

Business Process Intelligence Challenge 2020: Analysis and evaluation of a travel process

Sabine Klein¹, Johannes Lahann¹, Lea Mayer¹, Dominic Neu¹, Peter Pfeiffer^{*,1},
Adrian Rebmann², Martin Scheid^{*,1}, Brian Willems¹, and Peter Fettke¹

¹ German Research Center for Artificial Intelligence and Saarland University,
Campus D3 2, 66123 Saarbruecken, Germany `firstname.lastname@dfki.de`
<https://www.dfki.de/>

² Data and Web Science Group, University of Mannheim, Mannheim, Germany
`rebmann@informatik.uni-mannheim.de`

Abstract. The 2020 Business Process Intelligence Challenge provides data from the travel management system of the Eindhoven University of Technology and encourages the participants to tackle several questions regarding qualitative and quantitative aspects of the process flow. In this report, we show our analysis of the log files beginning with a conceptual approach, which includes the manual creation of a baseline process using the given textual description and some basic data descriptions. Based on this, we define our understanding of the log files, and the process followed by some basic descriptive analysis, which allows to get a deeper understanding of the process based on the data. This allows us to answer some of the first questions, which mostly relate to the quantitative aspects of the travel process. Furthermore, we apply some more advanced process mining and data analytics techniques, such as process conformance checking, machine learning and techniques, which are provided by the newly implemented RMM4Py framework to deal with the other questions of the challenge. Using those methods, we also try to find further noticeable process behaviour, which is not covered by the provided questions. This is followed by an overview of our results and the discussion of limitations and further recommendations, wherever applicable.

Keywords: Process Mining · Conformance · Travel Process.

1 Introduction

The digitization of business processes is one of the most important topics of the last decades. More and more processes are digitized and through this, they became analysable by algorithms and descriptive technologies. The resulting data is often stored in, so called, event logs. These consist of a sequence of events. Each of these events can also contain various additional attributes, such as the users who executed an activity or other resources, which can be analysed. Data Mining algorithms take the data which came from the information systems, e.g. workflow engines, and analyse them according to predefined or conspicuous

behaviour. Using the current state-of-the-art algorithms makes it possible to find problems in the process flow which otherwise, does not become visible due to the complexity of the process or the amount of data which needs to be processed.

In the 2020 BPI Challenge [4], we have a deep look into one of those processes: the travel process of the Eindhoven University of Technology (TU/e), a University with a research focus on engineering science and technology, which uses its own system for processing, approval and accounting of business trips. It is interesting to note that this system is available for university employees, as well as external staff, but these processes differ in the way they are carried out. The task of this year's challenge is to find misuse of the system and conspicuous process behaviour such as long sequences, bottlenecks or conspicuous recurring drags as well as suspicious employee combinations. In addition, we have looked at the attachments between the various partial logs to be able to track possible overlaps and relevant infringements. Figure 1 gives an overview of our work within the BPI Challenge and the structure of the report.

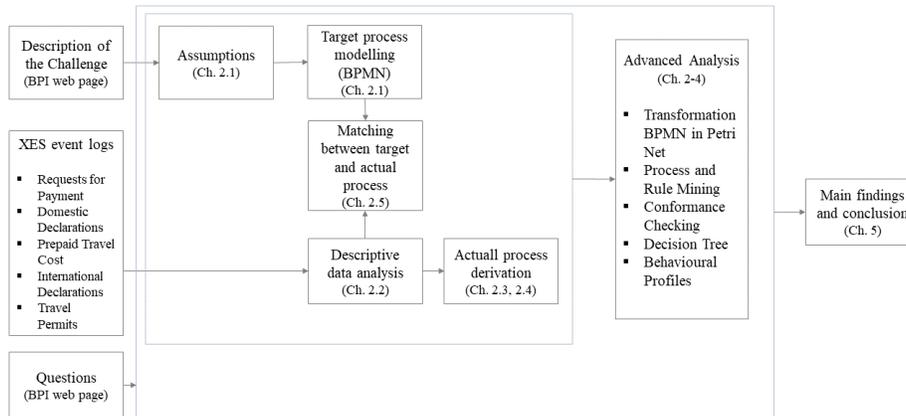


Fig. 1: BPI-Challenge - Overview

We start our report with a basic analysis of the given textual description of the process. Using this, we issue some requirements and assumptions to the process description without consideration of the data. This makes it possible to model the process using the BPMN 2.0 standard[5]. On the other hand, we create a descriptive data analysis for each of the three event logs with a focus on a first description of the provided data[4]. Subsequently, analyses were carried out regarding possible dependencies between the logs and special features of its were addressed. Furthermore, for the following analysis we mapped the activities from the log to the modelled functions in the BPMN models which were created

from the given process description to get a single point of truth from a conceptual and data mining point of view. Section three answers the questions which came with the challenge in detail using the prepared data and process models. Section four includes some analysis which goes beyond the questions and tries to identify further relevant contents or abnormalities in the log with advanced methods. The last section concludes the report and gives some suggestions on how to address the findings of the analyses.

2 Analysis

Before answering the questions, we analysed the information, available from the challenge description and the data, which came as five event log files. For analysing the data, we mainly used RMM4Py, a python framework for reference modelling and process mining as well as functionalities from PM4Py 1.3.0 [2]. Whenever required functionalities were not available from RMM4Py 1.0.0 or PM4Py, we implemented them ourselves using python and several libraries like pandas. To model the corresponding processes using BPMN, Camunda 3.7.2³ was used. All analyses, answers to the questions and a more detailed description of the implementation will be made available via Jupyter notebooks⁴.

2.1 Process Understanding

The subject of the investigation are the travel arrangements and the reimbursement of expenses of TU/e employees and non-TU/e employees. In the following, the target process will be described, which was determined solely on the basis of the textual description and represented in BPMN models. Ambiguities or lack of information should be taken up accordingly in the following assumptions. These assumptions are introduced subsequently and will be maintained throughout the work.

Assumptions

A1 In the context of travel accounting for TU/e employees, a distinction must be made between the process request for payment (see figure 4), the reimbursement of costs without travel expenses, and the activity "request payment" within the context of the reimbursement of costs with travel expenses, namely domestic declaration, international declaration and pre-paid travel costs. Both reimbursements are represented in the same way in the BPMN, consisting of "Request Payment" and "Make Payment".

³ <https://camunda.com/de/>

⁴ <https://bpm.dfki.de/bpi-challenge-2020-notebooks/>

A2 Various declaration documents - all declarations (domestic and international declarations, pre-paid travel costs and requests for payment) follow a similar process flow - in the BPMN identified as sub-process "handle request" (see figure 3). In all cases, the travel administration, budget owner, supervisor and director are involved.

A3 Travel permits: The travel authorization follows a slightly different process compared to the declaration documents. No payment is made; instead, after approval, a trip with planned trip dates (expected start and end date) can take place. However, all previous steps up to the approval are the same, refer to BPMN sub-process "handle request".

A4 "After a travel permit is approved, but before the trip starts, employees can ask for a reimbursement of pre-paid travel costs" - the option of declaring pre-paid travel costs can be used for both domestic and international travel.

A5 "Several requests can be submitted independently of each other. After the trip ends, an international declaration can be submitted, although sometimes multiple declarations are seen for specific cases" - the option to submit several declarations at once is shown in the BPMN with the loop symbol. The task will be repeated until the condition for the termination of the process handle request is met or until a predefined number of repetitions has been reached. As there is no information on the maximum number of requests, the termination criteria is the processing of all pending requests.

A6 "Next to travel declarations, there are also requests for payments. These are specific for non-TU/e employees. Are there any TU/e employees that submitted a request for payment instead of a travel declaration?" - This information from the listed questions of the BPI-Challenge is included to describe the process understanding. According to our understanding, TU/e employees have the possibility to get reimbursed for costs that are related or unrelated to a trip via a request for payment. However, this form of reimbursement is only intended for costs that have no reference to a trip. Non-TU/e employees, on the other hand, should be reimbursed for their trip via a request for payment.

In the following, the two types of trips, domestic and international trips, as well as the associated processes will be described based on the assumptions made above-noted. Figure 2 shows the corresponding BPMN.

Domestic travel For planning and executing a domestic travel, no additional approval (travel permit) is necessary before booking and starting the trip. The trip can be booked and commenced, and reimbursement can be requested at two points in time, depending on the type of costs. This will be explained in more detail below.

Several request types (pre-paid travel costs, domestic declaration):
First of all, a distinction can be made between two types of costs: pre-paid travel

costs and all costs incurred during the trip (domestic declaration). While the reimbursement of pre-paid travel costs can only be requested prior to the trip, the domestic declaration can only be made after. After the trip, the employee also receives a reminder generated by the system. The domestic declaration is the final travel expense report, meaning the request for reimbursement of all costs incurred. As the two applications, domestic declaration and pre-paid travel costs, follow a similar process flow, they will be summarized below under the sub-process "handle request" (A2, A4). The procedure for these reimbursement requests is explained in more detail below:

International travel Compared to domestic travel, international travel requires an additional approval, the so called travel permit, before booking and starting the trip. This will be further explained in the following.

Travel permit: After the employee has submitted the request, it is sent to the travel administration for approval. Once approved, the request is forwarded to the budget owner and lastly to the supervisor. If the budget owner and supervisor are the same person, then only one of these steps is required. In some cases, the director must also approve the request afterwards. If the request is approved, the trip can be booked and started, otherwise the employee has the option of dropping the request or resubmitting it after revision (A3).

Several request types (pre-paid travel costs, international declaration): While the reimbursement of pre-paid travel costs can only be requested before beginning the trip (A4), international declarations can only be made after the trip is completed. In special cases, several international declarations can be submitted for a particular travel permit (A5). After the end of a trip, the employee also receives a reminder generated by the system. The international declaration is the final travel expense report, meaning the request for reimbursement of all costs incurred. The procedure for these reimbursement requests is the same as for domestic travel. Also, if the request is rejected, the employee has two options, as with all requests: The employee can resubmit the request after revision or dismiss the request. As these processes are very similar to the previous application procedures (pre-paid travel costs, domestic declaration), this procedure (international declaration) will also be described using the sub-process "handle request" (A2).

Request for payment This type of reimbursement covers costs that are not related to a travel case of a TU/e employee. This includes, for example, expenditure on project-specific hardware or representation costs. It also covers the reimbursement of costs related to travel, in case of a non-TU/e employee (A6). The description of the travel planning, processing and accounting of a non-TU/e employee is not included in the description of the challenge. In the following, the

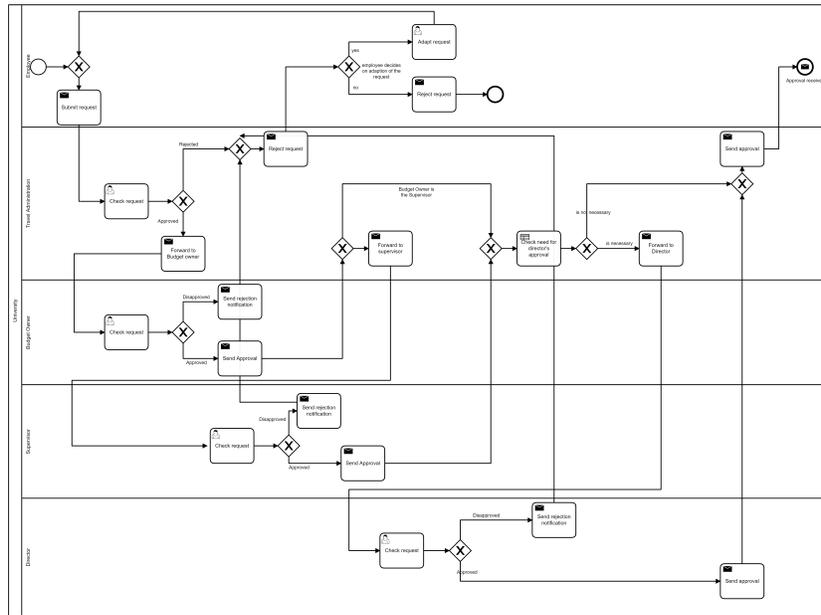


Fig. 2: BPMN - Handle request

processes of the request for payment are explained in more detail. Figure 2⁵ shows the corresponding BPMN.

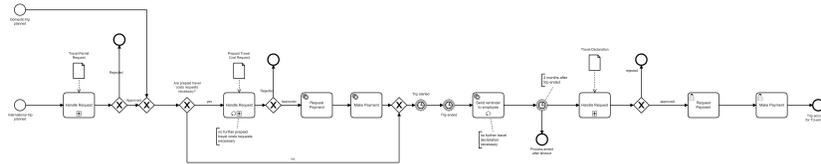


Fig. 3: BPMN - Domestic and international declaration

After submission, the request for payment is sent to the administration. If approved, the request is then forwarded to the budget owner, followed by the supervisor. If the budget owner and supervisor are the same person, only one of these steps is executed. In some cases, the director must also approve the request afterwards. If the request is approved, a payment request is made (Request Payment) and a payment is issued to the applicant (Make Payment). However, if the request is rejected, the applicant can resubmit the revised request or dismiss

⁵ The full scale images are available for download at <https://bpm.dfki.de/bpi-challenge-2020-notebooks/>.

it.

As the first part of this process is very similar to the previous application procedures (pre-paid travel costs, domestic declaration, international declaration), this part of the process (first part of request for payment) will also be described using the sub-process "handle request" and the handling of payment (A1, A2). The Request for payment BPMN is visualized in Appendix A⁶.

2.2 Descriptive data analysis

Table 1: Meta-data of the different logs

Log	#cases	#events	#activities	#case attributes
Permit	7,065	86,581	51	168
InternationalDeclarations	6,449	72,151	34	18
DomesticDeclarations	10,500	56,437	17	5
PrepaidTravelCost	2,099	18,246	29	17
RequestForPayment	6,886	36,796	19	9

For an initial overview, table 1 shows some general statistics of the different log files. It can be seen that the permit log is composed of more activities and several more case attributes, than any other log. The event level data, on the other hand, is identical in all five logs. It comprises an ID-field, an activity name, a timestamp, a resource and an organizational role.

The resource is always one of two possible values *SYSTEM* or *STAFF MEMBER*. Likewise, the organizational role has eight possible values (*EMPLOYEE*, *SUPERVISOR*, *MISSING*, *UNDEFINED*, *DIRECTOR*, *PRE_APPROVER*, *ADMINISTRATION*, *BUDGET OWNER*) and is the same across all logs, except in domestic declarations, where the director is never involved.

The timestamps of events in the logs start at 09.01.2017 for the domestic declarations, the requests for payment and the prepaid travel cost. Events in the permit log, as well as the international declarations log start at 04.10.2016. The latest recorded date is different in each log: 21.02.2019 for prepaid travel cost, 17.06.2019 for domestic declarations, 08.08.2019 for request for payment, 09.05.2020 for international declarations and 31.08.2021 for the permit log. The case attributes are explained in more detail for each log individually hereafter.

Domestic Declarations This log has 5 case attributes. From these, the *id* and *concept:name* are identical. The attribute *DeclarationNumber* is the *id* incremented by one, except in 452 cases, where the field value is *UNKNOWN*. The case attribute *BudgetNumber* is always *budget 86566* and the last attribute is the

⁶ The full scale images are available for download at <https://bpm.dfki.de/bpi-challenge-2020-notebooks/>.

Amount, a float value between 0 and 3,292.54 with a median of 43.42 (Further details in appendix B).

International Declarations This log has more case attributes than its domestic counterpart but similar to the domestic declarations log, the *id* and the *concept:name* are identical. Also the *DeclarationNumber* is the *id* incremented by one in 6,189 cases whereas in 260 cases, it is *UNKNOWN*. Next to these, there exist four columns with IDs referring to the related permit, namely: *Permit travel permit number*, *travel permit number*, *Permit ID* & *Permit id*. The *travel permit number* and *Permit travel permit number* are identical in 5,970 cases and different in 479. In the differing cases, *Permit travel permit number* is 23 times *UNKNOWN*. *Permit ID* and *permit id* are identical in 6,001 cases and never *UNKNOWN*. When the two IDs are different, *Permit id* is always *travel permit 423*. Furthermore, in 5,970 cases the *travel permit number* is the *Permit ID* incremented by one while in 448 cases, it is decreased by one, leaving 31 cases, where both numbers are completely different.

In addition, there are five numerical case attributes *Amount*, *RequestedAmount*, *OriginalAmount*, *AdjustedAmount* & *Permit RequestedBudget*. The first three are always identical, whereas the *AdjustedAmount* is different, only in one single case, where the *Amount* of 0 is adjusted to 100.49. The requested budget is usually higher than the amount.

Next to these, the process has 6 categorical attributes attached to it, which are encoded as IDs. The attributes include 6 different tasks, 207 budgets, 719 permit budgets, 825 projects, 27 organizational entities and 34 activities (Further details in appendix C).

Prepaid Travel Costs The case attributes of this log resemble the other two logs. The unique case identifier *Rfp_id* is identical with the *concept:name*. Again, the *RfpNumber* is the *Rfp_id* incremented by one in 1,999 cases but *UNKNOWN* in 100 cases. The same holds for the *Permit id* and *Permit travel permit number*. The number is the ID in 1,931 cases and is different in 168 cases, where the number is *UNKNOWN*. The *RequestedAmount* has a median of 515.04. While 99 % of the cases remain below 3,500, 7 cases exceed 10,000 and 3 cases even surpass 400,000. The *RequestedBudget* on the other hand, has a higher median of 1,477.44 but is at max 13,451.80. The *Cost Type* contains no information gain, as it is always 0. Next to these, there are nine categorical attributes with anonymized IDs, similar to those in the international declarations (further details in appendix D).

Request for Payment This event log is structurally similar to the prepaid travel costs. The unique case identifier *Rfp_id* is identical with the *concept:name* and the *RfpNumber* is the *Rfp_id* incremented by one as well in 6,321 cases and *UNKNOWN* in 565. The *Cost Type* is always 0 as well and there are 4 categorical attributes with IDs: *Project*, *Task*, *OrganizationalEntity* & *Activity*. The only

numerical attribute is the *RequestedAmount* where 99 % of the values are under 2,500 with a median of 94.25 but 3 cases with values higher than 10,000 and a max value of 982,873.80 (Further details in appendix E).

Permit Contrary to all the other logs, the permit log has not only significantly more case attributes, but is sparse as well. Empty fields can only be found in this log, with a mean sparsity of 0.099, nine out of ten case attribute fields are empty. Furthermore, this log contains multiple indexed attributes, where attribute names are concatenated to an index, e.g. *dec_id_1*, *dec_id_2*, *dec_id_3*, etc. Possible explanations for this will be given in section 2.3.

The three attributes unique to the permit log are *Overspent*, *OverspentAmount* & *TotalDeclared*. *Overspent* is a boolean value which is true in 32% of the cases and *OverspentAmount* is the corresponding float value which is between -14,952.33 and 1,457,490.15 with a mean of -106.95. *TotalDeclared* ranges from 0 to 1,458,578.75 and has a mean of 1,050.68.

While inspecting the different attributes of the permit log, we identified a small encoding inconsistency in the *RequestedAmount*. In the xes-file, the field is sometimes tagged as a string with one to five digits and sometimes it is tagged as a float value with multiple digits and a decimal point. The strings, when longer than 3 digits, are divided by a comma. We therefore inferred that the string tagged values are actually integers with a comma as a thousands separator and converted the field accordingly.

2.3 Derivation of Relations between Logs from Data

After describing each log individually, we will now analyse, how the logs intersect and what relations can be drawn. The total number of distinct activities across all logs is 59. By comparing this number with the distinct activities per log, it becomes apparent that these do overlap. By their name, 3 main groups can be identified, where each group is related to a subprocess: permit activities, declaration activities and request for payment activities as these start with a reference to the subprocess, e.g. *Permit SAVED by EMPLOYEE* or *Declaration REJECTED by SUPERVISOR*. Additionally, there are 5 supporting activities: *Start Trip*, *End Trip*, *Send Reminder*, *Request Payment & Payment Handled*, where the latter three are always executed by the organizational resource *SYSTEM*. It is important to specify that the log files do not only contain the activities related to one subprocess, but rather multiple subprocesses. The *PermitLog* file contains activities of all 3 subprocesses, *InternationalDeclaration* contains the declaration subprocess as well as the permit subprocess and the *PrepaidTravelCost* file contains the request for payment activities plus the permit activities. *DomesticDeclaration* encompasses only declaration related activities whereas the *RequestForPayment* file includes only request for payment activities.

International trips Since the *PermitLog* contains activities of all three subprocesses, we analysed if other attributes are also contained. An overview of the

related attributes is given in appendix F. Unfortunately, the attribute mapping between the logs is highly confusing. Therefore, we searched for an attribute that connects all logs as comprehensively as possible. We found that all *InternationalDeclaration.Permit ids* map on a value from *PermitLog.id*. Furthermore, all values from *PrepaidTravelCost.permit travel permit number* map on a value of *PermitLog.travel permit number*. We noted that both mappings are not bijective (1:1). Several InternationalDeclarations and PrepaidTravelCost cases can map on one PermitLog case, since a part of the permit cases is related to multiple declarations or prepaid travel costs. This is consistent with the description of the data set on the homepage which proclaims that the possibility of multiple declarations for one permit exists. However, from the relations we see that multiple reimbursements of prepaid travel costs are also existing. As there is only a very small number of cases where this mapping between those three logs is not consistent, we assume that *PermitLog.id* and *PermitLog.travel permit number* are foreign keys connecting declarations, permits and reimbursements of prepaid costs for a international travel.

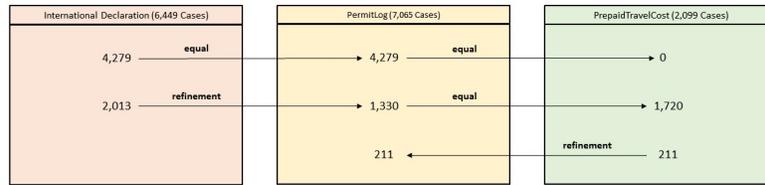


Fig. 4: Mappings between InternationalDeclaration, PermitLog and Prepaid-TravelCost

Afterwards, we analysed what the differences between the connected processes in each log are. Figure 4 shows the relation between the three logs. In 4,279 InternationalDeclaration cases the control-flow in InternationalDeclaration is identical to the control-flow of the process with the same key in PermitLog. These cases do not have any connection to the PrepaidTravelCost Log. In 2,013 cases, the process in PermitLog has more activities than the process in InternationalDeclaration. We call these cases a refinement. That is, the control-flow of one case can be transformed to one of another case, by only adding activities. All those permit cases have an equal control-flow as their related cases in PrepaidTravelCost (1,720 cases). In 211 cases, the related PermitLog case for a PrepaidTravelCost is also a refinement.

Apart from that, we found that in 448 cases, the *InternationalDeclaration.Permit id* refers to the PermitLog case with *Permit.Permit id* 423, but the permit does not link to these declarations (neither through case attribute nor by the subprocess activities). In those 448 cases, the *InternationalDeclaration.Permit ID* refers to a different permit, which cannot be found in the PermitLog. We will further

analyse this in the questions section. Furthermore, there are several cases in all three logs that have no connection to each other.

Other relation Contrary to our expectations, we could not find other relations between different logs by comparing attribute values pairwise. Especially between *DomesticDeclaration* and *PrepaidTravelCost*, no relation was found. We assumed that a reimbursement of prepaid travel costs is also possible for domestic declarations (see section 2.1 Assumptions). We also tried to find a relation by investigating the timestamps of the activities "Request Payment" and "Payment Handled" in both logs, as those are the only activities both logs share. Using this method, we found around 700 cases where the activities have been executed in the same timestamp. As this mapping was not as precise as expected, we stopped further investigation of this relation.

However, there are attributes like *id* in *PermitLog* that seem to store encoded information. Unfortunately, we were not able to find any meaningful relation or rule to decode that attribute and connect it to any other.

There are *PermitLog* cases that have no relation to any *InternationalDeclarations* or *PrepaidTravelCost* cases. Furthermore, there are *PrepaidTravelCost* cases that have no relation to *PermitLog*, i.e. no relation to a travel nor a permit. Most of them have *Permit travel permit number* of *UNKNOWN* or *travel permit 22777*.

2.4 Implications from Data

From the descriptive data analysis and the relations found between the different logs, several findings about the process could be gained.

International Declarations Approval and Reimbursement When planning an international trip, first a new entry in *PermitLog* is created once an employee files a request. This *PermitLog* case holds all relevant information about the travel, its declarations and reimbursements of prepaid travel costs. Furthermore, an entry in *InternationalDeclarations* is always created and is linked to the entry in *PermitLog* by its *Permit ID*. In both logs, the same activities are recorded. The process starts with the approval process where several roles (*ADMINISTRATION*, *BUDGET OWNER*, *DIRECTOR* etc.) have to authorize the travel permit. Afterwards, reimbursements for costs that had to be paid before the trip starts can be requested. If an employee decides to do so, a new entry in *PrepaidTravelCost* is created with the *Permits travel permit number* as field *permit travel permit number*.

International trips can basically be split in two groups. Those, for which reimbursements of prepaid travel costs were not requested and those for which reimbursements were carried out. Depending on this, the permit case has two different variants. If reimbursements for prepaid travel cost were not requested for that trip, the control-flow in *InternationalDeclarations* and *PermitLog* is the same. If prepaid travel costs were requested, the permit case has additional

activities from the case in *PrepaidTravelCost*. Both processes, the process for reimbursements of prepaid travel costs in the log *PrepaidTravelCost* and the process for reimbursements of travel costs in *InternationalDeclarations* have an approval part which is similar to the approval process for the general approval of a international travel in *PermitLog*. The *PermitLog* case combines the information from all three logs and holds additional information.

There can be multiple declarations and reimbursements of prepaid travel costs for a single travel and therefore a single *PermitLog* case. For each filed declaration, the permits case attributes *DeclarationNumber_0* to *DeclarationNumber_16* are filled with the respective *DeclarationNumber* from *InternationalDeclarations*. The declarations are assigned to *DeclarationNumber_x* from lower to higher indexes. For each prepaid travel cost request, the *PermitLog* case attributes *RfpNumber_0* to *RfpNumber_14* are filled in the same manner as the declarations. In both processes *InternationalDeclarations* and *PrepaidTravelCost*, there are two activities *Request Payment* and *Payment Handled* indicating that money is requested and paid. As *PermitLog* combines the activities from all three subprocesses, it has multiple occurrences of these two activities. One can determine the process these activities belong to by checking the event *id*.

Domestic travels Our initial assumption that prepaid travel cost can be reimbursed for all trips is not reflected in the data. We found no connection from the *DomesticDeclarations* to the *PrepaidTravelCost*. Moreover, the *PrepaidTravelCost* always contains case attributes relating to permit attributes. Since a permit is only filed for international declarations, *PrepaidTravelCost* seem to be only applicable to international declarations.

2.5 Matching of Process Understanding (BPMN) to Log Data

Following the presentation of the process understanding in section 2.1 and the data understanding in section 2.2, the two findings are now compared to each other. The objective was to align the process understanding, which is based solely on the textual description of the BPI Challenge 2020, and the data understanding, which is based only on the examination of the log files.

To start, we compiled a list of all activities found in the five logs. Next, duplicated entries of identically named activities in different logs were eliminated. Afterwards, the activities from the log were matched to their corresponding activities from the BPMNs. The sub-process "handle request" in figure 3 made it possible to simplify the presentation of all declarations and request approval steps.

All BPMN-activities could be aligned with one log-activity, while some log-activities remained without a counterpart:

- **PRE_APPROVER**: An involvement of this entity is never mentioned in the textual description

- **FOR_APPROVAL & FINAL_APPROVED:** The textual description does not distinguish between the steps *FOR_APPROVAL*, *APPROVAL* and *FINAL_APPROVAL*
- **SAVED by EMPLOYEE:** There is no indication that a declaration can be saved by an employee.
- **REJECTED by MISSING:** Lastly, the text does not mention the possibility of a rejection by a missing entity.

Otherwise, the understanding of the process coincided with the understanding from the descriptive analysis of the data. The detailed matching can be found in appendix G.

3 Questions

Q1: What is the throughput of a travel declaration from submission (or closing) to paying? The majority of the travel declarations has been paid and a throughput time can be calculated. For the international declarations, 95.94 % of the 6,449 declarations have been paid, whereas the domestic declarations have a slightly slower payment rate of 95.62 % of the 10,500 cases. The remaining cases are either never approved or just saved as drafts that have not been submitted. To measure the duration of one declaration, we divided the processing time into approval and payment time. The first duration measures the time delta between the initial submission by the employee and the payment request. The latter measures the delta between payment request and payment handling. In general, a declaration has a mean approval time of 9 days and 5 hours with a standard deviation of 16 days and 18 hours. This large deviation will be explained later on, by some extended executions. The 75% quantile is at 9 days and 22 hours. The payment is in average executed in 3 days and 13 hours with a considerably lower standard deviation of 2 days and 6 hours.

Q2: Is there a difference in throughput between national and international trips? In comparison to the domestic declarations, the international declarations have a higher mean approval duration of 11 days whereas the domestic declarations are on average handled within 8 days. The payment process shows no crucial time difference. In figure 5 it can be seen that the majority of cases are executed in the summer. The decrease of cases in August coupled with a slight increase in approval duration might indicate a holiday season. In line with the textual description, there are less cases in 2017, since only two departments participated during that year. The few records in 2019 might indicate that only running instances were recorded. This assumption is reflected in the data, as all activities in 2019 refer to the end of a declaration process, which also explains the linearly growing approval time.

Q3: Are there differences between clusters of declarations, for example between cost centers/departments/projects etc.? For the analysis of

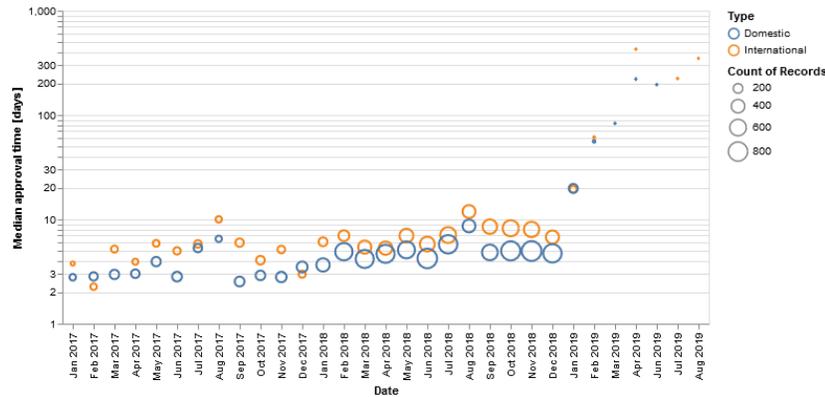


Fig. 5: Declaration throughput

clusters and differences in organizational units, projects or tasks, we restrict the data to the international declarations, as only this log contains the additional information.

The amount of declarations submitted by each unit varies notably. While half of the units submit 1 to 20 declarations in total, the leading organizational units submitted more than 1,400 declarations each.

On the project level, there are some unusually high median handling times (approval and payment combined): for project 4339 more than 21 days & projects 5126, 3352, 6277, 3595, 19312, 3062, 11292, 750, 17142 with more than 14 days. All aforementioned projects are related to at least 10 declarations. Projects with less declarations were not considered due to the lack of statistical significance.

The task 429 is connected to most of the declarations (3,910), while a further 2,238 cases have an *UNKNOWN* task number. The other 4 tasks are matched to less than 15 declarations each with a lower median execution time.

A small correlation between amount and median execution time is identifiable. Amounts of less than 769 have a median execution time of 9 days, while amounts between 767 and 1,538 have a median execution time of 11 and higher amounts have a median execution time of 13 days and higher.

Q4: What is the throughput in each of the process steps, i.e. the submission, judgement by various responsible roles and payment? The throughput of each process step roughly follows the general throughput of the process, although a clear change of process can be observed between 2017 and 2018. While there are some actions of a pre-approver in 2017, they are replaced in 2018 by actions of the administration and budget owners (see appendix H). While approval and payment activities show a similar curve, appendix H concentrates only on approval activities. The number of actions includes approval as well as rejection, since there was no clear difference in the behaviour of the two groups.

While evaluating the throughput times, we noticed approval steps with an execution time of just several seconds. For example, the *Declaration APPROVED by ADMINISTRATION* has a median execution time of 3 minutes, but in 696 cases it is executed less than 5 seconds after *Declaration SUBMITTED by EMPLOYEE*. It is not very likely for this quick reaction time to be possible in practice. Either the submitting employee and the administration are the identical person, or the approval step is done automatically by the system. Another example are *Declaration APPROVED by SUPERVISOR* and *Declaration FINAL_APPROVED by DIRECTOR*. These events can occur in the same second, which suggests that one event is set automatically by the system. In consequence, the log is likely to contain real human interaction as well as system actions without a flag to distinguish between the two.

Q5: Where are the bottlenecks in the process of a travel declaration?

To determine the bottlenecks in the process of a travel declaration, we calculated the duration of an activity for each case. Based on this, we calculated statistics over all cases including mean, median, min, max and std. We examined domestic declarations and international declarations separately. The figure in appendix J lists the activities with a median duration exceeding one day for both, domestic and international declarations. We select the activities with the highest bottlenecks based on the median and not based on the mean duration, since we do not want to put too much emphasis on outliers.

In figure J the activities with a high median duration for both domestic and international declaration are those regarding the payment of the travel. Also, for the international declaration there is a relatively large idle time between the permit request and the submission of a declaration by the employee. Furthermore, it is noticeable that the activities start trip and end trip have very high median and mean duration. This makes sense, since the travel request is often made a few days before the start of the trip.

Q6: Where are the bottlenecks in the process of a travel permit (note that there can be multiple requests for payment and declarations per permit)?

To answer Q6, we followed the same procedure as described in Q5, while examining the permit log. The results are in line with those from Q5 as shown in figure 6. The activities Start trip, End Trip and Payment Handled have a high median duration greater than one day. Next to that, the activity *Declaration SUBMITTED by EMPLOYEE* has a high median and mean duration. If we look at the involved roles we can see that *SUPERVISOR*, *EMPLOYEE* and *UNDEFINED* need the most time (figure 7).

Q7: How many travel declarations get rejected in the various processing steps and how many are never approved? To answer Q7, we had to examine all activities that suggest a rejection or an approval of the declaration. Therefore, we took all activities into account that include the terms "reject" and

	mean	median	min	max	std
Permit FINAL_APPROVED by SUPERVISOR	24 days 02:01:23.083227	10 days 10:31:47	0 days	391 days 02:50:11	34 days 09:48:33.502620
End trip	20 days 08:55:51.645576	6 days 10:24:20	0 days	470 days 12:45:45	38 days 17:24:26.546282
Start trip	5 days 16:04:26.835951	4 days 00:00:00	0 days	1102 days 00:00:00	17 days 21:54:51.613274
Request Payment	2 days 18:43:58.479547	2 days 23:39:25	0 days	29 days 01:27:49	2 days 04:22:31.666242

Fig. 6: Activities of the permit log whose median duration exceeds one day

	mean	median	min	max	std
UNDEFINED	9 days 21:15:03.844302	3 days 05:35:56	0 days	353 days 06:00:40	19 days 05:04:33.810751
EMPLOYEE	9 days 13:43:04.013729	3 days 15:07:50	0 days	1102 days 00:00:00	26 days 09:40:32.943077
SUPERVISOR	2 days 12:56:46.926397	1 days 00:12:28	0 days	86 days 23:27:32	3 days 20:54:23.368203

Fig. 7: Roles in the permit log whose median duration exceeds one day

”approved”. To this end, we firstly computed a case activity matrix that counted the occurrence of each activity for each case. This allowed us to count the occurrences of rejection and approval activities. We again analysed domestic and international declarations separately. Figure 8 lists all rejections and approvals per activity within the domestic declarations. Afterwards we inspected the domestic declaration on case level. In total, 365 out of 10,500 domestic declarations have never been approved. Interestingly, out of these 365, the declarations *90815* and *95149* have not been approved, however the payment was completed.

Declaration REJECTED by MISSING	91.0
Declaration REJECTED by PRE_APPROVER	86.0
Declaration REJECTED by EMPLOYEE	1365.0
Declaration REJECTED by SUPERVISOR	293.0
Declaration REJECTED by ADMINISTRATION	952.0
Declaration REJECTED by BUDGET OWNER	59.0

(a) Rejections

Declaration FINAL_APPROVED by SUPERVISOR	10131.0
Declaration APPROVED by PRE_APPROVER	685.0
Declaration APPROVED by ADMINISTRATION	8202.0
Declaration APPROVED by BUDGET OWNER	2820.0

(b) Approvals

Fig. 8: Rejections and Approvals in the Domestic Declarations

For the international declarations, we considered approvals and rejections both during the permit request and the travel declaration. We started again inspecting the activities that contain ”REJECT” or ”APPROVED” and counted their occurrences. Appendix I lists all rejections and approvals within the international declarations.

Then we had a look at the cases. Out of the 6,449 cases, in 5,150 the permit and the declaration had been approved. For 614 cases the permit and the declaration has not been approved. For 279 cases the permit has been approved but the declaration has never been approved and for 406 the declaration has been approved without approval of the permit.

Q8: How many travel declarations are booked on projects? For international declarations the mapping to which project a declaration belongs is stored in the *Permit Projectnumber*. 2,333 declarations are assigned to the project *UNKNOWN*, which probably means that they are not booked on any project. The other 4,116 international declarations are assigned to specific projects. The projects with the most booked travel declarations are *projects 426, 3442 and 8761*. Unfortunately, we were not able to reproduce the relation between domestic declarations and projects based on the given event logs.

Q9: How many corrections have been made for declarations? To answer this question we looked at two things. First, we checked if the activity *Declaration submitted by employee* is executed multiple times. Second, we evaluated if the cases end properly with a payment or if they are aborted. For domestic declaration there are 1,166 re-submissions appearing in 1,019 different cases. Of these 1,019 cases, 22 cases include only rejections and no approvals and 997 cases include both rejection and approval activities. 978 cases are properly finished with a payment handled. For the international declarations, we also distinguish between re-submissions of permits and re-submissions of declarations. There are 254 re-submissions of permits in 234 cases and 1,724 re-submissions of declarations in 1,405 cases. 1,345 of the resubmitted declarations are properly completed with the activity *Payment Handled*, 56 cases are rejected and have not been approved, 4 cases have been approved and rejected and have not been completed with a payment. If we look at the resubmitted permits, the 254 re-submissions occur in 234 cases. 17 of these cases have been rejected and never been approved. 213 cases have been successfully completed and 4 cases have been finally rejected or the payment is still pending.

Q10: Are there any double payments? To answer this question, we initially collected ideas what a double payment could be. From the data, one can see that a payment is indicated by the activity *Payment Handled*. It occurs together with activity *Request Payment* that is usually executed before *Payment Handled*. A double payment could have happened, (1) if *Payment Handled* happened twice within a single case. As some international trips have multiple declarations and prepaid travel cost reimbursements, we also checked, (2) if the same declaration or prepaid travel cost was used twice within one *PermitLog* case or (3) if *Payment Handled* was executed more often than *Request Payment* within a single permit. Furthermore, we analysed (4) whether the number of *Payment Handled* activities corresponds to the number of different filed prepaid travel cost and declaration reimbursements.

From all five logs, only the *PermitLog* has cases where *Request Payment* or *Payment Handled* occurs more than once within a case. As the *PermitLog* captures all activities from *InternationalDeclarations* and *PrepaidTravelCost*, the payment activities of both processes are shown in *PermitLog*. As there are no relations from *DomesticDeclaration* or *RequestForPayment* to any other log, double payments did not happen for those.

In two *PermitLog* cases (IDs 36421 and 76646) the number of *Request Payment* and *Payment Handled* activities differs. In the first case, *Payment Handled* was executed twice but *Request Payment* only once. After checking the related field *InternationalDeclarations* processes, we saw that in permit case 36421 the activity *Request Payment* has not been executed but *Payment Handled*. In the second permit case, *Request Payment* has been executed three times but *Payment Handled* only twice. Although *Request Payment* was not executed in one of the declarations of permit case 36421, the suspicious declaration got approved by the supervisor. Thus, the payment could still be correct. In the second case, one payment that was requested, was not handled. Therefore, no double payment could be found by checking for (1) and (3).

No cases could be found where the same declaration number or prepaid travel cost was used multiple times within the same permit (2) by checking the attributes *DeclarationNumber_0* to *DeclarationNumber_16* and *RfpNumber_0* to *RfpNumber_14*. However, there are 7 *PermitLog* cases, where *UNKNOWN* was used multiple times as declaration or prepaid travel cost number. Furthermore, there are 87 cases where the number of *Payment Handled* activities does not match the number of filed prepaid travel cost or reimbursement by checking the IDs in *PermitLog* (4). We further analysed those cases but did not find any case, where more *Payment Handled* activities were executed as there were declaration or prepaid processes. In all cases, less *Payment Handled* activities were performed, indicating that some declarations or prepaid travel costs processes got rejected. Overall, we conducted that there is no solid evidence for a double payment, except permit case 36421, where the approval seems to be suspicious.

Q11: Are there declarations that were not preceded properly by an approved travel permit? Or are there even declarations for which no permit exists? We will answer the first question first and then the second. As a permit is only required for international declarations, we will not check domestic ones.

Declarations are approved in several steps by multiple different roles (*ADMINISTRATION*, *BUDGET OWNER*, *PRE_APPROVER*, *SUPERVISOR* and *DIRECTOR*). For now, we do not know what the proper approval process is (neither for permit nor any other process in the data). Therefore, we started with very simple assumptions and extended them. First, we checked if each permit with declaration activities has at least one activity *Permit APPROVED by ...* or *Permit FINAL_APPROVED by* This assumption is true. Note that there are 553 cases that only have the *Permit FINAL_APPROVED by ...* activity but no other. As all such *PermitLog* cases are approved, we next checked if there are cases where the permit was approved first, but then rejected. As an employee can resubmit the permit after being rejected, we went through all loops until no more rejecting was found. Finally, we did not find any case that was not approved after being rejected. Even after multiple submission-rejection-loops, the permit always got approved. Furthermore, we found travel cases, where the reimbursement of prepaid travel cost, the declaration or the travel itself started

and were even finished before the permit got finally approved⁷. This is not in accordance with the target process.

Next, we tried to find rules, when a permit has to be checked by a certain role (e.g. trips with costs above 500 must be approved by a director). Unfortunately, we did not find such an attribute in the *PermitLog* cases nor did we find any hard condition, explaining the different approval sequences (e.g. *ADMINISTRATION* → *BUDGET OWNER* → *DIRECTOR*). However, this does not mean that there is no proper preceding process. From what we knew from the data description and target process, we found that most permits are approved in multiple steps but 553 are approved only by the a director or supervisor without previous steps. We assume that there are certain rules to follow in the approval process.

From section 2.3 we know that for all international trips a permit is assigned. However, there are 449 international declarations that have the same permit assigned but neither have the same control-flow nor a prepaid travel cost case assigned. They all map on travel permit 423. Different to all other declarations, the attribute values from *InternationalDeclarations* are different from the values in *PermitLog*. Therefore, the 449 international declarations have a permit assigned but the permit is not the correct one (travel permit 423 could be a dummy permit).

Q12: How many travel declarations are submitted by the traveler and how many by a mandated person? To answer this question we firstly inspected the activity names. There is only one starting with *Declaration SUBMITTED by* and that is *Declaration SUBMITTED by EMPLOYEE*. So the activity names give no indication, if the declaration was submitted by the traveler or a mandated person. Next, the *org:role* and the *org:resource* where considered. Unfortunately, the activity of submitting the declaration is always executed by the same role and resource (*EMPLOYEE* & *STAFF MEMBER*). For the international declarations we searched for differences between the submission activity and the start trip activity, but found no indication as well. We therefore conclude that this question is not answerable with the given data.

Q13: How many travel declarations are first rejected because they are submitted more than 2 months after the end of a trip and are then re-submitted? This evaluation can only be done for international declarations, as there is no travel date in the domestic declarations log. Of the 6,449 international declarations, 6,141 were submitted in time, i.e. less than 2 months after the last travel day, whereas 308 declarations are submitted after this deadline. Of those, 234 are only submitted once, leaving 74 declarations that were submitted after the deadline and resubmitted after the initial rejection.

Q14: Is this different between departments? Yes, there are differences between departments: 13 departments never submitted a declaration too late,

⁷ *travel permit 40269, 45030, 73110, 73125, 84065, 8770*

while 6 departments submitted less than 5% of their declarations too late and another 5 departments less than 10%. Front runners are department 65470 with 10% late submissions, department 65465 with 12,5% and department 65480 with 55% (see appendix K). When a department submitted declarations after the deadline of two months, they usually tried to resubmit a subset of the late submissions. Only the three departments 65465, 65470 & 65467 never tried to resubmit a late declaration that was rejected priorly.

Q15: How many travel declarations are not approved by budget holders in time (7 days) and are then automatically rerouted to supervisors? To answer this question several assumptions had to be made, in order to calculate a deadline. The textual description of the process and the related questions give no indication, as to when the 7 days deadline starts counting. We therefore took two mutually exclusive assumptions and tried to answer the question for both:

- The time starts when the declaration is submitted.
- The time starts when the declaration is approved by the administration.

For both assumptions, the question appeared unanswerable, as there is no clear indication in the data, if and when a declaration is redirected. This is due to two insights: The budget owner can still approve or reject declarations, even if the deadline has passed and on the other hand, the supervisor can approve a declaration before the 7 days deadline and thereby eliminate the need for an approval of the budget owner. Furthermore, the text specifies that the budget owner and the supervisor can be the same person. In this case only one approval step is needed, but the text does not specify, which activity will be recorded in the the log. Next to these complications, there was no indication in the data when the budget owner should be involved. The fields *Permit BudgetNumber* and *BudgetNumber* are always filled, and the *Permit RequestedBudget* is not correlated with the involvement of the budget owner. Hence, there is no way to find process violations. Similar problems arise when evaluating director activities.

Q16: Next to travel declarations, there are also requests for payments. These are specific for non-TU/e employees. Are there any TU/e employees that submitted a request for payment instead of a travel declaration? We have not been able to identify any information in the data that would allow us to detect with absolute certainty any incorrectly submitted request for payments. However, we compared the requested amounts of the *RequestForPayment* log and the amounts listed in the *DomesticDeclaration* and *InternationalDeclaration* logs. We revealed some suspicious overlaps. There are 116 matches between request for payments and domestic declarations and 18 matches between request for payments and international declarations. Apart from one amount that is 0.00 the numbers look rather rare, and it is therefore improbable that the same amount is used multiple times. If we add the time perspective to our analysis, we can further reduce the list of potential anomalies. If

a travel declaration and a request for payment show exactly the same amount, this is noticeable. However, if the two transactions were made at two completely different points in time, it seems to be just a coincidence. We have therefore further filtered the data for whether they were made in similar time periods. More specifically, we checked whether the payment activities *Request Payment* and *Payment Handled* of a request for payment and a declaration were executed within a period of five days. After this further filtering, 55 domestic declarations and 7 international declarations remained.

4 Advanced Analysis

In addition to the general descriptive analyses that we used to answer the questions, we applied additional algorithms to extract anomalies in the process flow as well as deviations of the process flow that is mapped in the data and the modelled processes that contain the explicit rules of the textual description. For this purpose, process conformance approaches and rule mining algorithms were applied. Using these techniques we tried to get further insights into the travel process.

4.1 Conformance Checking

In order to check the conformance of the data we use some typical process conformance algorithms as Process Alignment and Token Based Replay (TBR) algorithms to check if the processes which result from the provided data files comply with the process which was described on the website and then, was modelled as a BPMN 2.0 process model. The first step is to transform the process models from section 2.1 into a petri net. We assume that the description contains our basic truth and represents it in its entirety. Since no other requirements are otherwise specified for the process, these are adopted 1:1 under the described assumptions in section 2.1.

Token Based Replay (TBR) provides results, but these are difficult to transfer to the overall context. It replays the log and compares it to a petri net step by step. If an activity in the model is performed only in the log but not in the model, this is called log move. The other direction, a move in the model but not in the log is called model move. A move in the model and in the log is called synchronous move. TBR can be used in several characteristics using several parameters but has also several limitations which can strongly influence the quality of the results[6].

Overall, the fitness is very low around 0.1 over the entire log. This is mainly due to the imprecise specification of the process on the website. This leads to very imprecise results. Also the missing events, which are only in the data but not in the textual description and therefore in the model to be tested, lead to a very bad fitting. The TBR clearly shows that these missing activities have a great influence on the model. To be able to deliver better results, the petri net,

which is derived from the BPMN model, would have to be further revised in several steps to really represent the actual process.

Alignments are another solution to fit a given process model to data and check if the log data is in conformance with the model. Alignments use search-like algorithms to match the process data step by step to a model by searching e.g. for paths [1].

The same behaviour could be found when using alignments. First a smaller model was used to tune the model and then it was extended to the whole data set. Since the identical data set and the petri net, which is based on the process description of the website, were used, the same problems occur, especially the different degrees of granularity and the insufficient description lead to an unsatisfactory fit of the model.

4.2 Decision Tree Generation

Mining a Decision Tree (DT) enables us to uncover hidden structures in the event data [3]. Looking at the Decision Tree in figure 9⁸ unveils that if *Permit.TotalDeclared* is over 1,388, a lot of cases (1,351 cases) have a much simpler representation as most of the samples follow a straight way down (5,461 samples). This is also indicated by the higher gini score on the left hand side.

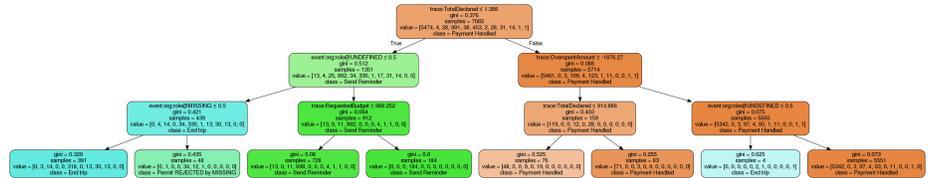


Fig. 9: Decision Tree mined using Gini impurity and a max depth of 3 with concept:name as target

By looking at a DT which uses the role of an user as target (see appendix M) one thing that attracts attention is a sub-tree that only includes a missing role. It also becomes clear that the various roles have very different tasks, some of which overlap with those of the employees. This could lead to general irritations of the roles in the system and should therefore be investigated with the help of a precise role description in consultation with the system managers. It has been noticed that special roles correlate more strongly with resources of the system, e.g. the probability that an employee is involved is very high. On the other hand, there are exceptions to this rule which need to be examined more closely. These could be, for example, incorrectly assigned users or system bots that act like users.

⁸ The full DT can be accessed via <https://bpm.dfki.de/bpi-challenge-2020-notebooks/>

Another possible DT aims at splitting according to process execution times. We have tried to find temporal regularities or anomalies, but we assume a minimum run-time of 10 days up to 200 days for the mining of the Decision Tree. However, no regularities worth mentioning could be found here.

4.3 Behavioural Profiles

Behavioural profiles [7] are based on relations that define the arrangement of two activities in each of the traces. These are either strictly arranged (i.e. activity *a* always follows activity *b*), have no strict arrangement, or they are mutually exclusive (i.e. activity *a* and activity *b* never occur together in a trace). The properties are computed per log and for each pair of activities.

The expressiveness of the following behavioural profiles is restricted, due to the nature of the approval process. The process allows for revisions of the same case, e.g. a declaration can be rejected by a director after being approved by the administration and the supervisor, then be resubmitted and again be approved by the different entities. Since behavioural profiles look at traces of an event log, relations on a per revision basis may not be visible, e.g. the relation that a supervisor approval is always preceded by an administration approval is only true per iteration, but not on a case perspective.

Domestic declarations In the domestic declarations log, there exist three *Declaration FOR_APPROVAL by ...* activities. Those relating to the *SUPERVISOR* or *PRE_APPROVER* are always preceded by a submission of the employee and followed by a *Declaration REJECTED by MISSING*. Therefore, these events are rightfully mutually exclusive to payment activities. Interestingly, although most activities strictly proceed the payment activities, there are declarations that have a *Declaration REJECTED by EMPLOYEE* after *Payment Handled*. Also, there have been *Declaration REJECTED by MISSING* after *Request Payment*, but before *Payment Handled*, which implies an organizational entity within the payment process, that is able to veto a declaration. This is not reflected in the textual description of the process. Furthermore, there is an activity *Declaration SAVED by EMPLOYEE* that is mutually exclusive to all activities, except the payment activities, i.e. declarations were paid without any approval step. An exemplary behavioural profiles matrix is shown in appendix N.

International declarations This log contains the permit and the declaration sub-process. Ideally, the permit is approved before the trip starts. Nearly all permit related activities can occur after the event *Start trip*, which means that the international trip was taken without a finalized permission. There are even *Declaration SUBMITTED by EMPLOYEE*, before *Permit SUBMITTED by EMPLOYEE* although there are no approval steps, before the permit is submitted. Also, the payment of a declaration can occur before and after *End trip*, which is inconsistent with the textual description of the process. And like in the

domestic declarations, there are payment requests that are followed by a *Declaration REJECTED by MISSING*. Moreover, there are payment requests that occur after the payment has been handled.

Prepaid travel costs A similar behaviour can be observed in the requests for prepaid travel costs. There are requests being submitted before all permit activities are finished and again, payments are firstly handled and then requested. There are cases as well, where a *Request Payment* is followed by *Declaration REJECTED by MISSING*.

Request for payment This log is not related to any permit, but otherwise it behaves similar to the prepaid travel costs: *Request payment* after *Payment Handled* and requested payments are REJECTED by MISSING, but there are no cases, where a request is rejected by an employee after it was payed. Then again, if cases are approved by the budget owner, the payment is immediately handled, without a payment request.

Permit log contained no additional relations, but the concurrency of requests for prepaid travel cost reimbursement and declarations could be observed.

5 Main Findings and Conclusion

In this report, we described our findings from the analysis of a travel process in the scope of the BPI Challenge 2020. The task was to answer specific questions of the organizers, as well as to reveal unknown interdependencies and suspicious behaviour in the event logs. For this purpose, we set up some assumptions about the process in section 2.1, as the description on the homepage left some open questions and room for interpretation. In line with these assumptions a target process model was developed, which functioned as a baseline for the following work. In section 2.3, we analysed the relations between the logs and found that only three logs are connected. Furthermore, the content of the data is hard to capture, which made it very hard to map the data to the target process. This also became apparent by some characteristics that are in the target process but could not be found in the data. We could, e.g. not find relations between the travel process of domestic declarations and prepaid travel cost or between international declarations and request for payment.

Based on the insights and assessments we drew from the event logs and the corresponding process description of the BPI Challenge, we answered the specific questions posed by the challenge owner in section 3. Apart from one question that was unanswerable because crucial information was not accessible, we were able to provide a solution for the other questions. We collected further interesting information and relations while solving the questions. For example, the concept drift that was mentioned in the textual description of the process, was clearly

recognisable during the analysis, so that we had to separately elaborate the behaviour in 2017 and 2018 for a few questions.

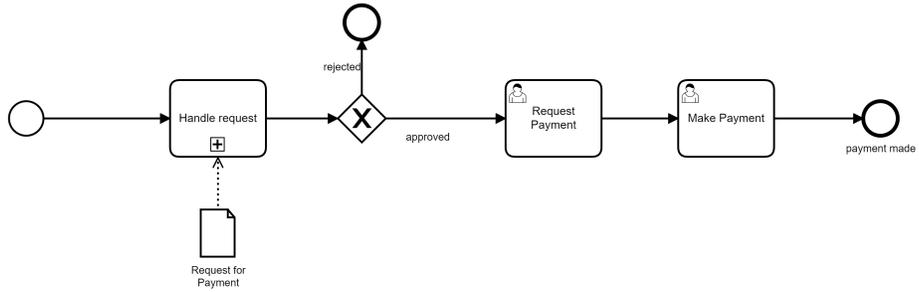
In the further course of the work, we tested some more advanced techniques like Conformance Checking algorithms, Decision Trees and Behavioural Profiles to find out more about the given process data. However, since some of these technologies require an explicit process description of high quality, they may be of little use. Unfortunately, the target process description and the derived actual process are very imprecise. Mining of Decision Trees and Behavioural Profiles of processes can be useful for evaluation and also help to implement rules in the system and refine workflows in them. However, it should be noted that a better process description could have achieved significantly better results.

To gather better results, the target process should be described in more detail to get deeper insights in the regulations and characteristics of the process flow. In this regard, the relations between the five logs and the systems which produce the logs, that capture different (sub-) parts of the travel process, should be made clear. We suspect, that the intention of TU/e was to separate the travel process into sub-processes to avoid redundancies. However, splitting the data causes a lot of disorder. Additional, the given roles which are associated with the log are more organizational entities and do not state the role of a employee (e.g. manager or director). This would be helpful to mine an organizational tree from the data and discover the division structures to assure the permit flow.

We motivate to use a clear naming convention for logs, activities and attributes to keep the system maintainable. There is, e.g. a log called Request-ForPayment and several activities with request for payment that are not related to each other. Furthermore, a clear event to indicate a rejection or approval of a permit, travel declaration or other processes would be desirable. Likewise, an unambiguous identifier for related processes in different information systems (logs) would be of great use.

In summary, the entire work on the BPI-Challenge has led to many discussions, suggestions and ideas that are being pursued in various directions. There, the semi-automatic evaluation of logs and ensuring process performance play a major role at our ongoing research.

A BPMN - Request for payment



B Field description for domestic declarations log

field	Info. Density	Special values	unique values	type
id	1	-	56.437	ID
org:resource	1	-	2	categorical
concept:name	1	-	17	categorical
time:timestamp	1	-	45.403	datetime
org:role	1	UNDEFINED MISSING	7	categorical
case:id	1	-	10.500	ID
case:concept:name	1	-	10.500	ID
case:BudgetNumber	1	-	1	ID
case:DeclarationNumber	1	UNKOWN	10.049	ID
case:Amount	1	-	8.326	numerical

C Field description for international declarations log

field	Info. Density	Special values	unique values	type
id	1	-	69.073	ID
org:resource	1	-	2	categorical
concept:name	1	-	34	categorical
time:timestamp	1	-	51.270	datetime
org:role	1	UNDEFINED MISSING	8	categorical
case:Permit travel permit number	1	UNKNOWN	5.596	ID
case:DeclarationNumber	1	UNKNOWN	6.190	ID
case:Amount	1	-	6.100	numerical
case:RequestedAmount	1	-	6.100	numerical
case:Permit TaskNumber	1	UNKNOWN	6	ID
case:Permit BudgetNumber	1	-	207	ID
case:OriginalAmount	1	-	6.100	numerical
case:Permit ProjectNumber	1	UNKNOWN	825	ID
case:concept:name	1	-	6.449	ID
case:Permit OrganizationalEntity	1	-	27	ID
case:travel permit number	1	-	6.033	ID
case:Permit RequestedBudget	1	-	5.259	numerical
case:id	1	-	6.449	ID
case:Permit ID	1	-	6.028	ID
case:Permit id	1	-	5.608	ID
case:BudgetNumber	1	UNKNOWN	719	ID
case:Permit ActivityNumber	1	UNKNOWN	145	ID
case:AdjustedAmount	1	-	6.101	numerical

D Field description for prepaid travel costs log

field	Info. Density	Special values	unique values	type
id	1	-	16.205	ID
org:resource	1	-	2	categorical
concept:name	1	-	29	categorical
time:timestamp	1	-	14.070	datetime
org:role	1	UNDEFINED MISSING	8	categorical
case:Rfp_id	1	-	2.099	ID
case:Permit travel permit number	1	UNKNOWN	1.318	ID
case:Task	1	UNKNOWN	435	ID
case:OrganizationalEntity	1	-	26	ID
case:RequestedAmount	1	-	2.049	numerical
case:Activity	1	UNKNOWN	5	ID
case:Permit TaskNumber	1	UNKNOWN	5	ID
case:Permit BudgetNumber	1	UNKNOWN	139	ID
case:Permit ProjectNumber	1	UNKNOWN	403	ID
case:Project	1	UNKNOWN	29	ID
case:concept:name	1	-	2.099	ID
case:Permit OrganizationalEntity	1	UNKNOWN	22	ID
case:Permit RequestedBudget	1	-	1.296	numerical
case:Cost Type	1	-	1	ID
case:Permit id	1	UNKNOWN	1.327	ID
case:Permit ActivityNumber	1	UNKNOWN	58	ID
case:RfpNumber	1	UNKNOWN	2.000	ID

E Field description for request for payment log

field	Info. Density	Special values	unique values	type
id	1	-	36.796	ID
org:resource	1	-	2	categorical
concept:name	1	-	19	categorical
time:timestamp	1	-	30.141	datetime
org:role	1	UNDEFINED MISSING	8	categorical
case:Rfp_id	1	-	6.886	ID
case:Project	1	UNKNOWN	79	ID
case:Task	1	UNKNOWN	597	ID
case:concept:name	1	-	6.886	ID
case:Permit OrganizationalEntity	1	-	36	ID
case:Cost Type	1	-	1	ID
case:RequestedAmount	1	-	6.449	numerical
case:Activity	1	UNKNOWN	6	ID
case:RfpNumber	1	UNKNOWN	6.322	ID

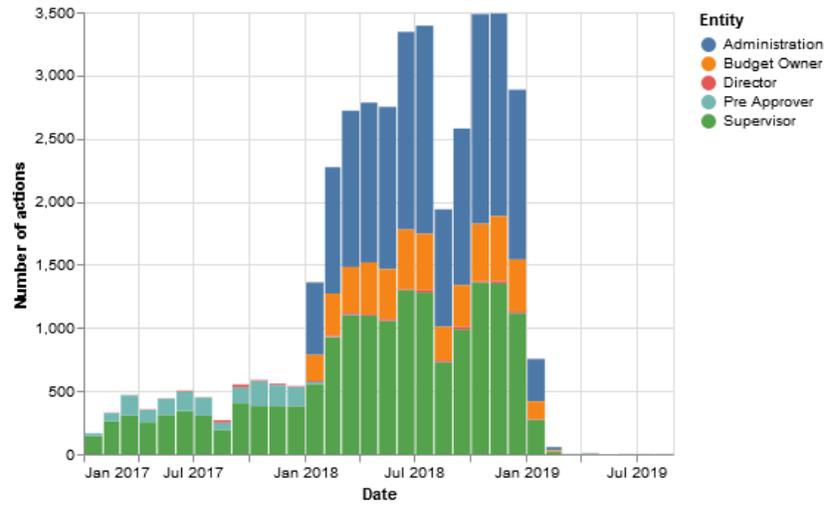
F Attribute relations of the permit log

International decl.	Permit	Prepaid travel Cost
	Activity_14	← Activity
Permit ActivityNumber	← ActivityNumber	→ Permit ActivityNumber
Permit BudgetNumber	← BudgetNumber	→ Permit BudgetNumber
	Cost Type_14	← Cost Type
id	↔ dec_id_16	
DeclarationNumber	↔ DeclarationNumber_16	
Permit ID	↔ id	↔ Permit id
Permit id	↔ id	
Permit OrganizationalEntity	← OrganizationalEntity	→ Permit OrganizationalEntity
	OrganizationalEntity_14	← OrganizationalEntity
	Overspent	
	OverspentAmount	
	Project_14	← Project
Permit ProjectNumber	← ProjectNumber	→ Permit ProjectNumber
RequestedAmount	→ RequestedAmount_16	→ RequestedAmount
Permit RequestedBudget	← RequestedBudget	→ Permit RequestedBudget
	Rfp_id_14	↔ Rfp_id
	RfpNumber_14	↔ RfpNumber
	Task_14	← Task
Permit TaskNumber	← TaskNumber	→ Permit TaskNumber
	TotalDeclared	
Permit travel permit number	↔ travel permit number	↔ Permit travel permit number
travel permit number	↔ travel permit number	

G Matching of data fields to the BPMN-model

Data View (Log)	Process View (textual description, BPMN)		
Activity	Lane	Activity	Notes
Declaration REJECTED by SUPERVISOR	Supervisor	Send rejection notification	-
Declaration APPROVED by ADMINISTRATION	Travel Administration	Forward to Budget owner	-
Declaration APPROVED by BUDGET OWNER	Budget Owner	Send Approval	-
Declaration APPROVED by PRE_APPROVER	-	-	No Pre Approver in Description
Declaration APPROVED by SUPERVISOR	Supervisor	Send approval	-
Declaration FINAL_APPROVED by DIRECTOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Declaration FINAL_APPROVED by SUPERVISOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Declaration FOR_APPROVAL by ADMINISTRATION	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Declaration FOR_APPROVAL by PRE_APPROVER	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Declaration FOR_APPROVAL by SUPERVISOR	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Declaration REJECTED by ADMINISTRATION	Travel Administration	Reject request	-
Declaration REJECTED by BUDGET OWNER	Budget Owner	Send rejection notification	-
Declaration REJECTED by DIRECTOR	Director	Send rejection notification	-
Declaration REJECTED by EMPLOYEE	Employee	Reject request	-
Declaration REJECTED by MISSING	-	-	No Information in Description
Declaration REJECTED by PRE_APPROVER	-	-	No Pre Approver in Description
Declaration REJECTED by SUPERVISOR	Supervisor	Send rejection notification	-
Declaration SAVED by EMPLOYEE	-	-	No Information in Description
Declaration SUBMITTED by EMPLOYEE	Employee	Submit request	-
End trip	no lane	Trip ended	-
Payment Handled	no lane	Make Payment	-
Permit APPROVED by SUPERVISOR	Supervisor	Send Approval	-
Permit SAVED by EMPLOYEE	-	-	No Information in Description
Permit APPROVED by ADMINISTRATION	Travel Administration	Forward to Budget owner	-
Permit APPROVED by BUDGET OWNER	Budget Owner	Send Approval	-
Permit APPROVED by PRE_APPROVER	-	-	No Pre Approver in Description
Permit APPROVED by SUPERVISOR	Supervisor	Send Approval	-
Permit FINAL_APPROVED by DIRECTOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Permit FINAL_APPROVED by SUPERVISOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Permit FOR_APPROVAL by ADMINISTRATION	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Permit FOR_APPROVAL by SUPERVISOR	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Permit REJECTED by ADMINISTRATION	Travel Administration	Reject request	-
Permit REJECTED by BUDGET OWNER	Budget Owner	Send rejection notification	-
Permit REJECTED by DIRECTOR	Director	Send rejection notification	-
Permit REJECTED by EMPLOYEE	Employee	Reject request	-
Permit REJECTED by MISSING	-	-	No Information in Description
Permit REJECTED by PRE_APPROVER	-	-	No Pre Approver in Description
Permit REJECTED by SUPERVISOR	Supervisor	Send rejection notification	-
Permit SUBMITTED by EMPLOYEE	Employee	Submit request	-
Request For Payment APPROVED by ADMINISTRATION	Travel Administration	Forward to Budget owner	-
Request For Payment APPROVED by BUDGET OWNER	Budget Owner	Send Approval	-
Request For Payment APPROVED by PRE_APPROVER	-	-	No Pre Approver in Description
Request For Payment APPROVED by SUPERVISOR	Supervisor	Send Approval	-
Request For Payment FINAL_APPROVED by BUDGET OWNER	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Request For Payment FINAL_APPROVED by DIRECTOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Request For Payment FINAL_APPROVED by SUPERVISOR	-	-	No dif. btw. FINAL_APPROVAL and APPROVED
Request For Payment FOR_APPROVAL by ADMINISTRATION	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Request For Payment FOR_APPROVAL by SUPERVISOR	-	-	No dif. btw. FOR_APPROVAL and APPROVED
Request For Payment REJECTED by ADMINISTRATION	Travel Administration	Reject request	-
Request For Payment REJECTED by BUDGET OWNER	Budget Owner	Send rejection notification	-
Request For Payment REJECTED by EMPLOYEE	Employee	Reject request	-
Request For Payment REJECTED by MISSING	-	-	No Information in Description
Request For Payment REJECTED by PRE_APPROVER	-	-	No Pre Approver in Description
Request For Payment REJECTED by SUPERVISOR	Supervisor	Rejected	-
Request For Payment SAVED by EMPLOYEE	-	-	No Information in Description
Request For Payment SUBMITTED by EMPLOYEE	Employee	Submit request	-
Request Payment	no lane	Request Payment	-
Send Reminder	no lane	Send reminder to employee	-
Start trip	no lane	Trip started	-

H Actions undertaken by the various entities



I Rejections and Approvals in the International Declarations

Declaration REJECTED by ADMINISTRATION	1549.0
Declaration REJECTED by BUDGET OWNER	40.0
Declaration REJECTED by DIRECTOR	4.0
Declaration REJECTED by EMPLOYEE	1780.0
Declaration REJECTED by MISSING	103.0
Declaration REJECTED by PRE_APPROVER	84.0
Declaration REJECTED by SUPERVISOR	126.0
Permit REJECTED by ADMINISTRATION	83.0
Permit REJECTED by BUDGET OWNER	31.0
Permit REJECTED by DIRECTOR	1.0
Permit REJECTED by EMPLOYEE	231.0
Permit REJECTED by MISSING	43.0
Permit REJECTED by PRE_APPROVER	25.0
Permit REJECTED by SUPERVISOR	92.0

(a) Rejections

Declaration APPROVED by ADMINISTRATION	5037.0
Declaration APPROVED by BUDGET OWNER	1834.0
Declaration APPROVED by PRE_APPROVER	612.0
Declaration APPROVED by SUPERVISOR	256.0
Declaration FINAL_APPROVED by DIRECTOR	252.0
Declaration FINAL_APPROVED by SUPERVISOR	6039.0
Permit APPROVED by ADMINISTRATION	4839.0
Permit APPROVED by BUDGET OWNER	1763.0
Permit APPROVED by PRE_APPROVER	534.0
Permit APPROVED by SUPERVISOR	641.0
Permit FINAL_APPROVED by DIRECTOR	640.0
Permit FINAL_APPROVED by SUPERVISOR	5381.0

(b) Approvals

J Activities of the declaration logs whose median duration exceeds one day

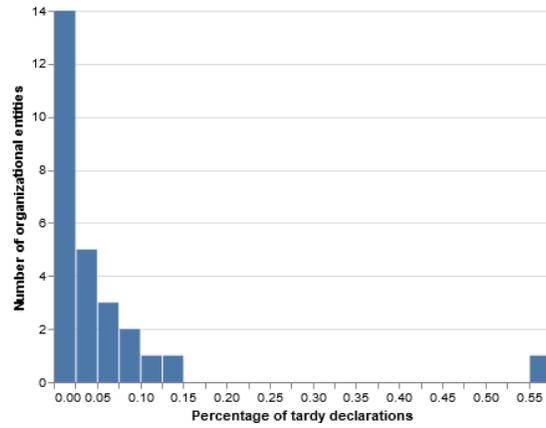
Domestic Declarations

	mean	median	min	max	std
Payment Handled	3 days 11:18:45.580285	3 days 04:53:58	0 days	284 days 06:35:30	3 days 20:46:59.133506
Request Payment	3 days 00:37:28.589142	1 days 02:02:55.500000	0 days	234 days 16:19:13	7 days 01:17:25.589948

International Declarations

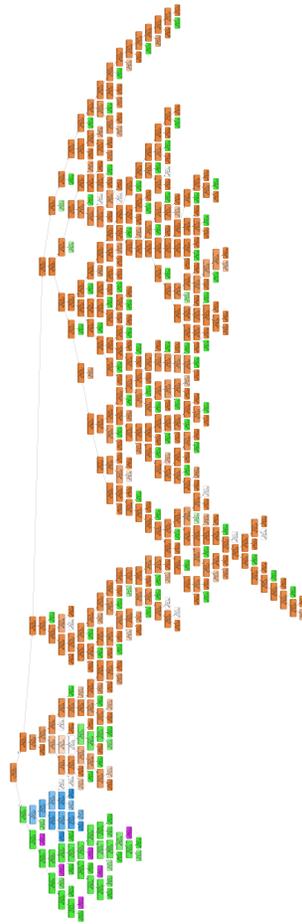
	mean	median	min	max	std
Start trip	42 days 18:09:20.167933	21 days 06:16:51	0 days	454 days 08:43:15	67 days 19:11:44.550750
Declaration SUBMITTED by EMPLOYEE	11 days 01:00:43.141727	3 days 16:33:32	0 days	332 days 00:19:06	25 days 02:23:59.832393
End trip	7 days 13:35:30.122344	4 days 00:00:00	0 days	660 days 06:27:57	25 days 16:42:41.073873
Payment Handled	3 days 09:58:04.998139	3 days 04:17:04	0 days	273 days 08:12:15	4 days 18:45:47.090087
Request Payment	3 days 01:31:09.422391	1 days 02:53:27	0 days	266 days 02:22:37	6 days 02:28:07.318592

K Percentage of tardy declarations per organizational entity



M Full version of the decision tree by role

The full scale images are available for download at <https://bpm.dfki.de/bpi-challenge-2020-notebooks/>.



N Behavioural Profiles for the domestic declarations

		DomesticDeclarations															
Declaration SUBMITTED by EMPLOYEE			→	→					+					→		→	
Declaration FINAL_APPROVED by SUPERVISOR			→	→					+					+		+	↑
Request Payment	←	←	→	→	←		←		←	↑	←	←	←	+	↑	+	↑
Payment Handled	←	←	←	→	←	↑	←		←	↑	←	←	←	+	↑	+	↑
Declaration APPROVED by PRE_APPROVER			→	→					+		→	→	→	+	+	+	+
Declaration REJECTED by MISSING			→	→					+		→	→	→	←	↓	←	+
Declaration REJECTED by PRE_APPROVER			→	→					+	+	→	→	→	+	+	+	+
Declaration REJECTED by EMPLOYEE									+					+		+	→
Declaration SAVED by EMPLOYEE	+	+	→	→	+	+	+	+	+	+	+	+	+	+	+	+	+
Declaration REJECTED by SUPERVISOR			→	→					+					+		+	→
Declaration APPROVED by ADMINISTRATION			→	→	←	←	←		+					+		+	
Declaration APPROVED by BUDGET OWNER			→	→					+					+		+	←
Declaration FOR_APPROVAL by SUPERVISOR	←	+	+	+	+	→	+	+	+	+	+	+	+	+	+	+	+
Declaration REJECTED by ADMINISTRATION			→	→	+	↑	+		+					+		+	→
Declaration FOR_APPROVAL by PRE_APPROVER	←	+	+	+	+	→	+	+	+	+	+	+	+	+	+	+	+
Declaration REJECTED by BUDGET OWNER		→	→	→	+	+	+		+					+		+	→
Declaration FOR_APPROVAL by ADMINISTRATION		↓	↓	↓	+	+	+	↑	+	↑		↓	↓	+	↑	+	↓

|| interleaving relation
 → strict order relation
 + exclusiveness relation

References

1. Adriansyah, A.: Aligning observed and modeled behavior. Ph.D. thesis, Department of Mathematics and Computer Science (2014). <https://doi.org/10.6100/IR770080>
2. Berti, A., van Zelst, S.J., van der Aalst, W.: Process mining for python (pm4py): bridging the gap between process-and data science. arXiv preprint arXiv:1905.06169 (2019)
3. de Leoni, M., van der Aalst, W.M., Dees, M.: A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems* **56**, 235 – 257 (2016). <https://doi.org/https://doi.org/10.1016/j.is.2015.07.003>, <http://www.sciencedirect.com/science/article/pii/S0306437915001313>
4. van Dongen, B.: Bpi challenge 2020. 4tu.researchdata. collection. (2020). <https://doi.org/https://doi.org/10.4121/uuid:52fb97d4-4588-43c9-9d04-3604d4613b51>
5. ISO: Information technology — object management group business process model and notation (2013)
6. Rozinat, A.: Process mining : conformance and extension. Ph.D. thesis, Industrial Engineering & Innovation Sciences (2010). <https://doi.org/10.6100/IR690060>, proefschrift.
7. Weidlich, M., Polyvyanyy, A., Desai, N., Mendling, J., Weske, M.: Process compliance analysis based on behavioural profiles. *Information Systems* **36**(7), 1009 – 1025 (2011). <https://doi.org/https://doi.org/10.1016/j.is.2011.04.002>, special Issue: Advanced Information Systems Engineering (CAiSE'10)