,12642	617
vendorID_0366	792380	-102440	2662,709877	324
vendorID_0388	0	-1,72138E+12	5,73797E+11	5
vendorID_0470	895930	-989147	7905,188153	287
vendorID_0471	3115	-2256	571,4615385	13
vendorID_0472	4148408	-4020389	28042,66392	848
vendorID_0473	18535150	-18280293	35909,28633	1383
vendorID_0534	25658848	-54941723	194569,5029	1040
vendorID_0535	2262953	-20316230	125991,9216	204
vendorID_0536	5674676	-22298552	139294,9255	443
vendorID_0537	1455558	-9008432	64520,88333	180
vendorID_0538	17399799	-76445802	369888,6975	519
vendorID_0539	10692699	-16927479	77928,1352	429
vendorID_0540	6421962	-23458070	149774,8382	346
vendorID_0541	3174116	-29793713	189900,0261	307
vendorID_0595	170	-100	51,14285714	7
vendorID_0977	300692	-304862	5073,844444	270

In the final analysis the same tools are utilized, however a different approach is chosen. Similar to the previous analysis cases are analyzed towards significant deviations between the net values but in this analysis the net values of the RGR and CI events are compared. As an initial selection we identified 51,672 cases, that have at least one RGR event, but no CI event and 870 cases with the opposite event constellation, both were excluded from further analysis. Next, if the sum and average of RGR net values deviates from the sum and average of the CI net values and the minimum RGR net value does not explain the CI net value sum or average the case is checked for potential rounding deviations. Only cases for which all these tests are conclusive, a final analysis step is conducted. This last step includes the computation of the duration and identification of the corresponding vendor.

This results in the identification of 51 unique vendors responsible for 1,003 cases consisting of 53,682 events that might represent conformance violations. Vendor_0230 is involved in 208 of these cases and 19,272 of the events, in total the deviation of the average RGR and CI net values amounts to an absolute value of 13,180,269,844€. Vendor_0230 dominates the analysis results and thus was excluded from the visualization of the deviations in figure 13.

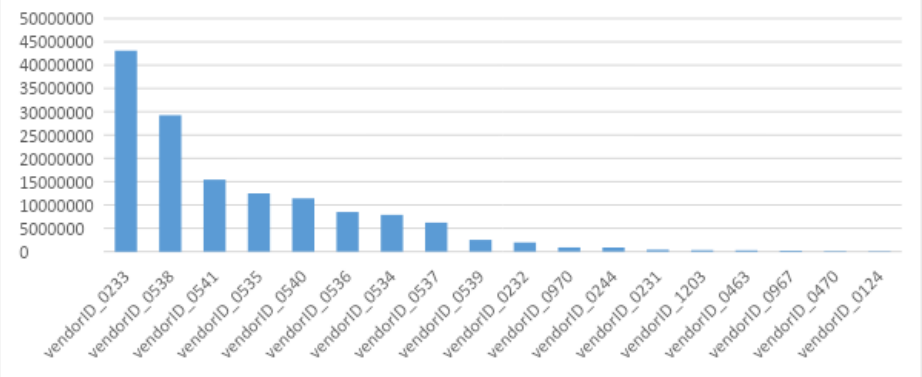

Fig. 13. Total absolute deviation with Create Invoice and Record Goods Receipt (Own Depiction).

It is noteworthy that approximately 99% of the potential compliance deviations identified in this analysis are of the document type “Standard PO” and of the case item category “3-way match, invoice after GR). Also, only 2 cases consisting of 24 events are part, less than 1%, are part of the spend area “Packaging”.

In future research it might be worthwhile to conduct a more detailed analysis of these 51 vendors, some data points concerning the potential compliance deviations grouped by vendors can be found in table 2.

Table 2: Vendors with potential compliance deviation with focus on Create Invoice and Record Goods Receipt.

VENDOR	ABS DEV OF AVG NV	AVG DURATION IN DAYS	COUNT CASES	COUNT EVENTS
vendorID_0230	13180269844	1	208	19272
vendorID_0233	43131753,33	36	73	165
vendorID_0538	29226585,29	7	90	514
vendorID_0541	15433107,28	9	48	234
vendorID_0535	12485612,83	13	48	146
vendorID_0540	11420484,64	4	51	282
vendorID_0536	8492909,884	5	47	407
vendorID_0534	7875849,22	1	47	1594
vendorID_0537	6213041,5	24	42	124
vendorID_0539	2510849,718	3	64	1057
vendorID_0232	1953079	43	18	47
vendorID_0970	891915,6923	14	1	24
vendorID_0244	875152	11	11	22
vendorID_0231	395524,4459	8	101	1443
vendorID_1203	303109	18	2	4
vendorID_0463	271586	23	2	4
vendorID_0967	216576	11	3	6
vendorID_0470	133337,9218	8	80	1174
vendorID_0124	107339	40	3	6
vendorID_0812	80645	38	2	4
vendorID_0295	69785	41	2	4
vendorID_0433	60265	36	3	6
vendorID_0183	58050	30	2	4
vendorID_0684	56119	43	2	4
vendorID_0514	42535	40	1	2
vendorID_0279	41093	36	2	4
vendorID_1344	32445,66667	3	1	10
vendorID_0117	31870	28	1	2
vendorID_0658	26025	36	1	2
vendorID_0166	25548,33333	21	2	7
vendorID_0653	23944	18	1	2
vendorID_0271	21733	11	1	2
vendorID_0502	21040,5	29	3	9
vendorID_0483	16550	32	2	4
vendorID_0141	16420	32	1	2
vendorID_0314	11867	34	3	6
vendorID_0171	11601	17	1	2
vendorID_0496	11068	30	1	3
vendorID_0461	8678	45	1	2
vendorID_0195	6826	36	1	2
vendorID_0062	6287	42	1	2
vendorID_0712	6100	10	1	2
vendorID_0467	3880,903656	2	6	101
vendorID_0151	3457,578947	3	1	67
vendorID_0595	2986,5	40	12	32
vendorID_1094	2663	12	1	2
vendorID_0125	1939,75	24	1	7
vendorID_0214	1359	39	1	2
vendorID_0111	555,8571429	12	1	8
vendorID_0119	520	11	4	8
vendorID_0182	440	54	1	2

4 Acknowledgements

All Authors contributed equally to this paper. We would like to express our gratitude to all sponsors for making the BPI Case Challenge 2019 possible, thus facilitating creative approaches for real Process Mining problems. We would especially like to thank Celonis SE for providing us with licenses of their Process Mining Software which was our main tool to conduct our findings.

5 Conclusion

This paper outlines the underlying purchase process flows of a Dutch company operating in the field of coatings and paints. We analyzed and depicted the different processes in order to generate a solid understanding of the company's processes. As a first step irregularities in the data sample were identified and used to define a scope for further analysis. Based on the selected scope and research question, a portion of the data was analyzed with Celonis. The previously defined scope was applied, and the resulting process-categories were documented modeled to establish a baseline. In the further analysis, a specification was performed, as the student category inherits the possibility to focus on particular aspects. To ensure a structured and goal-oriented approach we formulated a research question to orientate and navigate through the upcoming challenges: "Are there any optimization possibilities in the area of compliance, throughput time regarding the top 10 vendors? ".

The investigation of the throughput times of the invoicing process showed, that the activity flow from RGR to CI is with great distance the most often occurring bottleneck. This opens up the possibility for further internal investigation, whether the payment process is supposed to take such a long time.

Throughout the analysis process several challenges had to be overcome. One of these is the uniqueness of most processes and as such the difficulty of abstracting and summarizing without losing sight of important details. The analysis section is based on the assumption, that seemingly inexplicable deviations in the net value can be used to identify compliance violations. In total three analysis methods were used to identify potentially deviating cases and subsequently grouping the results by the corresponding vendors. Especially infrequent or unique cases can be covered by the sheer number of cases. Finally drawing on the established models, the scope and with regard to the research question, the purchase documents of particular vendors were selected in a deep analysis. Here several irregularities were observed. For example, do Vendor_136, Vendor_104 and Vendor_106 follow the happy path to a comparatively low degree. Additionally, a root cause analysis showed that the process will also be extended if RGR occurs after CPO instead of after VCI.

Again, the process after the invoice is received until the invoice is paid takes a long time in almost all cases. Finally, we found for the top ten vendors in the packaging spend area, only minor issues with regard to invoice conformance occur.

Reviewing the entire analysis process leads to the conclusion, that the vast amount of cases from many different spend areas is challenging to oversee and understand. For

some cases it could be helpful to have background knowledge making connections between activities easier to understand. Generally, it makes sense to differentiate between the four types of data flows. However, it could be useful to further differentiate and structure according to the most important vendors so that rework especially in the payment process is limited as much as possible.

References

1. Aalst, Wil M. P. & Weijters, A. (2004). Process mining: A research agenda. Computers in Industry. 53. 231-244.
2. Aalst, van der, W. M. P. (2011). Process mining: discovery, conformance and enhancement of business processes. Berlin: Springer.
3. Tiwari, A., Turner, C. J, & Majeed, B. (2008). A review of business process mining: State-of-the-art and future trends. Business Process Management Journal, 14(1), 5-22.
4. Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L., & Xie, X. (2018). Optimal Process Mining for Large and Complex Event Logs. IEEE Transactions on Automation Science and Engineering, 15(3), 1309-1325.