

# Performance, Variant, and Conformance Analysis of an Academic Travel Reimbursement Process

Luise Pufahl<sup>1</sup>, Richard Hobeck<sup>1</sup>, Paul Binetruy<sup>2</sup>, Wepan Chada<sup>2</sup>, Mykola Dytar<sup>2</sup>, Kim Julian Gille<sup>2</sup>, Marta Slarzynska<sup>2</sup>, Fabian Stiehle<sup>2</sup>, Ingo Weber<sup>1</sup>

<sup>1</sup> Software & Business Engineering, Technische Universität Berlin, Berlin, Germany  
`{firstname.lastname}@tu-berlin.de`

<sup>2</sup> Technische Universität, Berlin, Germany  
`{firstnameInitial.lastname}@campus.tu-berlin.de`

**Abstract.** Process mining techniques enable organizations to exploit business process data logged in their IT systems to analyze process performance, compare process variations, and detect compliance issues. The travel reimbursement process, a traditional process in almost every organization, also plays a key role in the academic administration. It needs to be fast and effective, allowing smooth business trip handling while preventing misuse of academic funds. In this report, we analyze an academic travel reimbursement process of a Dutch university for internal (i.e., for employees) as well as external (i.e., for external visitors) reimbursements. The most common process flow is discovered, and its flow time, as well as its bottlenecks, are analyzed. Furthermore, different process variants are presented and compared by departments and projects. Finally, the results of the conformance analysis are presented where the to-be process is compared to the real-world data and common deviations are discussed. The report results offer the domain experts insights for further analysis and process improvements.

**Key words:** BPI Challenge, Travel reimbursement, Process mining, Variant analysis

## 1 Introduction

Annually, the process mining community is presented with the Business Process Intelligence (BPI) Challenge: a case-study type initiative providing publicly available real-life event logs to test process mining techniques on. It challenges students and professionals from industry and academia to investigate process data and apply their process mining skills to try and answer a set of questions about the data.

Combining methods from the fields process analysis, data science and business process management, process mining offers a set of techniques to unveil information about processes' actual behavior in practice based on real-life event logs (*discovery*). Once the actual behavior of the process is known, we can measure the process performance, find bottlenecks and investigate variations of the process for their causes. The discovered process model can be checked for deviations

from the the to-be models (*conformance checking*) which helps to take measures process improvements (cf. [1]).

The 2020 data set<sup>1</sup> covers event logs of the reimbursement of travel expenses process as well as travel permit-issuing at Eindhoven University of Technology (TU/e). Reimbursement processes are central for organizations’ endeavor to handle and typically compensate travel costs incurred by their staff and externals, e.g., visitors who were invited by the organization. If not set up properly, however, they can also pose the risk of double-billing expenditures or delaying travel plans, thereby causing additional costs or loss of business and research opportunities – prospects that justify examining their performance with scrutiny. The logged data comprises five event logs, as shown in Table 1. It covers the years 2017 and 2018, with 2017 being a pilot year in which the process was still undergoing changes. Due to the latter reason, in this report the 2017 data is generally not considered in the analysis, and cases that started in 2017 are filtered out.

Table 1: BPI Challenge data sets

| Data set                         | Total  |        | 2017  |        | 2018+ |        |
|----------------------------------|--------|--------|-------|--------|-------|--------|
|                                  | Cases  | Events | Cases | Events | Cases | Events |
| <i>DomesticDeclarations</i>      | 10,000 | 56,437 | 1,740 | 10,062 | 8,260 | 46,375 |
| <i>PrepaidTravelCost</i>         | 2,099  | 18,246 | 323   | 2,442  | 1,776 | 15,804 |
| <i>InternationalDeclarations</i> | 6,449  | 72,151 | 1,552 | 14,886 | 4,897 | 57,265 |
| <i>TravelPermits</i>             | 7,065  | 86,581 | 1,481 | 15,923 | 5,584 | 70,658 |
| <i>RequestsforPayments</i>       | 6,886  | 36,796 | 1,108 | 4,976  | 5,778 | 31,820 |

For each business trip, employees of the TU/e have to make at least one declaration, distinguishing between international and domestic declarations.

The data model for the relevant business objects is shown in Fig. 1. As per the documentation provided with the data sets, international trips should only be started once a travel permit has been obtained. While the *InternationalDeclarations* log focuses on the international reimbursement process from the viewpoint of the declaration (i.e., the instance notion is the declaration), the *PermitLog* focuses on the travel permits. In the latter, permits exist with one, several, and even no related declaration(s). In the *InternationalDeclarations* log also cases without any permit can be observed. Further, for one permit multiple requests for payment of prepaid expenses can exist (present in 1,314 of the 7,065 permit cases, with up to 15 times for a single permit case). The latter are contained in the *PrepaidTravelCost* log. Additionally the provided *RequestsForPayments* log includes cases of requesting reimbursements of external visitors or students.

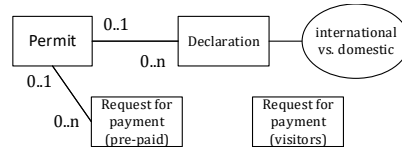


Fig. 1: Data model of the process documents and their relations.

<sup>1</sup> <https://icpmconference.org/2020/bpi-challenge/> (last accessed: 31-8-2020)

In the process of composing this report, we followed the Process Mining Project Methodology ( $PM^2$ ) [2] where applicable, guiding process mining projects through six stages: *Planning*, *Extraction*, *Data Processing*, *Mining and Analysis*, *Evaluation* and *Process Improvement and Support* (out of scope). To obtain our results, we largely used the ProM framework<sup>2</sup> as well as the tools Celonis Intelligent Business Cloud (Academic Version)<sup>3</sup> and Disco<sup>4</sup>, both with academic licenses.

This report comprises our submission to the BPI Challenge 2020. Section 2 revolves around discovering process models from the data sets, revealing the real-life process flow, and analyzing the process performance. Section 3 details process variants and compares them across projects and departments, while Section 4 provides insights into the process conformance with respect to the to-be process. Section 5 discusses the findings of the previous sections and draws a conclusion.

## 2 Process Discovery and Performance Analysis

In this section, we will discover the most common process flow for travel reimbursements at TU/e from the real-world event data. We split our discovery and analysis results into the internal process for employees of TU/e (Section 2.1), and reimbursements for students or external visitors of the TU/e (Section 2.2). Unless stated otherwise, in the following we refer only to data from 2018 on and insights based on that data. Additionally to the discovered control flow, we also present performance data of the process and its different stages (permit, declaration etc.). Rejections of permits, declaration, etc. have a negative impact on the overall process performance because it requires rework for several workers involved in the process (e.g., employee, administrators, managers). Thus, we give insights on how often rejections occur and discuss improvement possibilities.

### 2.1 Internal Travel Reimbursement Process

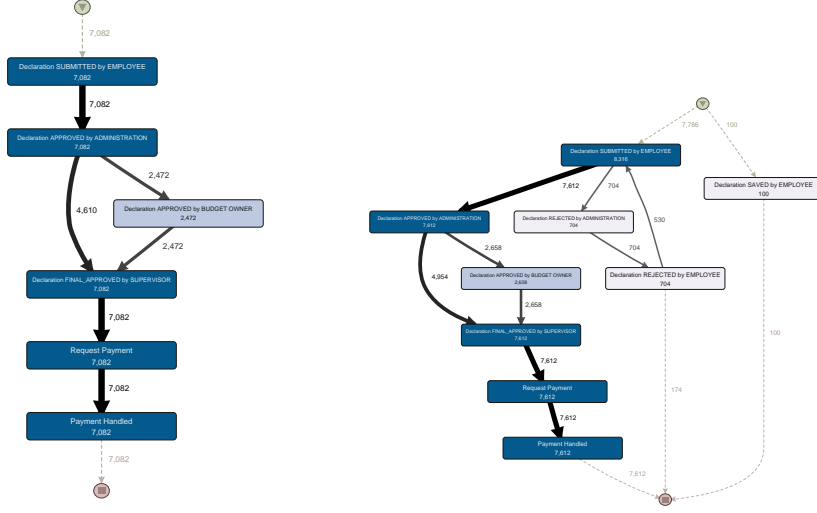
The internal travel reimbursement process for employees of TU/e has two different process variants, the domestic and the international reimbursement process. In comparison to the domestic travel reimbursements, the international process is more complex because a permit is necessary and payments can be requested in advance, also multiple times. For a domestic trip within the Netherlands, only a declaration has to be made after the trip.

We discovered the models shown in this section with Disco, and used the variation filter to explore the most common behaviors. A variation is a certain type of a trace that can be observed in a log. As such, we discovered the happy path (cf. [3]) showing the two most common variations, and a more realistic one, also showing typical deviations from the happy path.

<sup>2</sup> <http://www.promtools.org/doku.php>

<sup>3</sup> <https://www.celonis.com/academic-alliance/>

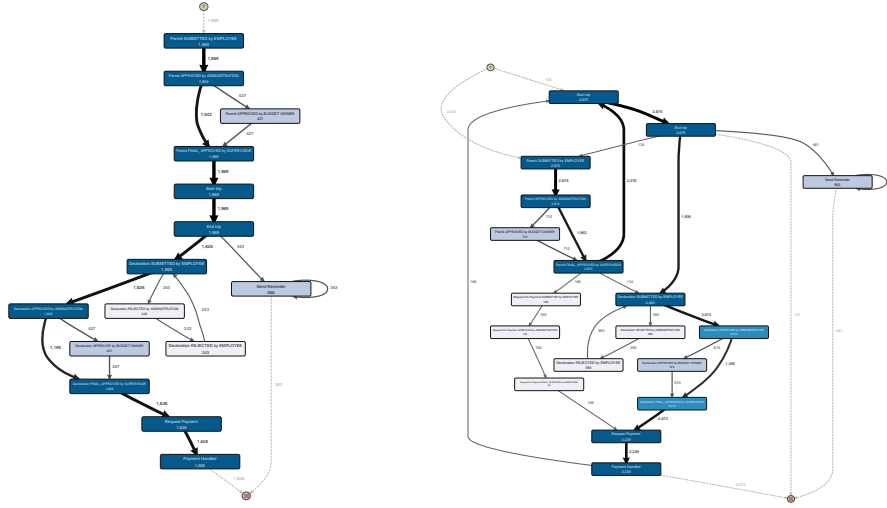
<sup>4</sup> <https://www.fluxicon.com/disco/>



(a) Happy path model (85% of cases, two variations) (b) More realistic model (95% of cases, six variations)

Fig. 2: Process Model mined from the *DomesticDeclarations* log using Disco.

**Discovered process models.** In the happy path model for the domestic reimbursement process in Fig. 2a, we can see that most cases (85%) follow a successful flow where a submitted declaration is approved by all relevant parties. About a third of the cases are additionally approved by the budget owner, i.e., the budget owner was not the supervisor. The more realistic model in Fig. 2b covering 95% of the cases shows that 704 cases of the visualized 7,886 cases are rejected by the administration. Afterwards they are also rejected by the employees. Mostly they are resubmitted under the same declaration number, and are finally successfully accepted and paid. Some cases end with the rejection of the employee. Additionally, some declarations are simply saved and never processed any further. For discovering the international reimbursement process, we selected the *Travel Permits* log as this is the initial document needed to perform such a trip. The 448 cases without a permit will be discussed later in the conformance analysis. In contrast to domestic travel, the happy path model for the international reimbursement process shown in Fig. 3a covers only 35% of the cases. As explained above, this process includes more phases and an employee can submit a request for prepaid reimbursement at any point after having a fully approved travel permit, resulting in more variations. Furthermore, the expected workflow is not always followed for international travel, which we also cover later in detail in the conformance analysis (Section 4). In the happy path model, the permit gets approved, followed by the trip. After the trip, the employee hands in a declaration, which is approved. Approximately 1 out of 9 declarations is rejected



(a) Happy path model (35% of cases, (b) More realistic model (47% of cases, eight variations)

Fig. 3: Process Model mined from the *TravelPermits* log using Disco.

and resubmitted. Less than a fifth of the observed cases has a reminder send to the employee twice, to remind them to hand in their declaration. All of these reminders seem to be ignored, as *Send Reminder* is the last activity in all of the cases. Looking at concrete traces in the log, we observed that a reminder is sent out by the system usually two months after the end of the trip, exactly on the first of the month.

The more realistic model shown in Fig. 3b already shows the common behavior of 47% of the cases in the log. Additionally to handing in a permit first, we can observe in the more realistic view that there is a minority of cases where the trip is started before a permit was submitted, as well as before the permit was approved. Furthermore, in this model few employees request payments in advance (166 cases out of the visualized 2,675 cases). If a payment is requested, this happens directly after the final approval of the permit and before the start of the trip. Still, it is noticeable that 121 simply end after the trip and 481 cases after two reminders without any declaration. This is not only due to the cut-off of the event data because also permits includes trips, which have already started in January.

**Throughput time analysis.** In order to analyze the performance<sup>5</sup> and identify potential bottlenecks of the internal travel reimbursement processes, we used enhanced process models with the median and mean duration between the events. For better visibility and analysis, four models were created, i.e., for domestic

<sup>5</sup> We interpreted all questions about throughput to mean throughput time, or duration.

travel declarations (cf. Fig. 4a), for international travel declarations (cf. Fig. 4b), for permits (cf. Fig. 5a) and finally for the pre-paid travel cost payments (cf. Fig. 5b). By representing the median (shown in larger font size in the models) and the mean, we can analyze which durations are influenced by outliers.

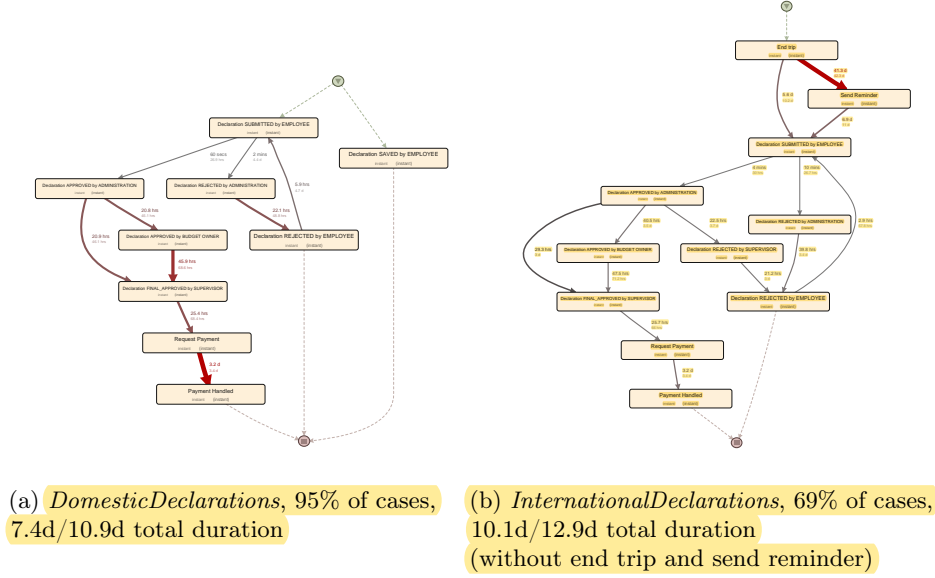


Fig. 4: Declaration processes with median and mean durations based on the *DomesticDeclarations* and *InternationalDeclarations* log.

For the throughput time analysis of the declaration process we consider the more “realistic” variations where a declaration is mostly approved and then sent for payment, but might be rejected by the administration. In the next paragraph, we focus also on the rejection by other roles. The overall process takes usually (median) 7.4 days for Domestic declarations and 10.1 days for international declarations by considering the *Declaration Submitted by Employee* activity as start point and *Payment Handled* as end point. In both processes, it mostly takes only a few seconds to minutes to get the declaration accepted or rejected by the administration. The average value for this duration is much higher. By checking the data in more detail, we observed that almost half (2,435 out of the 4,897 cases) of the international declarations are processed in less than 2 minutes by the administration. This might be an indication that the administration uses a service for automatically checking the documents, and only critical cases need to be checked additionally by an employee. Another clear finding is: if the budget owner is involved in checking a case, more time is needed and, even more time is spent until the final check of the supervisor. For example, in the

domestic declaration process the median duration from administration approval to supervisor approval increases from 20.9 hours without a separate budget owner to 20.8 hours plus 45.9 hours with a separate budget owner. A related question in this regard by this year’s challenge is: *How many travel declarations are not approved by budget holders in time (7 days) and are then automatically rerouted to supervisors?* An exhaustive answer would require deep domain knowledge, however, we can check in which cases the supervisor was directly involved after the administration and where the duration between those two roles is longer than 7 days. Thus, we filtered all cases where final approved/approved/rejected by a supervisor is directly following the approval by the administration with a duration of more than 7 days. It resulted in 211 domestic declarations (2,55%) and 363 international declarations (7.41%). Whereas in the domestic process this is a relatively small minority of cases, in the international process it might be relevant for further analysis. Still, this analysis is limited as we cannot determine whether the budget owner would have been involved at all. Comparing Figs. 4a and 4b, the approval process for the international declarations takes longer than for the domestic ones. The reason might be that more and different types of documents need to be checked in *InternationalDeclarations*, whereby the process of obtaining relevant approvals becomes more time-consuming. Moreover, also the end of the trip and the reminders can be observed in the shown model of the international declaration process. The median time for employees declaring their travel cost is 5.6 days after the trip. In 136 cases (not shown in the given performance model), employees need a reminder, and then hand it in 6.9 days later (median). In both declaration process variants, we observed that as soon as a declaration was finally approved by a supervisor, the time needed for requesting payment, i.e., initiating reimbursement process, is in most cases more than one day and on average almost three days. As this relatively long duration is not fully explainable, it might be considered a bottleneck, which should be further analysed.

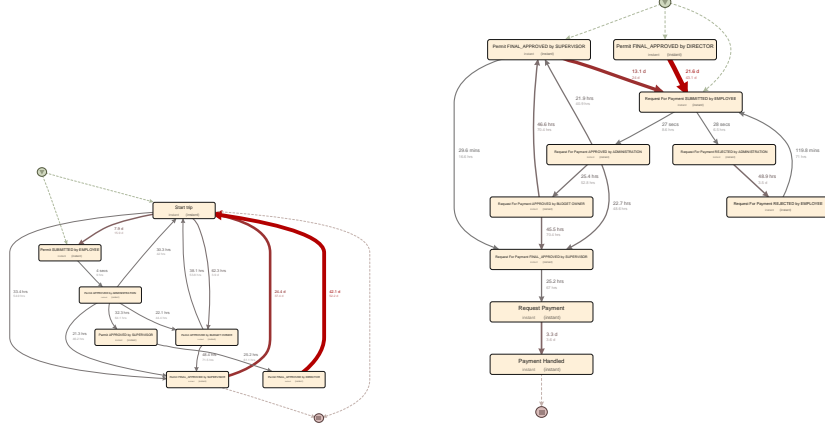
As described in the previous sections, a permit is required for international travel. It needs to be obtained before the actual trip happens, but it also has occurred after the start of the trip. This could mean that either the start date changed, or the permit was obtained during the actual trip. We can observe that employees often hand in the permit 7.9 days after the trip started, as shown in Fig. 5a. We next focus on the cases in which a permit was submitted before the actual trip. Employees usually submit their travel permit around a month before their actual trip (29.5 days). The overall permit process for 64% of cases reached a final approval after 44.7 hours (or 1.9 days, median). Similar to the travel declarations, the approval of the administration is received very quickly, here mostly in a few seconds. Moreover, the process also takes more time if more roles are involved (e.g., with budget owner vs. final approval by supervisor, the process needs an additional 49 hours) as shown in Fig. 5a. For cases where a supervisor is responsible for the final approval, employees usually submit their permit requests closer to their departure date (mostly 25 days in advance) comparing to cases

for which a director’s approval is required (mostly 45 days in advance). This numbers also include the 1.9 days to reach the approval.

As part of the travel reimbursement process, travelers can also claim for a reimbursement of prepaid travel expenditures, provided their permit is already approved. Employees usually requested payment in advance 13.1 days after their permit was finally approved by a supervisor, or after 21.6 days in case of a director’s final approval (see Fig. 5b). The total duration of pre-paid travel cost process activities, starting from requesting a pre-payment until such payment was handled, had a median of 8.3 days. Again, we can observe that more involved roles lead to longer processing times. When a pre-paid travel cost request needs to be additionally approved by a budget owner, it takes more time (median of 48.2 hours) to reach supervisor approval. Similarly to the declarations, for the majority of pre-paid travel cost requests it takes 25 hours (median) to request payment when the final approval is already obtained; payment is then handled in 3.3 days (median).

In summary, we learnt:

- Most cases are efficiently handled and documents are processed in a few days.
- The administration often needs only some seconds/minutes to process a document and takes hours in some cases. The former might be due to an involved automatic service.
- Supervisors need more time to finally approve the request if a budget owner is involved (applies to both the prepaid and declarations logs).



(a) Permits, 64% of cases,  
44.7hrs/3.4d total duration  
(without start trip)

(b) Prepaid travel cost, 80% of cases,  
8.3d/10.3d total duration  
(without permit approval)

Fig. 5: Permit and prepaid travel cost processes with median and mean durations based on the *PermitLog* and *PrepaidTravelCost* log.



- Finally approved declarations and prepaid cost requests need more than a workday to be transitioned to Request Payment. It might be possible to reduce this time by automation.
- In the international declaration process, it occurs in 7.4% of the cases vs. 2.6% in the domestic process that the time between the final approval of administration and the processing by the supervisor is more than 7 days.

Table 2: *DomesticDeclarations*: Number of rejections per role and median duration of the rejected cases by the respective role from submitted until payment or process end.

| Declaration SUBMITTED (total = 9,139, cases 8,160) |         |                  |                        |
|--|---------|------------------|------------------------|
| Rejected by...                                     | # total | # case           | median duration (days) |
| ADMINISTRATION                                     | 948     | 842              | 11                     |
| BUDGET OWNER                                       | 59      | 58               | 11.3                   |
| SUPERVISOR   | 222     | 217              | 15.0                   |
| Total corrections                                  | 1,119   | 979 (+323 never) | -                      |

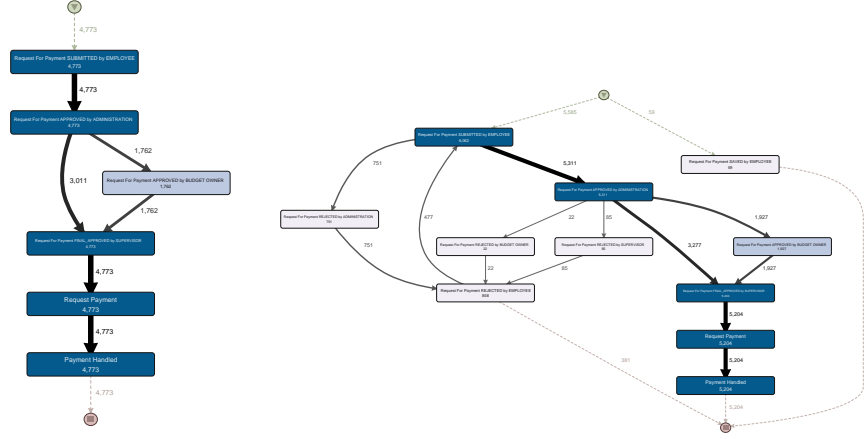
**Occurrences of rejections.** Each rejected document in the travel reimbursement process causes additional work for both the employee, as well as for the different roles responsible for approval. In this part, we analyze the number of rejections per role as well as the median processing duration incurred by a rejection. To this end, we used a filter in Disco and selected only the cases that included the respective reject activity. Starting with domestic reimbursements, Table 2 shows that 8,160 unique declarations were submitted (not including the 100 saved declarations), and in total 9,139 were processed, which includes the resubmissions after rejection. Further, Table 2 shows, for each role, the total number of rejections as well as the total number of rejected cases. Additionally, we show the median duration of the cases being rejected by the respective role from the submission until the end of the process. Note that we do not list the rejections by the employees, because they usually reject their submission as a consequence of it being previously rejected by another role. We can observe that most of the submitted declarations are rejected by the administration. Still, only a little more than 10% of the 8,160 submitted declarations are rejected. The budget owners as well as the supervisors only reject a few cases, here the larger share of the cases is rejected by the latter. By comparing the number of rejected cases (# cases) with the total number of rejections (# total) for each role, we can see that only a fraction of the cases is rejected multiple times. A rejection always leads to longer processing times. In the previous section, we have learnt that the processing of a domestic declaration mostly takes 7.4 days as shown in Fig. 5a. With a rejection by the administration the median duration between submission until process end takes 11 days. The longest time can be observed for the rejections by the supervisor, 15 days. 979 cases were rejected, reworked,

and finally accepted, some with repetitions. A minority of submitted declarations (323 cases) were never approved.

Table 3: *InternationalDeclarations*: Number of rejections in the permit log and median duration of the rejected cases by the respective role from submitted until final approval or process end.

| Rejected by...  | # total | # case             | median duration (days) |
|---|---------|--------------------|------------------------|
| Permit SUBMITTED (total = 5,768, cases = 5,573)               |         |                    |                        |
| ADMINISTRATION  | 126     | 124                | 7                      |
| BUDGET OWNER  | 50      | 49                 | 10.6                   |
| SUPERVISOR  | 91      | 91                 | 10.3                   |
| DIRECTOR  | 2       | 2                  | 31.5                   |
| Total corrections   | 193     | 183 (+75 never)    | -                      |
| Request for prepaids SUBMITTED (total = 1,713, cases = 1,092) |         |                    |                        |
| ADMINISTRATION  | 228     | 209                | 11.3                   |
| BUDGET OWNER  | 7       | 7                  | 13.2                   |
| SUPERVISOR  | 20      | 19                 | 15.8                   |
| Total corrections   | 184     | 170 (+71 never)    | -                      |
| Declaration SUBMITTED (total = 6,006, cases = 4,329)          |         |                    |                        |
| ADMINISTRATION  | 1,512   | 1,258              | 14.3                   |
| BUDGET OWNER  | 36      | 36                 | 20.2                   |
| SUPERVISOR  | 93      | 90                 | 20.2                   |
| DIRECTOR  | 1       | 1                  | 15                     |
| Total corrections   | 1,494   | 1,199 (+152 never) | -                      |

In the international process, it is noticeable that only very few permits get rejected by the different roles. Still a rejection leads to longer processing times: filtering the rejected cases which got finally accepted, a reject by the budget owner leads to 10.6 days in contrast to 44.7 hours median duration (cf. Fig. 5a). A higher portion of prepaid requests were rejected in contrast to permits, most often by the administration. The median duration of such requests for payments is increased in case of a rejection (8.3 days median duration including also the payment, cf. Fig. 5b). 157 requests for payments are resubmitted by the employee and 33 are never finally accepted. In comparison to the domestic declarations, many more international declarations (around 25%) are rejected, also most often by the administration. In the BPI challenge, the question was raised: *How many travel declarations are first rejected because they are submitted more than 2 months after the end of a trip?* By filtering cases where the declaration was submitted 60 days of more after trip end, we found 101 cases getting directly rejected by the administration after submission but also several ones which were accepted. Thus, the duration might not be a relevant reason for rejecting a declaration. International declarations, which are usually processed in 10.1 days (median), cf. Fig. 4b need 4 days more if they were rejected by the administration and around two times as long for the other rejects. 1,200 cases are reworked by the employees and finally successfully approved; only a few are never approved.



(a) Happy path, 82% of cases, two variants (b) More realistic model, 97% of cases, 12 variants

Fig. 6: Process models mined from the *RequestsForPayment* log using Disco.

## 2.2 External Travel Reimbursement Process

**Discovered process models.** To discover the *external travel reimbursement* process, we used the *RequestsForPayment* log. The BPI website mentions that request for payments are also related to hardware reimbursements. We assume the process is also used to reimburse travel expenses by external guests and students of the TU/e, and assume that the log was prepared, such that it only includes traces related to travel reimbursements.

As with the previous process models, we filtered the log to only include cases after 2017, which resulted in 5,778 cases and 31,820 events remaining in the log. The happy path model in Fig. 6a shows that the majority of the requests (82% of the cases) are handled successfully, too. We found that little more than half the cases (about 58.52%) also needed approval by the budget owner, whereas the remainder of cases was approved by the administration and supervisor only. Fig. 6b shows the more realistic process model, including 97% of the cases. Here, the corresponding rejection activities for each of the approval steps are visible (rejected by administration, budget owner, supervisor, and employee). Also, the *SAVED by EMPLOYEE* activity is displayed here. As shown in our more realistic model, 477 resubmissions occur, after the request was rejected by the employee. Out of these, 431 requests finally were approved after one or at most two resubmissions.

**Throughput time analysis.** Fig. 7a shows our realistic process model of the *RequestForPayment* log annotated with the median and mean durations. The median case duration is 8.4 days and the average is 11.9 days, which is similar to

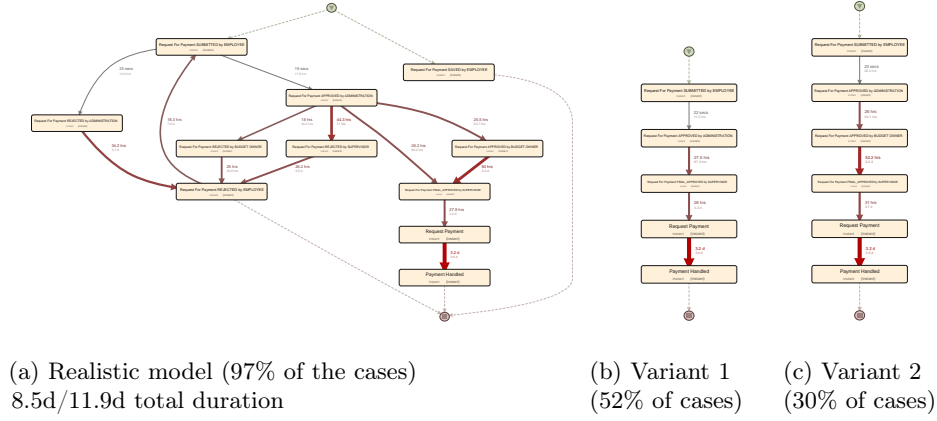


Fig. 7: *RequestForPayment* process with median (and mean) throughput times

the processing of the prepaid costs. For the positive approval flow, this duration varies between 7.3 days (Fig. 7b) and 11.3 days (Fig. 7c), depending on whether or not the request has to be approved by the budget owner in addition to the supervisor. As for the internal process, we can see that the final approval by the supervisors takes almost twice as long (median of 50 hours compared to 26.2 hours) if the request was approved by the budget owner first. This is similar to our previous findings and might be worth to analyze the reasons behind this additional longer processing times in case of an handover from budget owner to supervisor. We also find that a rejection leading to a (successful) resubmission more than doubles the median case duration to 16.2 days. This is also due to a longer time-span between rejection and the resubmission. For some outliers, we found durations of 27 and 193 days until the employees resubmitted their requests. One could analyze, if a time limit for the re-submission is appropriate here. The median processing time for cases with real rejections is 40.5 hours. Similar to the internal processes, it takes more than a day between the final approval and initiating the payment as well as the handling of payments takes 3.2 days. It seems that the same payment system is used for all processes.

**Occurrences of rejections.** As with the previous processes, we summarized the rejection occurrences for the *external travel reimbursement* process in Table 4. Again, we can see that the administration rejects the vast majority of cases. A much smaller amount is rejected by the supervisors, who still reject about three times as many cases as the budget owner. Looking at the median processing times of the request for payments, we find that the administration's reject is not leading to an increase of the throughput time (cf. Fig. 7a). The rejection by the budget owner and the supervisor leads to longer processing times. Finally, 501 cases needed rework at least once, and 414 cases were never approved. These amount to a 15.84% rework rate when looking at all the submitted cases.

Table 4: *RequestsForPayment*: Number of rejections per role and median duration of the rejected cases by the respective role from submitted until payment or process end.

| <i>RequestsForPayment</i> (total = 6,324, cases = 5,778) |         |                  |                        |
|--|---------|------------------|------------------------|
| Rejected by...   | # total | # case           | median duration (days) |
| ADMINISTRATION   | 834     | 753              | 7.3                    |
| BUDGET OWNER   | 47      | 47               | 13.1                   |
| SUPERVISOR   | 133     | 128              | 18.1                   |
| Total corrections  | 1,069   | 501 (+414 never) | -                      |

### 3 Variation Analysis

To break down the complexity of the travel reimbursement processes at TU/e, we conduct a variation analysis and group the process flow into predefined categories. This will help to get more detailed insights on specific process streams, and to analyze causes for deviations. We compare the departments and projects with a high volume of declarations and aim at finding out the differences between clusters of declarations, for example between departments/projects. We also examine how many travel declarations are booked on projects by looking into the influence of unknown project numbers on the process. Thus, only logs holding information about project numbers (*Case-Permit ProjectNumber* and/or *Case-Project*) or departments (*Case-Permit OrganizationalEntity* and *Case-OrganizationalEntity*) were suitable for this analysis. Those included the *InternationalDeclarations* (discussed in Section 3.1), the *PrepaidTravelCost* (in Section 3.2) and the *RequestForPayment* log (in Section 3.3).

#### 3.1 International declarations

The *InternationalDeclarations* log has the attributes *Case-Permit OrganizationalEntity* and the *Case-Permit ProjectNumber*. For each attribute, we analysed the process flows of the most frequent data instances, described in the following.

*Case-Permit OrganizationalEntity*. We extracted the six most frequently occurring organizational units. Their process maps show that all six units have similarly complex processes (given the number of activities and paths per process model) and follow various process variants (see Fig. 8). The organizational unit was always provided in the log; the unit was never unknown. The ratio of resubmissions after an initial rejection is similar throughout the selected departments (between 30%-40%), only *Department 65455* has a lower resubmission rate (8,6%). In the units 65455 and 65460, the budget owner is less involved, and the director almost not; in contrast, the budget owner is very present in the units 65456 and 65454, and the director in 65458.

*Case-Permit ProjectNumber*. Of the 4,897 traces in the log, for 1,858 (37.9%) the project numbers were *UNKNOWN*, and the remaining 3,039 cases (62.1%) had a

project number assigned. The generated process models follow different flows in the six most frequent projects when ordered by *Case-Permit ProjectNumber* as shown in Fig. 9. The process model of the submissions with unknown project numbers has the most complicated process map with the highest amount of events and paths. The ratio of resubmitted declarations is at 32.5%, while 3.2% of the submissions remain rejected. Unknown project numbers probably occur because the travel was paid for from an internal budget, or due to a data quality issue. The main contributing project (Project 426, 282 traces) has a diverse process map with 3,9% of the cases being rejected. In projects 426 and 3442, the budget owner is involved in approximately 50% of the cases. The process models of projects 1185, 1623, and 3442 are almost or completely similar to the happy path model as presented in the previous section only with some rejections by the administration.

### 3.2 Prepaid travel cost

The *PrepaidTravelCost* log has one attribute dedicated to the organizational unit (*Case-OrganizationalEntity*) and two identifiers for projects (*Case-Permit ProjectNumber* and *Case-Project*).

*Case-OrganizationalEntity*. When looking at the *PrepaidTravelCost* logs, the same six organizational units as in the *InternationalDeclarations* log appear most frequently, which is unsurprising given the link between the two processes. However, in the *PrepaidTravelCost* log the number of submissions per organizational unit is around 70% lower and the process flows are less diversified than their *InternationalDeclaration* counterparts. We also observed a lower level of resubmissions (around 10%) and a lower number of events per process map. (Due to those similarities within the set of models, we did not include a set of process maps for these variants.)

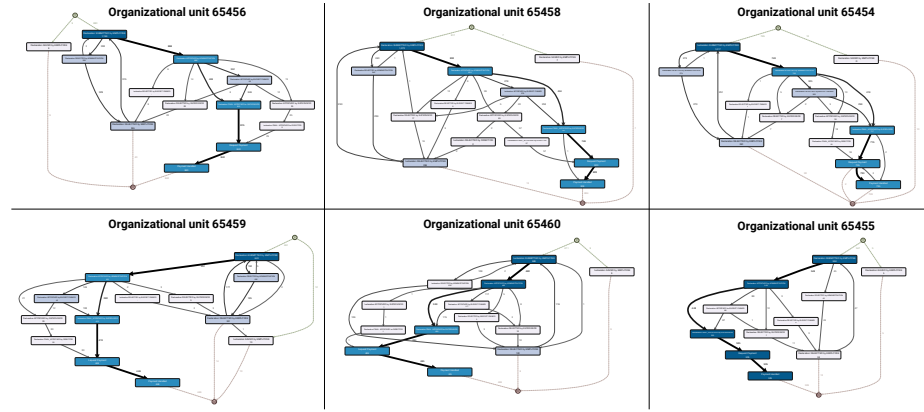


Fig. 8: Process maps of the six most frequently submitting organizational units (*Case-Permit OrganizationalEntity*) in *InternationalDeclarations*.

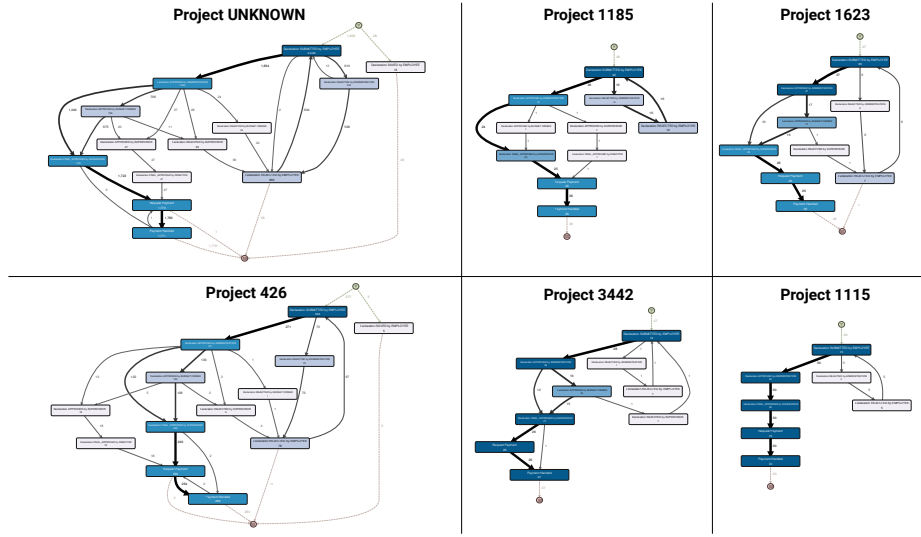


Fig. 9: Process maps of the six most frequently submitting projects (*Case-Permit ProjectNumber*) in *InternationalDeclarations* (including UNKNOWN).

*Case-Permit ProjectNumber*. The *PrepaidTravelCost* log documented 1,776 cases. For the *Case-Permit ProjectNumber*, 1,049 cases had a project number, while 727 cases had an unknown project number. Dividing the *PrepaidTravelCost* log by *Case-Permit ProjectNumber* reveals that, like in the *InternationalDeclaration* log, traces with unknown project numbers occur most often, followed by traces on *Project 426* (72 traces) (see Fig. 10). These two variations' project maps also incorporate the highest number of events and paths. The remaining less frequently occurring projects (18-22 traces) have much more streamlined processes with a significantly smaller number of traces and resubmissions (0% - 11%). Even *Project 1115* that managed to complete its declarations most efficiently, indicating that they have an expert handling the declarations, still has a rejection and resubmission rate of 16.7% with all resubmissions finally being handled. That hints towards the process being too complicated in the submission phase.

*Case-Project*. Of the 1,776 cases of the *PrepaidTravelCost* log, 22 cases (1.2%) had an unknown project number for the *Case-Project* attribute, 1,754 cases (98.8%) had a project number assigned. Dividing the log by the attribute *Case-Project* shows a different picture compared to the division by *Case-Permit ProjectNumber*. An unknown *Case-Project* attribute led to certain rejection of a declaration. The most frequent project is *project 503* (92,3% of traces) which has a chaotic process map. *Project 503* seems to be an accumulative project number concentrating the vast majority of traces, possibly used as a last resort if no project number was available for handing in a request, which is reflected in the diverse process flow. The other projects occur about two orders of magnitude less often and accordingly

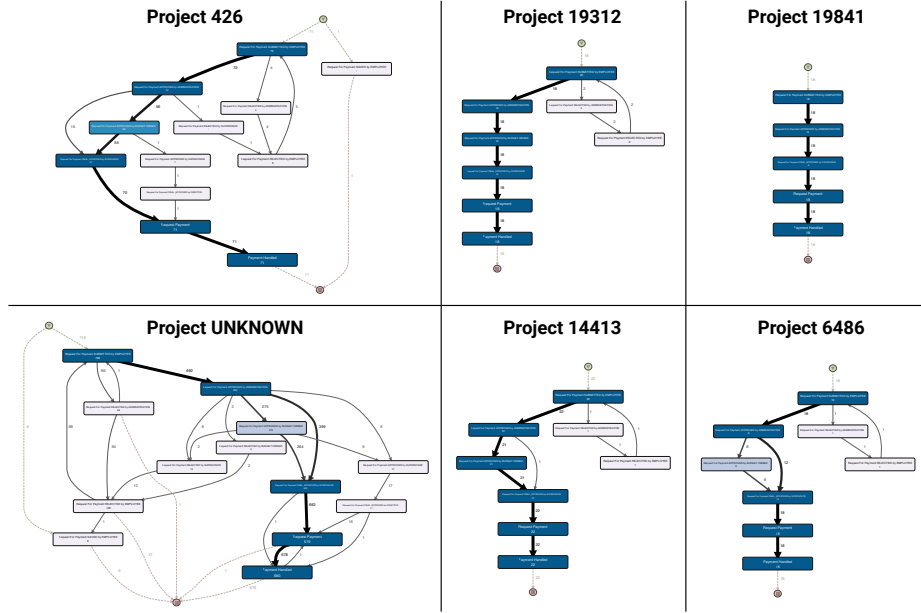


Fig. 10: Process maps of the six most frequently submitting projects (*Case-Permit ProjectNumber*) in *PrepaidTravelCost* (including UNKNOWN).

have more straight-lined flows resembling the happy path. (The process maps are not included due to the complexity of the process map of *Project 503*)

### 3.3 Request for payment

The *RequestForPayment* log has the case attribute *Case-Project*, like the *PrepaidTravelCost* log, but no attributes indicating the related organizational unit. It holds 5,778 traces of which 5,345 traces have a project number. The project number for the remaining 433 traces is unknown.

*Case-Project*. The traces are distributed more evenly over the projects than the *PrepaidTravelCost* log, and have no extreme outliers (249 - 857 traces among the most frequent projects). The process maps of the main contributing projects (depicted in Fig. 11) all have similar numbers of events and paths with similar rates of rejection (0.4% - 1.6%) and resubmission (5.7% - 10.8%). The exception is Project 503 with a resubmission rate of 21.0% due to a high ratio of *Request For Payment REJECTED by ADMINISTRATION*. Similar to the *PrepaidTravelCost* log, an unknown project number led to certain rejection. Unlike the *PrepaidTravelCost* log, the *RequestForPayment* log does not appear to have an accumulative project. Across all projects, the budget owner is involved in less than 50% of the cases. The director is involved 5 times in project 148052, in the others only once or not at all. This could indicate that large projects, with frequent travel, have



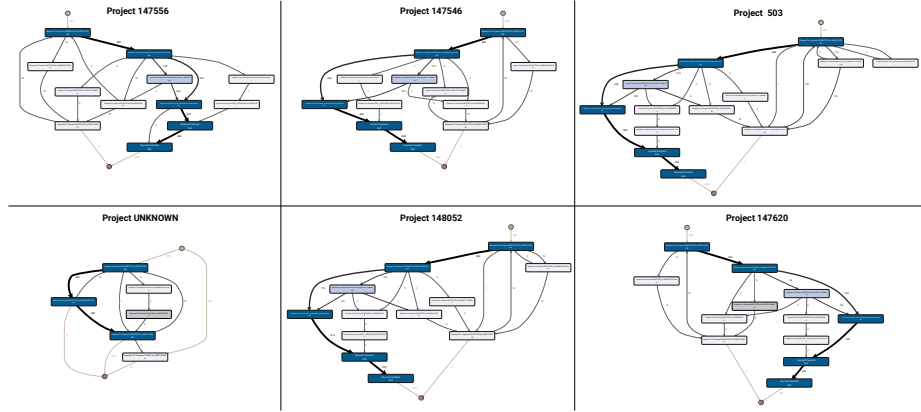


Fig. 11: Process maps of the six most frequently submitting projects (*Case-Project*) in *RequestForPayment* (including UNKNOWN).

Table 5: Number of cases with known / unknown project numbers per event log.

| Logs                      | Case-Permit ProjectNumber |         | Case-Project |         |
|---------------------------|---------------------------|---------|--------------|---------|
|                           | Known                     | Unknown | Known        | Unknown |
| InternationalDeclarations | 3,039                     | 1,858   | -            | -       |
| PrepaidTravelCost         | 1,049                     | 727     | 1,754        | 22      |
| RequestForPayment         | -                         | -       | 5345         | 433     |
| Total                     | 4,089                     | 2,584   | 7,099        | 455     |

dedicated budget owners (e.g., project leaders) who are relatively often distinct from the supervisors.

Additionally, the most frequent projects in the *RequestForPayment* log have a high number of cases, which could indicate that for these projects costs other than travel costs of external visitors' expenses are requested (like the mentioned hardware costs). We did not see a way to make a clear distinction on basis of the available information.

*How many travel declarations were booked on projects* has a conditional answer. Information about the project was only available in three logs and divided into two attributes (see Table 5). Considering the *Case-Permit ProjectNumber* attribute 4,095 bookings had a project assigned, 2,578 did not. For the *Case-Project*, 7,095 projects were booked with a project number, 459 were not, but all of the latter were rejected.

## 4 Conformance Analysis

In this section, we present the most interesting deviations discovered through our conformance analysis. First, based on the process flow description given, we introduce our reference models for the internal and external reimbursement

process at the TU/e in Section 4.1 and the resulting fitness scores. Then, we discuss major deviations that can be observed across all logs in Section 4.2. Finally, we present our most prominent log-specific findings in Section 4.3. For our conformance analysis we utilized ProM for fitness calculations, and Celonis and Disco to investigate deviations.

#### 4.1 Reference models and Fitness results

One major question of this year’s BPI challenge is whether the real executions given in the event logs follow the intended reimbursement flow, which was given on the BPI website. The provided description still leaves room for interpretation. For instance, it is not clarified when a director needs to give approval, or what role is logged in the system if the *Budget Owner* and *Supervisor* are the same person. We filled these blanks by studying our discovery results (Section 2). Still, we made sure to capture the explicit requirements stated on the website.

For the conformance checking, we analysed the individual processes of the different documents instead of using the overall travel permit process to allow for a more fine-grained analysis. Additionally, we provide some analysis results for the Travel Permits log at the end of this section as the different case notion chosen there allows for some interesting additional insights.

The approval process flow for submitting a Permit, Declaration, or Request for Payment can all be captured by a subprocess as shown in Fig. 12 and replacing *DOCUMENT* with the corresponding document type. For the *DomesticDeclarations* (Fig. 13 a) and *RequestForPayment* log (b), no permit is necessary, and after the declaration or request for payment is made, the payment is handled. For the *InternationalDeclarations* log (c) and *PrepaidTravelCost* log (d) a permit is mandatory. For *PrepaidTravelCost*, a Permit has to be obtained before a Payment can be requested. For *InternationalDeclarations*, a permit has to be obtained

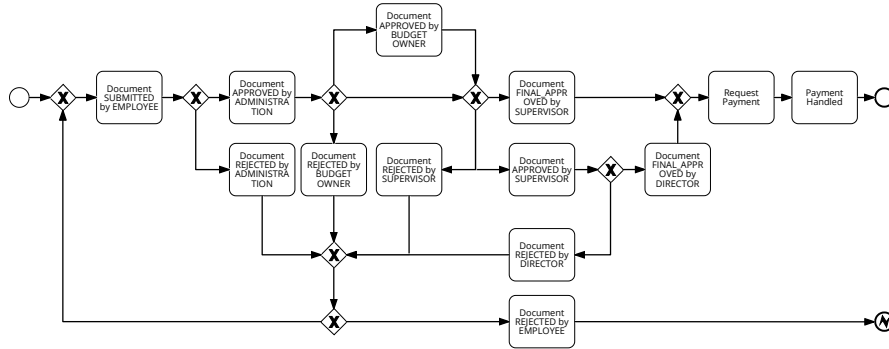
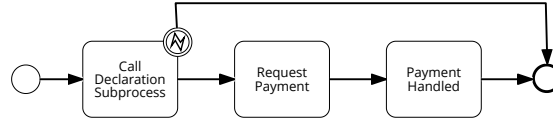


Fig. 12: The approval process flow for submitting a Permit, Declaration or Request for Payment can all be captured by replacing *DOCUMENT* with the corresponding document type.

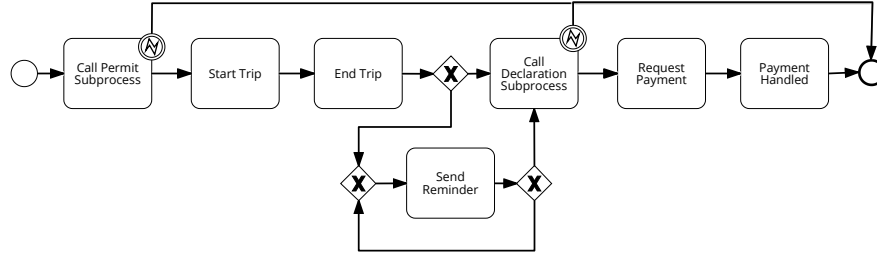
before a trip can be started. The process models for the different travels are shown in Fig. 13.

To calculate the fitness, we used the ProM plugin “Replay a log on a Petri Net for Conformance Analysis”. To enable this, we transformed the shown BPMN diagrams into Petri nets, and we filtered cases started in 2018 with Disco before processing them in ProM. Using its default settings in ProM, we yielded the fitness values shown in Table 6.

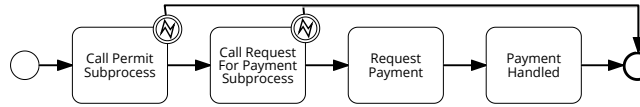
Logs without the permit process (*DomesticDeclarations* and *RequestForPayment*) have higher fitness than the ones with the permit process. To gain more insights, we investigated all main deviations. We will later discuss permit related deviations that lead to this lower fitness score, and now proceed to highlight our most prominent conformance findings, summarized in Table 7.



(a) Model for *DomesticDeclarations*, and *RequestForPayments* (When replacing Declaration with Permit)



(b) Model for *InternationalDeclarations*



(c) Model for *PrepaidTravelCost*

Fig. 13: Models using the subprocess defined in Fig. 12. The subprocess can be acquired by replacing *DOCUMENT* in Fig. 12 by the subprocess name.

Table 6: Fitness of the logs for the given process models calculated in ProM.

| Data set                  | % Fitness |
|---------------------------|-----------|
| DomesticDeclarations      | 97.8%     |
| InternationalDeclarations | 89.8%     |
| RequestForPayment         | 98.1%     |
| PrepaidTravelCost         | 91.9%     |

Table 7: Summary of key findings with the most prominent deviations, the number of cases affected and possible reasons.

| Deviation   | #Cases total (#2018) | Correlates with...  |
|---|----------------------|---|
| REJECTED by EMPLOYEE before resubmission.                       | 3,184 (2,920)        | All resubmissions across all logs.  |
| No Declaration Submitted, No Payment Submitted                  | 308 (238)            | Declaration SAVED by Employee activity, Request for Payment SAVED by Employee activity. |
| Permit acquired during travel (Between Start Trip and End Trip) | 89 (19)              | Trip length (time between Start Trip and End Trip) of 200+ days.                        |
| Permit acquired after travel (After End Trip)                   | 620 (471)            | -   |
| Travel without a Permit   | 448 (266)            | Project 426 and Permit number 424.  |
| Prepaid payment without a Permit                                | 243 (218)            | Project: 426, UNKNOWN, 18960 and Permit number 424.                                     |
| Trips without a submitted Declaration (Travel Permits Log)      | 1,356 (1,255)        | Reminder ignored or not sent.   |

## 4.2 Deviations observable in all processes

It became obvious that there are some deviations happening across all processes, which we present in this subsection.

**Conformance of resubmission flow.** All logs contain cases of requests that were resubmitted after being rejected by the employee. The process description on the website, however, states: “**Either** the employee resubmits the request, **or** the employee also rejects the request.” This deviation violates the either-or relationship described. It essentially introduces the REJECTED by EMPLOYEE event as another prerequisite to resubmitting the declaration.

By filtering for cases in which SUBMITTED by EMPLOYEE is followed anytime by SUBMITTED by EMPLOYEE and in which REJECTED by EMPLOYEE is directly followed by SUBMITTED by EMPLOYEE, we found 2,920 cases in all the logs (cf. Table 7). The detailed results per log and the influence on the average duration are given in Table 8. While this type of resubmission violates the process description given, it is clear that this is the actual behaviour for nearly all cases. We observe that this non-conforming flow adds an additional

Table 8: Non-conforming cases of the resubmission flow (2018 onwards) and the additional delay caused.

| Data Set  | % resubmissions with non-conforming flow | AVG + delay (days) |
|---|--|--------------------|
| Prepaid Travel Cost                             | 93% (158 of 169 cases)                   | 2.3 days           |
| Requests for Payments                           | 97% (515 of 530 cases)                   | 8.0 days           |
| Domestic Declarations                           | 99% (842 of 848 cases)                   | 4.0 days           |
| International Declarations                      | 99% (1192 of 1209 cases)                 | 0.0 days           |
| International Declarations: Request for Permits | 97% (156 of 161 cases)                   | 6.0 days           |

throughput delay in most cases. However, our sample size for the conforming resubmission flow is limited.

**SAVED by Employee activity.** Across all logs we observed that actors can also save their document, such as the declaration, or the request for payment without submitting them. We can observe that these cases correlate with all cases that never submit a Declaration or a Request For Payment respectively.

The *DOCUMENT SAVED* by Employee activity is not described in the process description and was accordingly not captured in our BPMN process diagrams, such that all these cases are viewed as a deviation. By filtering cases including the activity *DOCUMENT SAVED* by Employee, we found a few cases for each log as shown in Table 9 where the total number of cases, and cases only for 2018 are shown.

Table 9: Occurrence of specific SAVED by Employee activity

| Data Set                                   | # Cases (# 2018) |
|--|------------------|
| <b>Declaration SAVED by Employee</b>       |                  |
| DomesticDeclarations                       | 135 (100) cases  |
| InternationalDeclarations                  | 75 (58) cases    |
| <b>RequestForPayment SAVED by Employee</b> |                  |
| PrepaidTravelCost                          | 24 (19) cases    |
| RequestsforPayments                        | 74 (61) cases    |

We assume that this activity captures a save draft function in the system that leads to a request being saved for later editing. However, it is unclear why these drafts are not picked up again. Possible causes for this observation might be that the system is not able to relate drafts to final declarations, or that save draft events are not included in the log if the document is eventually submitted.

### 4.3 Specific deviations

After having discussed general conformance findings for all logs, we focus here on deviations relevant for specific logs and their root causes. *Note that, for this*

analysis, we also used cases that were started before 2018, and as such the findings might not apply to the final system.

**Permit acquired after Start Trip.** For international travel, we observe a considerable number of cases that start their trip without a valid permit. Most of them acquire a valid permit afterwards. They end their trip, acquire a permit, and then submit a declaration. In most cases, this process ends by either a successful payment or the End trip event. We decide to split our analysis into investigating both variants. Generally, the violations can be summarized in the sequences:

1. Start Trip → Permit Handling → Declaration Handling → End Trip
2. Start Trip → End Trip → Permit Handling → Declaration Handling

where *Permit Handling* and *Declaration Handling* denote the steps taken to acquire a permit or declaration and payment respectively (cf. Model Section).

*Method.* For this analysis, we consider all cases in the International Declarations log. For (1) we select all cases that start the process with Start Trip and end the process with End Trip. For (2) we select all cases that start the process with Start Trip and end with Payment Handled.

*Analysis.* For (1) we found 89 cases (started after 2018: 19 cases). All cases eventually acquire a permit and submit a successful declaration which is then paid out. We observe an interesting pattern: of the 89 cases, 53 alone start their trip at 31-12-2017 23:00h. We see that the average throughput time is 350 days – almost a year. We observe the same behavior for the timestamp 31-12-2016 23:00h. We see the corresponding end trip events on 30-12-2017 23:00 and 30-12-2018 23:00 respectively. We can generally see a correlation between the length of the trip, that is the throughput delay between Start Trip and End Trip and the occurrence of this pattern. If we select cases with a trip length of at least 200 days from all the cases in the log, we get 91 cases, of which 87 cases follow the flow as described in (1).

For (2) we obtained 620 cases (started after 2018: 471 cases). Contrary to (1), we identified 29 cases that never applied for a permit. These cases are a subset of cases that never apply for a permit but still travel. These cases are discussed next.

**Travel without a Permit.** *Method.* For this analysis we consider all cases in the International Declarations log. We select cases that do not pass any permit-related FINAL APPROVED activity.

*Analysis.* We obtain 448 cases from which 266 happened in 2018. Most of these cases occur around April 2018 (19 cases). None of these cases applied for a permit in the first place. Of the 448 cases, 426 go through a successful declaration process and are eventually paid out.

All 448 cases are linked to project number 426 through the field 'CASE-PERMIT PROJECTNUMBER'. Furthermore, we observe, that while these cases skip the permit process they are linked to Permit 424 through the field 'CASE-PERMIT

TRAVEL PERMIT'. Thus, all cases which travel but never acquire a permit, and also do not apply for one in the first place, belong to project 426. If we filter for all cases contained within project 426 we can obtain 450 of 464 non-conforming cases. We will later link Permit 424 and project 426 to other violations in the Prepaid Travel Cost log.

**Prepaid payment without a Permit.** *Method.* For this analysis we consider all cases of the *PrepaidTravelCost* log. We select only cases that not pass any permit-related FINAL APPROVED activity.

*Analysis.* We obtain 243 cases. These cases occur mostly throughout 2018 (218 of 243 cases). We can relate 85 cases to project number 426 through the field 'CASE-PERMIT PROJECTNUMBER'. Furthermore, we observe, that while these cases skip the permit process, they are linked to Permit 424 through the field 'CASE-PERMIT TRAVEL PERMIT'. Thus, these 85 cases can be related to the 448 cases that travel without permit that occur in the *InternationalDeclarations* log. The remaining cases are either linked to project UNKNOWN (155 cases) or 18960 (3 cases). Of the total 243 cases, 207 still get paid out.

**Trips without a submitted Declaration.** As we chose to model the process of International Declarations and Prepaid Travel Cost separately, we omitted the Travel Permits log (*PermitLog*) from our fitness calculation. However, we still conducted rule checking on the log. The most interesting deviation is the occurrence of travels without a follow-up declaration. This is a deviation that cannot be seen in the *InternationalDeclarations* log, as it only contains cases for which a declaration has been submitted.

*Method.* For this analysis we consider all cases of the Travel Permits log *PermitLog*. We select cases that do acquire a valid permit start and end their trip but do not submit a declaration.

*Analysis.* We acquire 1,356 cases from which 1,255 are related to trips in 2018. For 977 cases a reminder was sent but never acted upon from which also a smaller part (107) have also submitted a request for payment. These cases should be further analyzed.

For 324 cases, the process ends with End Trip, of which most ended their trip after 30-09-2018. If our assumptions from Section 2.1 (in which we observe that a reminder is sent out by the system usually two months after the end of the trip, exactly on the first of the month) are correct, the reminder is probably yet to be sent out and is not captured in our log. Still, 146 cases ended their trip throughout 2017 and 2018, and no reminder was sent at all.

## 5 Discussion and Conclusion

After having analysed the reimbursement process at the TU/e in detail in this report, we here list our main results and discuss limitations as well as future work. In summary, our observations are the following:

- The process discovery in Section 2 showed that the reimbursement process at the TU/e seems to be an efficiently handled process where documents are processed within a few days, and are handed over quickly between the different involved roles. In most of the cases, employees receive payments for their request within 11 to 14 days. Especially the administration is able to check requests in very short timeframes, which might be due to an automatic system. In the future, it might be an opportunity to check its applicability also for the other roles if legally and organizationally permissible.
- For the declaration as well as the request for payments, it was surprising that the time between the final approval and starting the payment procedure seems to take at least one day, which might be due to system interfaces. From a business perspective, this delay seems unjustified. This might be a point for potential improvements.
- An interesting observation is that the time to reach supervisor approval increases once a separate budget owner is involved. This can mean that supervisors who have separate budget owners are more busy and therefore usually need more time, or the handover from a budget owner to a supervisor is associated with more effort. This might be an interesting area for further analysis.
- Most rejections can be observed for declarations of international trips by the administration. It might be the case that employees need more support (additional training material, etc.) about required information and documents for international trips. We also found that a rejection in general leads to longer processing times. This is often due to a time span between the rejection and the resubmission. One could analyze whether employees would benefit from support in the resubmission.
- Even experienced submitters (by organizational unit and project) keep a high resubmission rate as shown in Section 3 (e.g., *Project 1115*, 16.7%), which supports the hypothesis that the submission might be rather complicated.
- Also, the two different project IDs were frequently unknown with a great imbalance between attributes (38.7% of *Case-Permit ProjectNumber*, 6.0% of *Case-Project*). Additionally, if declarations require a project number (i.e., for the *Case-Project* attribute), using accumulative default projects if the process number is unavailable should be prevented. Alternatively, the submission could be simplified by making the project number optional.
- Although some flexibility is allowed within the process, most of the recorded traces followed the prescribed behavior. When performing the conformance analysis, comparing the to-be process model with the logs (Section 4), we observed fitness values between 89.8% and 98.1%, with a tendency of processes involving a request for a permit scoring lower.
- Our conformance analysis via an alignment of logs and Petri nets revealed as the most common deviation a resubmission after the employee has already rejected a case. It should be checked whether this deviation is indented behavior.
- More serious deviations are travel without a permit (448 cases, which could be mainly observed for project 426), which were still approved later. Note, however, that this observation is partly based on data from 2017.



- Alarming might be the 1,356 international trips without a declaration. Unless this can be tied to a data quality issue, action might be warranted.

*Limitations and future analysis.* While working to solve the Challenge, we encountered considerable limitations. First of all, the *RequestForPayment* log’s description left room for interpretation on whether it only contains external travel reimbursements, or it includes other requests for payments, such as hardware/software costs. The log attributes gave no further hints for interpretation. In addition to that, we worked under the assumption that all cases were completed, due to the long tail of events past 2018, but no new instances starting.

In the discovery part in Section 2, we identified the number of rejections and their negative influence on the overall throughput time. However, we could not identify reasons for rejections due to the limited availability of additional data regarding the cases; only the unknown project attribute *Case-Project* led to definite rejections without resubmission. In the future, the TU/e could analyze reasons for rejections internally based on their complete data sets. The variation comparison in Section 3 relies to some extent on comparing different process maps visually, for which we used *Disco* – a tool that was not designed to make visually comparable process maps. In a future analysis, it might be interesting to generate clusters automatically based on similar process flows. The conformance analysis of Section 4 covered all sub-processes captured in the logs for the international travels. Still, the conformance analysis should in a future analysis be extended to the *PermitLog*, which is a perspective that could have offered additional insights. However, this would be more challenging as the log includes several 1-to-n relations, which needs then to be considered.

As stated in the introduction, we went through the first five phases of the ( $PM^2$ ) methodology [2] and explored the logs answering the Challenge questions. The last stage *Process Improvement and Support* is left to the process owners and domain experts at TU/e to bring our results to action.

**Acknowledgements.** We would like to thank the lecture participants of the Process Mining course 2020 for inspirations in their final project reports.

## References

1. van der Aalst, W., et al.: Process Mining Manifesto. In: Business Process Management Workshops. Volume 99, LNBIP. Springer (2012) 169–194
2. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P.:  $PM^2$ : A process mining project methodology. In: Advanced Information Systems Engineering. Volume 9097, LNCS. Springer (2015) 297–313
3. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Exploring Processes and Deviations. In: Business Process Management Workshops, Springer (2015) 304–316