

## THE APPLICATION OF BUSINESS PROCESS MINING TO IMPROVING A PHYSICAL ASSET MANAGEMENT PROCESS: A CASE STUDY

B.T. Greyling<sup>1\*\*</sup> & W. Jooste<sup>1</sup>

### ARTICLE INFO

#### Article details

Submitted by authors 27 Nov 2016  
Accepted for publication 2 Jun 2017  
Available online 31 Aug 2017

#### Contact details

\* Corresponding author  
btg@sun.ac.za

#### Author affiliations

1 Stellenbosch University,  
Department of Industrial  
Engineering, South Africa

# The author was enrolled for an  
MEng (Engineering Management)  
degree in the Department of  
Industrial Engineering,  
Stellenbosch University, South  
Africa

#### DOI

<http://dx.doi.org/10.7166/28-2-1691>

### ABSTRACT

Business process planning and control is important for effectively managing and improving processes relating to the management of physical assets. This is especially true when processes affect the uptime and value creation by physical assets. This article presents a case study where an asset management process is analysed using a technique called 'process mining', with which it is possible to investigate the process as it is being performed in the real world. By applying process mining instead of a traditional mathematical approach, real-world issues can be identified and corrected to improve the effectiveness of the given process. A process model is first constructed to investigate process execution patterns, after which dotted charts are used to identify problem areas within the process and to propose possible areas for improvement.

### OPSOMMING

Besigheidsprosesbeplanning en -beheer is belangrik vir die doeltreffende en die verbetering van prosesse wat verband hou met die bestuur van fisiese bates. Dit is veral waar wanneer prosesse beïnvloed word deur die operasionele tyd en waardeskepping van die fisiese bates. Hierdie artikel bied 'n gevallestudie aan waar 'n batebestuur proses ontleed word met behulp van 'n tegniek genaamd prosesmyn. Met prosesmyn, is dit moontlik om die proses te ondersoek soos dit uitgevoer word in die werklikheid. Deur die toepassing van prosesmyn in plaas van 'n tradisionele wiskundige benadering te volg, kan regte wêreld kwessies geïdentifiseer en reggestel word om die doeltreffendheid van die gegewe prosesse te verbeter. 'n Proses model word eers gebou om die uitvoering van die proses te ondersoek, en daarna word kolkaarte gebruik om probleemareas te identifiseer binne die proses vir moontlike verbetering.

## 1 INTRODUCTION

In modern organisations, physical assets play an increasingly important role. As these physical assets are essential for the value created within organisations, it has become vital for organisations to improve decision-making about physical asset management processes [1]. Furthermore, decision support is vital for improving operations, ultimately leading to increased efficiency and uptime of physical assets, and consequently to increased value creation [2][3].

The information systems used by asset-intensive organisations have the ability to record historical data for activity workflow. This data includes activities and events performed within the organisation that are status-bound. The status-bound characteristic is useful because it allows the start and end times of activities to be derived from the historical records [4]. Information systems such as enterprise asset management systems (EAMS), enterprise resource planning (ERP) systems, and workflow management systems (WFMS), all store the historical data in some form of historical or transaction log, commonly referred to as an event log.

Business process mining, or process mining, is a discipline that aims to use the data from the event log to gain analytical insight into the process. The goal is to apply process mining techniques and tools to improve existing processes. The improvement of processes then becomes routine and is embedded within the existing process improvement activities of the organisation [5]. Process mining is advantageous when compared with traditional process improvement activities, since the processes can be analysed by how they are performed in the real world. Although this is different from traditional methods, which use mathematical optimisation to improve processes theoretically, process mining does not require previously modelled or planned processes. Process mining tools are also able to construct a model just from the event logs captured from real-world activities [4] [6] [7]. Process mining models are primarily constructed by using Petri-nets or business process modelling and notation (BPMN). Petri-nets have a solid mathematical foundation, which makes them ideal for theoretical analysis. While both Petri-nets and BPMN are visual representations, BPMN is visually easier to interpret.

In this paper, process mining is used to analyse an asset management process within the petrochemical industry. A model is constructed to represent the reality of how the process is executed, followed by a dotted chart analysis and a key performance indicator (KPI) analysis. The results of the analysis are presented in a case study, in which the process under investigation is vital for the uptime of physical assets when reactive maintenance has to be performed. The objective is to map the real-world process as recorded by the organisation's EAMS, and compare it with the intended process. Problem areas are identified for improvement.

## 2 PROCESS MINING

Process mining revolves around the idea of first 'discovering' and then constructing a model from event log data [7]. The models built in this way are able to convey the real-world activities belonging to a process. When the organisation uses the event log from a process that already has a planned structure, the real-world process activities can be compared with the desired process structure [4] [8] [9]. In order to perform process mining, the following data attributes are required from the event logs:

- Event identification: These IDs are assigned to the activities within the process. IDs can either be generated as activities occur or have a fixed value for event types;
- Timestamp: Timestamps are assigned at the instant that an event status is triggered;
- Activity: A description of the event;
- Resource: Refers to the person or entity responsible for performing the event;
- Case identification: These IDs are assigned according to the process instance (case) to which the activities belong.

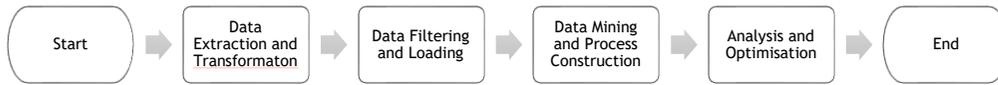
With these data attributes, activities can be modelled from different perspectives. These are [4]:

1. A control-flow perspective, which illustrates the process definitions and activity order;
2. A resource perspective, which gives an organisational view of the process structure and the role players who are involved;
3. A data perspective, illustrating the information generated within the process;
4. A task perspective, which is indicative of the organisational function involved in the process (i.e., the operations); and
5. An operational perspective, illustrating the process application and actions.

The control-flow perspective, in which the processes are holistically analysed, is presented in this paper. Firstly, a control-flow diagram is drawn of the process, illustrating the flow of activities required to finish a process instance. In this case, the asset management process from a petrochemical organisation is illustrated in Figure 2. When constructing the control-flow diagram from a given event log, it is important to note which attributes play a key role. With a control-flow diagram, the 'Case ID' provides a distinguishing field for which processes or cases are associated with an activity.

The process mining procedure starts from the business intelligence (BI) efforts of the organisation. BI is involved in the organisation's approach to capture, integrate, and clean enterprise data for decision-making [10]. The first two steps of the process mining procedure coincide with what is more

commonly referred to as an ‘extracttransform-load’ (ETL) process [11]. Since most organisations do not have proper BI configurations in place to support process mining efforts, the loading in this case is first interrupted by an additional filtering process to make the data suitable for the process mining procedure. The process mining procedure is presented in Figure 1.



**Figure 1: Process mining procedure**

The process mining process starts with data gathering or data extraction. It is well-documented that this activity is not trivial [9]. Most organisations are confronted with transactional data that is scattered across different departments during the implementation of an information system. Ideally, it should be stored in a central location - a data warehouse or data bank [10]. When data is scattered it often leads to difficulties, especially when the process being investigated calls for collaboration between different departments. It is in such cases that data is frequently configured across different data tables, requiring effort to merge data between the data tables in preparation for the process mining analysis.

Data filtering forms part of the procedure. With data being recorded in high volumes - and in some cases across different departments - there is the likelihood of erroneous data from the software, or due to input error from the user. This might also include data that should have been organised under a different process, or missing property contents, that might lead to unreliable results. It can also be argued that it is important to assume that the event log only contains events that have truly occurred [9]. Events that are not supposed to be included in the event log are impossible for process mining algorithms to detect because they are erroneous. ‘Noise’ in this case, therefore, does not exclusively refer to data that should not be included, but also to the ‘outliers’ that refer to infrequent events.

Software programs such as Disco offer filtering capabilities, such as a ‘timeframe filter’ that allows sectioning of process data for comparison before and after a certain point in time [12]. The ‘variation filter’ applies to the outlier concept explained above, where a certain percentage of variation from the norm can be filtered out. Heuristic mining, genetic mining, and fuzzy mining can also be used to filter out noise [9]. Crude errors are usually corrected before loading the data into the analytics software, while more intricate errors or noise can be handled by integrated algorithms.

## 2.1 Review of algorithms

To construct a model from an event log, the extraction and organisation of data first needs to be performed. In process mining, different types of algorithms are used to mine the data according to certain rules, depending on the desired behaviour [13]. Each of these algorithms has its own advantages and disadvantages. Viewed holistically, there are three main categories of data mining algorithms:

- Deterministic algorithms;
- Heuristic algorithms; and
- Genetic algorithms.

With deterministic models, all of the data necessary for the process mining outcome is known. One of the most notable properties of deterministic models is that the process mining output is constant for the given input variables from the event log. The model created during process mining will therefore always be repeatable. It can also be shown that deterministic algorithms are generally faster to execute than other algorithms [14]. An example of a deterministic algorithm is the  $\alpha$ -algorithm [7], which is concerned with the ordering and relational ties of events in an event log.

Heuristic algorithms are commonly used when a predefined algorithmic approach is unable to find an optimal solution. In cases like these, it is necessary to implement an approach that aims to look for a good solution (whether globally optimal or not) by trial and error [15]. Process mining use deterministic algorithms as the basis, supplementing these algorithms with Pareto principles (frequency indicative of importance) to be able to disregard paths or events that only lead to

unnecessary complexity. ‘Unnecessary complexity’ in this instance refers to the complexity that arises from a small percentage of undesired behaviours. As the algorithm attempts to compensate for this behaviour in the model, it adds complexity without its being useful during the analysis. Although this behaviour is excluded from the model, it is not completely forgotten, as it still plays a role once performance metrics are calculated.

Genetic algorithms are premised on the fundamental principles that govern natural selection in evolutionary theory [16]. With genetic algorithms, a solution to a problem is found by starting with an arbitrary starting point, and then searching for a better solution while disregarding inferior solutions. The search is done by combining attributes used for the previous solutions and also introducing random variations. The procedure of genetic algorithms follows a five-step process [17]:

1. Randomly generate solutions and evaluate the fitness of the generated model;
2. Crossover: Generate offspring (new solutions) by using the best solution in the previous step;
3. Mutation: Insert random variations in the solution space;
4. Fitness assignment: Assign a fitness value to solutions based on the outcome;
5. Selection: Select best solutions based on fitness, and use them for the next crossover step until the criterion has been satisfied.

The rationale of the procedure is that process models are generated randomly and are then iteratively reduced to find more satisfactory solutions by means of the mutation and crossover steps. It should be noted that the initial process models that are generated are not a representation of the event log. In cases where it is important to note the different path and the number of occurrences of a particular deviation, a fuzzy miner algorithm can be used to create a weight for the arcs, based on the occurrence of the particular path.

With every process-mining study, it is important to identify the needs of the study. This allows for the proper selection of algorithm, and ensures that the model is not under-fitted. As with all modelling processes, there is a constant trade-off that calls for a balance between the precision, fitness, simplicity, and generalisability of the model [9]. These four modelling criteria can be described as follows:

- Precision: The model does not allow for paths that differ from reality;
- Fitness: The ability to replay exactly what happened in the event log;
- Simplicity: Reduce the complexity of the model by eliminating unwanted paths; and
- Generalisability: The model is not exclusively for a certain event log, and can be used to make generalisations about other processes.

As it is not possible to meet all of these criteria at the same time, the importance of every criterion needs to be assessed before the process mining analysis. The ‘inductive visual miner’ (IvM) was introduced to deal with all the considerations involved in selecting the appropriate algorithms [18]. This tool set is available as a software plug-in for the process mining software ProM, and is able to simplify most of the iterative processes involved with setting different parameters and settings to obtain the desired model. It also has data-filtering capabilities that support the loading and formatting of the data. The IvM allows for the visual presentation of the model, and can animate the behaviour within the event log for the given process mining model.

### 3 CASE STUDY RATIONALE

The case study aims to apply process mining analysis to an existing asset management process in order to explore whether the process deviates from the planned process. It will further seek opportunities for improvement in cases where deviations are found, to illustrate how process mining can be used for process improvement.

The process selected for the process mining case study forms part of a petro-chemical organisation’s maintenance process. The process is embedded in the organisation’s EAMS, where the information system is used to schedule and manage the life cycle of maintenance work and the activity state changes within the process. From here on, ‘state changes’ will refer to the completion of one activity in the process and the progression to the next. The control-flow perspective of the process is illustrated in Figure 2, with the preferred path highlighted in grey.

## 4 ANALYSIS AND RESULTS

The ProM software is used to analyse the event log of the identified process. The data set contains recorded entries from 2 January 2015 to 8 June 2015. Although only six months of data is included, the data still included 198,642 records, which is sufficient to perform the analysis. A CSV file containing the event log data is imported and used to generate a report showing the initial attribute values for the event log. The summary of attributes is shown in Table 1, with the initial values shown in the ‘Before Filtering’ column.

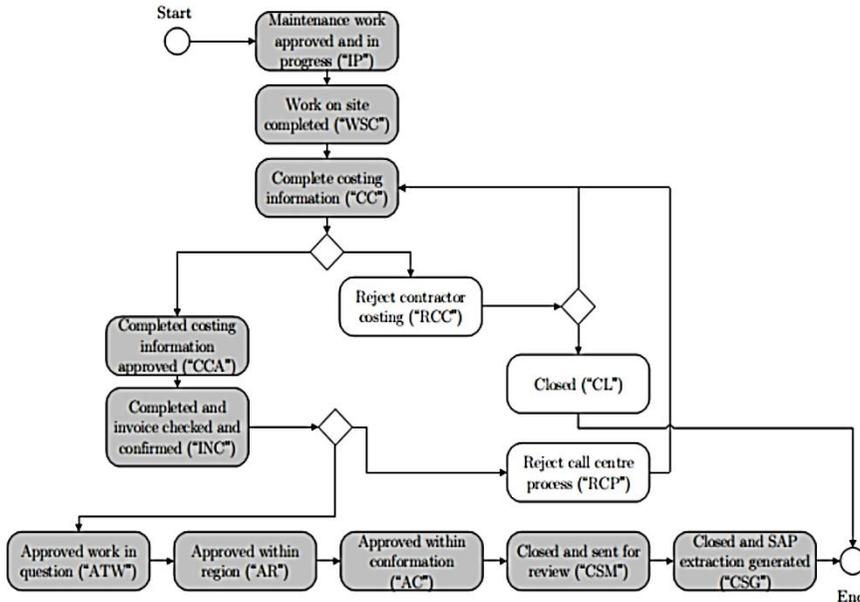


Figure 2: Base maintenance process (planned process)

A ‘log filter’ plug-in for the ProM software is applied to filter the event log before analysis. Firstly, a heuristic filter is used to filter out perceived noise. This is followed by a second filtering to remove all process activities that fall outside the selection time frame. The resulting event log attributes after filtering can be seen in Table 1, in the ‘After Filtering’ column.

Table 1: Summary of event log attributes

Attribute	Before Filtering	After Filtering
Processes	1	1
Cases	17414	15809
Events	216026	203309
Event Classes	27	26
Events Per Case	Min: 3, Mean: 12, Max: 51	Min: 4, Mean: 13, Max: 45
Event Classes per Case	Min: 2, Mean: 11, Max: 16	Min: 3, Mean: 11, Max: 16
Log Start Date	Fri Jan 02 6:05 2015	Fri Jan 02 6:05 2015
Log End Date	Mon Jun 08 11:35 2015	Mon Jun 08 11:03 2015

### 4.1 Process discovery and understanding

The next step in analysing the event log entails the construction of a control-flow model based on the given event log. The filtered data is first imported into the ProM software for analysis with the IvM tool set.

When using the IvM, the conformance settings are set for the model. Conformance is the measure that dictates the percentage by which a given event log matches a process model. Conformance can be deconstructed into two sub-metrics called ‘fitness’ and ‘appropriateness’. The difference between these two metrics should be noted, as they are integral to how process models are built. Firstly, ‘fitness’ relates to the mismatch between the constructed model and the given event log. It is calculated by considering all the process mismatches and deviations [19]. ‘Appropriateness’ is important when a model needs to be as concise and simple as possible to explain the behaviour

within the event log. It is important, therefore - especially in complex situations - that these two metrics are in balance. This will ensure that the model conforms sufficiently, while also being simple enough to explain the deviation in behaviour.

Two settings are available for adjusting the conformance: 'path filtering' and 'activity filtering'. The percentage of filtering can be set for each. Applying the software's default settings of 80 per cent path filtering and 100 per cent activity filtering for the first process mining iteration, the model in Figure 3 is discovered. The 80 per cent path filtering essentially applies the Pareto principle where 20 per cent of complexity is regarded as noise, and thus as not useful. One of the first observations that can be made is the high complexity at the start of the process, which is not indicative of the planned process illustrated in Figure 2.

In principle it is possible to find the original planned process when filtering settings are adjusted downward. This is assuming that the planned process was adhered to adequately for there to be a high enough process trace. The next step in the analysis is, firstly, to reduce the activity filter until only the original planned activities are shown. This point is reached at an activity filter value of 42 per cent. Secondly, the path filter setting is reduced until the planned process is isolated. This is achieved at 47 per cent. The mined process with the adjusted filters is shown in Figure 4.

Based on the results, the ProM software is able to extract metrics associated with discovered models by using a 'conformance checker' plug-in. These metrics are calculated based on the event log on which the model is based. The calculated values are shown in Table 2.

**Table 2 : Conformance test results**

Process model	IvM setting				Calculated values	
	Flexibility	Activities	Path	Fitness [f]	Precision [aaB]	Structural [aaS]
Figure 3	0.44	100%	80%	0.8348	0.7473	0.919
Figure 4	0	42%	47%	0.8356	1	1

At this point of the analysis, a summary of commonly occurring process traces can be extracted. The summary of traces shown in Table 3 is based on the original event history in the event log, and thus is not influenced by the path and activity filtering applied (Figure 4).

**Table 3: Discovered log traces**

Ranking	# of Instances	Log Trace (corresponding to the activity codes in Figure 2)
1	1822 (11.53% of Log)	AAA;IP;WSC;CC;CCA;INC;ATW;AR;AC; CSM;CSG
2	1289 (8.15% of Log)	AAA;IP;WSC;RCC;CC;CCA;INC;ATW;AR; AC;CSM;CSG
3	756 (4.78% of Log)	AAA;IP;WSC;CC;CCA;INC;ATW;AC;AR; CSM;CSG
4	440 (2.78% of Log)	AAA;IP;WSC;CL
5	415 (2.63% of Log)	AAA;IP;WSC;CC;RCC;CC;CCA;INC;ATW; AC;AR;CSM;CSG
6	385 (2.44% of Log)	AAA;IP;WSC;CC;RCC;CC;CCA;INC;ATW; AR;AC;CSM;RCP;CSM;CSG
7	364 (2.30% of Log)	PAA;IP;WSC;CC;RCC;CC;CCA;INC;AR; ATW;AC;CSM;CSG
8	363 (2.30% of Log)	WSC;CL;RWO;WSC;CC;CCA;RCC;CC;CCA; INC;ATW;AR;AC;CSM;RCP;CSM;CSG
9	334 (2.11% of Log)	AAA;IP;WSC;CC;CCA;RCC;CC;CCA;INC; ATW;AR;AC;CSM;RCP;CSM;CSG
10	323 (2.04% of Log)	PAA;IP;WSC;CC;RCC;CC;CCA;INC;ATW; AR;AC;CSM;CSG

Table 3 shows that the planned process trace only contributes to a total of 11.53 per cent of the entire event log. The next trace, contributing 8.15 per cent, only includes the 'RCC' activity, which involves rejecting the costing submitted by the contractor. The rest of the trace is identical to the planned process. Regarding the traces ranked third, fifth, and seventh, it is clear that the order of activities 'AR', 'ATW', and 'AC' are inconsistent. The inconsistency of these traces contributes 9.71 per cent towards the entirety of the log.

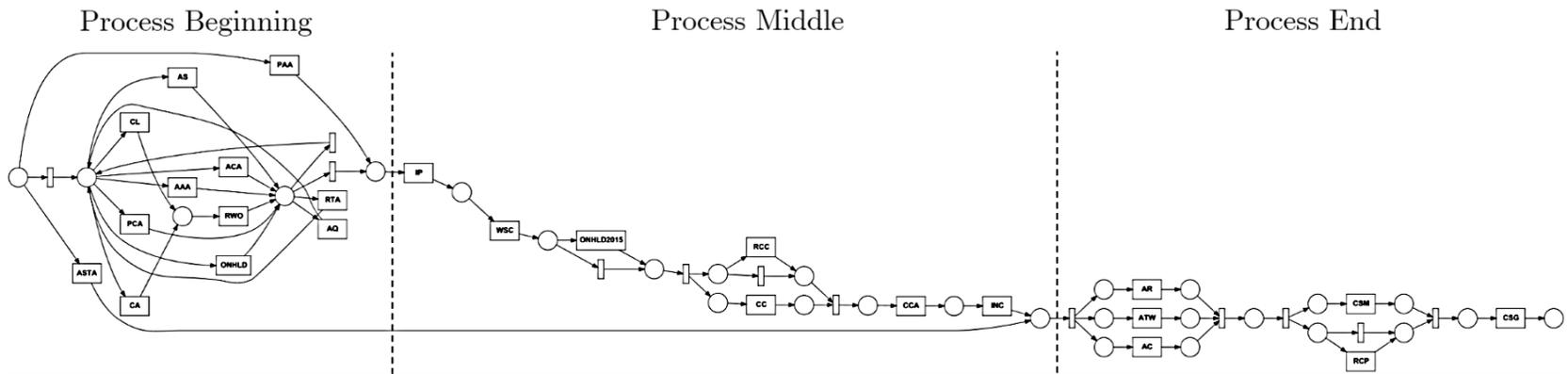


Figure 3: Mined process model with 100% activities and 80% paths



Figure 4 : Mined process model with 42% activities and 47% paths

The mined model can be further examined to gain insight into the real-world process flow. The initial model is divided into three sections (Figure 3). The analysis continues by starting at the end and progressing to the start of the model, as this is the order in which complexity increases. Inconsistencies are demonstrated in the mined process model where the model presents parallel activities compared with where activities should be sequential according to the planned process. The second split from the end in the Petri-net flow shows the parallel activities 'CSM' and 'RCP' and the by-pass of these activities. The Petri-net has a visualisation error in this case. Closer inspection of Table 3 reveals that when 'RCP' occurs, it falls into a loop with 'CSM', which always precedes 'CSG'. The process mining algorithm limits itself to a single representation per activity, which ultimately causes it not to show 'CSM' as a follow-up event. It is important to reduce the complexity of the mined model to remain useful and offer useful results. In this case, it can be derived that the end of the process suffers from non-conformance in the sequence of the 'AR', 'ATW', and 'AC' activities, resulting in the difficulty of completing 'RCP'.

An important conformance issue can further be seen regarding the 'RCP' activity. Referring back to the original planned process in Figure 2, this activity forms part of a joint-split from activity 'INC' through to 'AC'. The mined model (Figure 3) shows that 'RCP' mainly occurs after the 'AR', 'ATW', and 'AC' segment, and not after 'INC'. Investigation into this deviation should be considered to improve the process.

Another loop in the process model involves the activities 'CC' and 'RCC' in the process middle. While the 'CC' activity is the preferred activity according to the planned process, the 'RCC' activity forms part of the secondary parallel process. The 'ONHLD2015' activity is a placeholder for cases where the process was stopped during 2014 and planned to continue in 2015. This shows that a majority of the cases are interrupted after the 'WSC' activity during the transition to 2015. The result is that these process interruptions cause delays, and need to be reinitialised.

By considering the process beginning, it is clear that the majority of process instances start with the 'AAA' activity. The model illustrates that there are inconsistencies when considering lower frequency process traces. Process loops are also present early in the processes life cycle, indicating process uncertainty. This process uncertainty subsides once the 'IP' activity is reached. As some processes are cancelled or put on hold, process complexity increases, as shown by the presence of the 'CL', 'CA', and 'RWO' activities. The majority of waiting activities occur at the beginning of the process. This is mostly attributed to external contractor costing approval. It can be concluded that delays are largely due to external environmental factors. Improving relations with these external environmental factors will potentially improve process performance.

A summary of the activity occurrences is compiled to understand the behaviour of the process model and the real-world activities. A summary of all activity occurrences is shown in Table 4. It is evident that the top 11 ranked activities all form part of the planned process (Figure 2). From there onwards, process deviations occur with relatively low percentages. The most frequently occurring activities ('RCC' and 'RCP') both involve the rejection of a condition attributed to the maintenance process from external contractors over which control is limited. 'RCC' involves the rejection of costing provided by the contractor, while 'RCP' involves the rejection of a process from a call centre. Table 4 further shows that 2,818 cases were closed (activity 'CL'). It is shown (refer to 'Occurrence as end event') that 879 process instances were terminated by activity 'CL', while the remaining instances were re-opened by activity 'RWO'. It should further be noted that activity 'WSC' is shown occasionally to be a start event. This is due to projects being approved before 2015. These cases were not removed from the data to ensure that valuable information embedded in those instances is not lost.

In summary, this section reports on how process mining assists in exploring and discovering compliance and deviation within a process. The results are obtained from using only an information system's event log data, without the need for extensive insight into the actual real-world process.

**Table 4: Log summary of activity occurrence**

Rank	Event name	Total occurrences		Occurrence as start event		Occurrence as end event	
		Absolute	Relative	Absolute	Relative	Absolute	Relative
1	CC	25015	12.30%	-	-	-	-
2	CSM	18948	9.32%	-	-	-	-
3	CCA	16802	8.26%	-	-	-	-
4	WSC	16745	8.24%	2381	15.06%	-	-
5	ATW	16301	8.02%	-	-	-	-
6	AR	16049	7.89%	-	-	-	-
7	AC	16030	7.89%	-	-	-	-
8	INC	15408	7.58%	-	-	-	-
9	CSG	15053	7.40%	-	-	14930	94.44%
10	IP	14136	6.95%	1326	8.39%	-	-
11	AAA	10143	4.99%	9586	60.64%	-	-
12	RCC	10023	4.93%	-	-	-	-
13	RCP	4284	2.11%	-	-	-	-
14	CL	2818	1.39%	-	-	879	5.56%
15	PAA	2520	1.24%	2516	15.92%	-	-
16	RWO	1920	0.94%	-	-	-	-
17	ASTA	501	0.25%	-	-	-	-
18	AS	246	0.12%	-	-	-	-
19	PCA	218	0.11%	-	-	-	-
20	ACA	72	0.04%	-	-	-	-
21	CA	52	0.03%	-	-	-	-
22	ONHLD	11	0.01%	-	-	-	-
23	ONHLD2015	5	0.00%	-	-	-	-
24	AQ	4	0.00%	-	-	-	-
25	RTA	4	0.00%	-	-	-	-
26	RWC	1	0.00%	-	-	-	-

#### 4.2 Dotted chart analysis

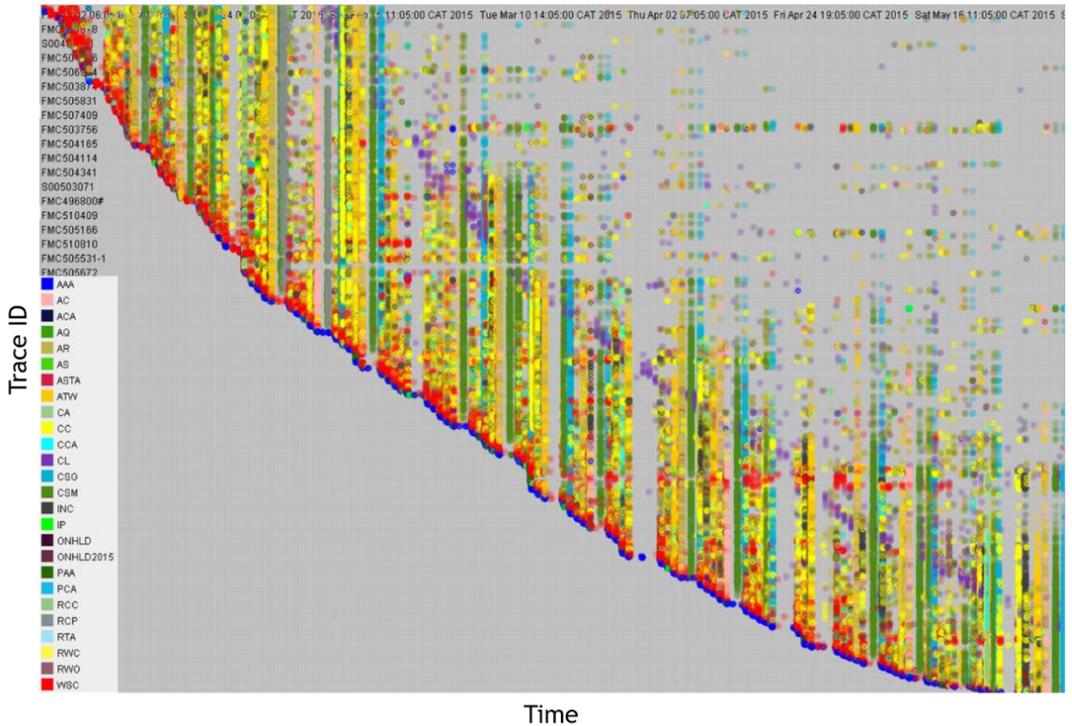
To interpret visually the behaviour in the event log, a dotted chart analysis is performed. The dotted chart displays the process trace along the y-axis and time along the x-axis. Every event is displayed as a dot on the chart coinciding with the process instance and the time at which it occurred. This allows for the interpretation that is not possible when only observing the event log entries.

Figure 5 shows the dotted chart, with the process instances sorted according to the date and time of the first event from the top. The time frame for this dotted chart ranges from 2 January to 8 June 2015. Events in the dotted chart are colour-coded, and a legend is provided. It can be seen in Figure 5 that there is a large concentration of events that occurred at the beginning of January. The rate of occurring processes declined as the year progressed towards June.

To illustrate the duration of the process instances, the cases are reordered according to the y-axis. The process instances remain relative to each other while the order is changed according to the instance's total duration. The overview of durations is shown in Figure 6, where the instances with the shortest durations are shown at the top while the instances with the longest duration are shown at the bottom. Given that the red dots represent 'WSC' and the yellow dots represent 'CC', it can be seen that the main cause of long process durations is the working time of 'WSC'.

#### 4.3 Key performance indicator (KPI) analysis

KPIs relating to activity duration and importance can be calculated to monitor and assist with improving the process. The log replay plug-in in the ProM software is used to calculate the KPIs relating to the activity durations. The event log consists of record events from January to the end of May 2015. It is therefore possible to analyse the data for each month. The event log is divided into separate monthly event logs and analysed separately.

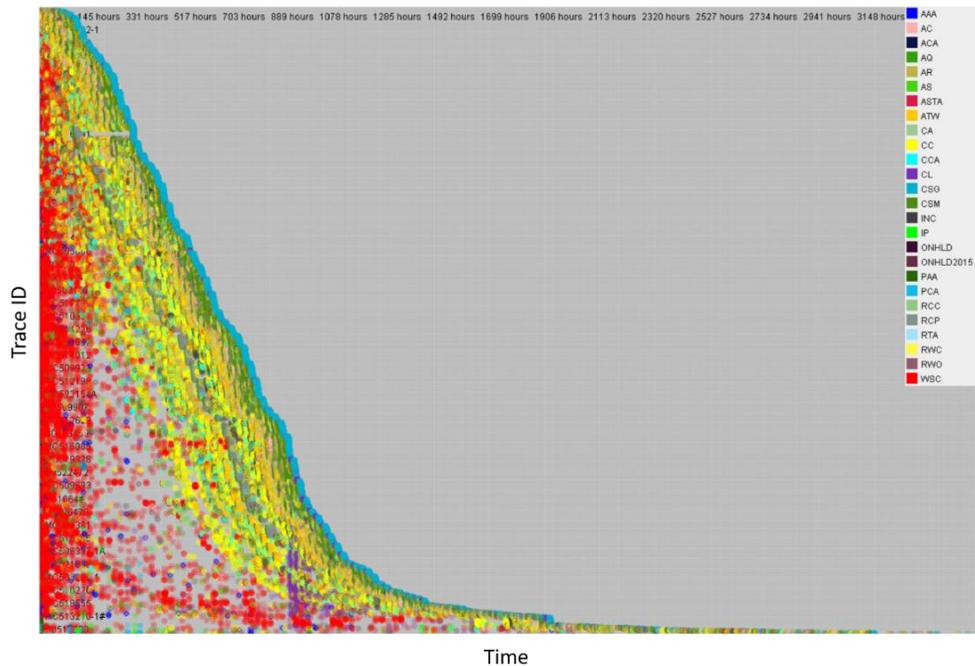


**Figure 5 : Dotted chart analysis (see online version for colour image)**

The results are presented in Figure 7, where it can be seen that process duration is the longest in January, with the longest time duration time, average, and spread about the median. The median remains constant throughout the period of January to May, while the maximum duration follows a downward trend. With the medians presented as diamonds within the box plots, it is seen that median activity durations become shorter over time.

Following the evaluated traces, analysis is conducted on the individual activities within the traces. This gives insight into what is happening within the processes and the factors influencing activity durations. The ProM software basic performance metric analysis is used to calculate the average durations of the activities. Figure 8a shows a summary of the median durations of the different activities.

From Figure 8a, it is seen that 'WSC', 'RTA', 'IP', 'CL', and 'ACHLD' are the five activities with the longest durations. Before activities for improvement can be identified, the influence of external environmental factors on activities needs to be considered. The criticality or importance of an activity also needs to be considered. Due to the effort and organisational resources required to improve activity durations, it is important to assess which activity duration improvement will have the highest impact. This criticality rating is determined by combining the duration of the activities with the frequency of occurrence.



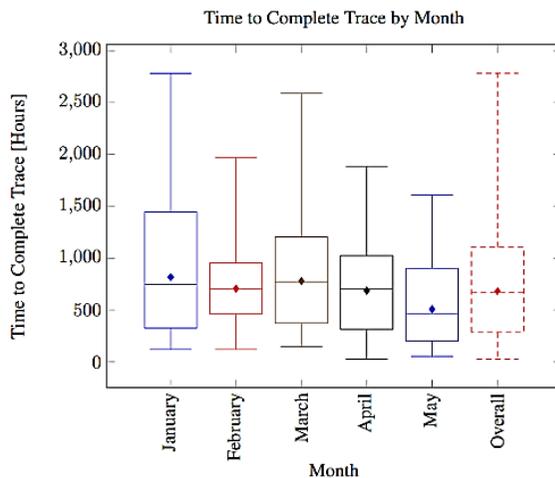
**Figure 6: Dotted chart analysis, with traces sorted by length (see online version for colour image)**

The activity criticalities are normalised to determine the most critical activities relating to the other activities. Equation 1 shows how the values are normalised, and Equation 2 how the criticality values are calculated. In these equations,  $x_i$  is the activity's average duration,  $\min(x)$  is the minimum time within the subset of activity durations, and  $\max(x)$  is the maximum time within the same subset.

$$\text{Normalisation} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

$$\text{Activity Criticality} = \text{Normalised Avg Working Time} \times \text{Normalised Frequency} \quad (2)$$

The normalised activity criticality values are shown in Figure 8b, where it is seen that the activities that offer the largest incentive for improvement are 'WSC' and 'IP'. The main consideration here is that the priority of improving the duration of an activity should be justified by its frequency of occurrence. For example, making the effort to improve activities 'RTA' or 'ACHLD' should be carefully considered, since they occur infrequently in the overall process.



**Figure 7: Monthly trace times box-and-whisker plot**

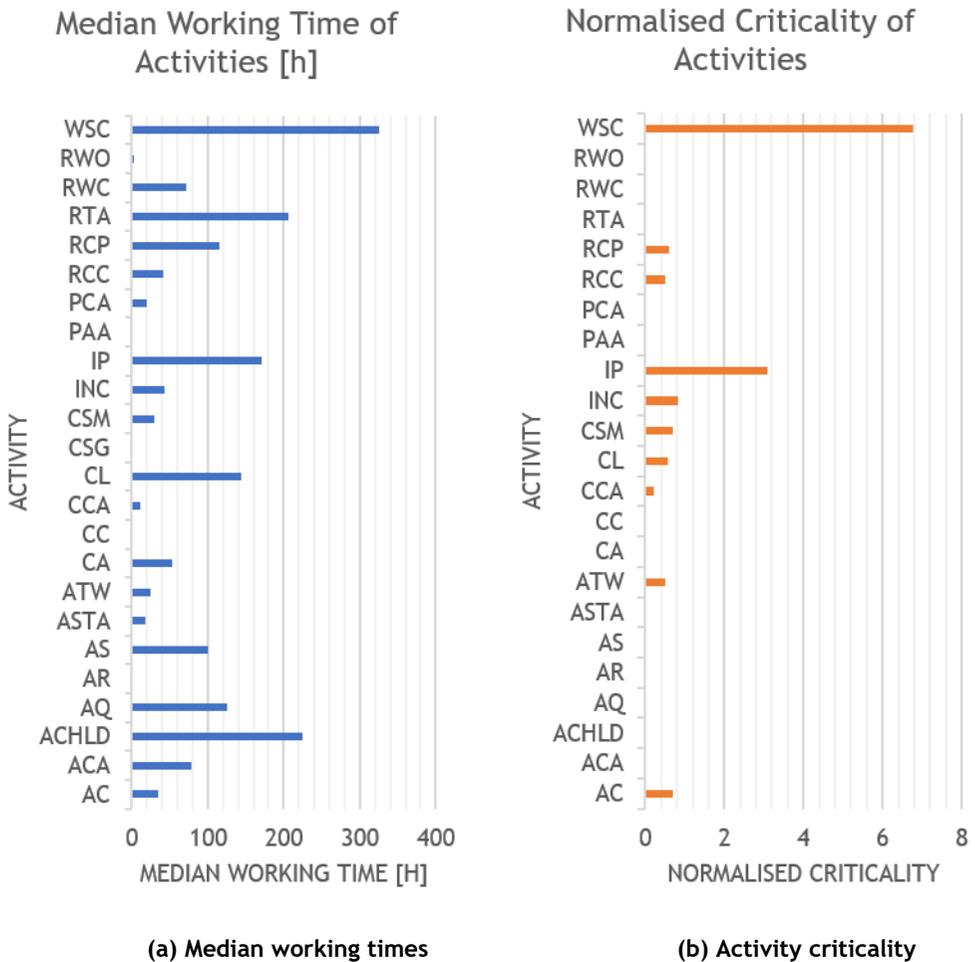


Figure 8: Activity KPI analysis (see online version for colour image)

## 5 DISCUSSION OF RESULTS

In this paper, the analysis of an asset management process is conducted by performing process mining. The analysis illustrates the explorative ability of process mining to gain a better understanding of a planned real-world process. Firstly, the data is gathered by performing data mining of the organisation's existing EAMS. This data is used to build a model aimed at explaining the behaviour of the real-world process. Deviations from the planned process are presented that indicate certain non-conformances in the process.

The individual processes are extracted and compared with the planned process. Through the analysis it is discovered that only 11.53 per cent of the performed process conformed to the planned process. To visualise the extracted event log data, dotted charts are presented that allow the analyst to discover trends within the processes - something that is not possible through mere observation of the event log or EAMS. From the dotted chart, variability in process duration becomes apparent: the analysis reveals that short processes are a result of the cancellation of activities, while long processes show that some activities get stuck in a loop.

KPIs are calculated to clarify the behaviour within the processes. Monthly box-and-whisker plots of the event log data are presented. These plots show a decrease in activity durations from January to May 2015, while individual activity measurements reveal how long activity durations are. Activity criticality is measured in support of the activity durations, to determine the activities on which improvement initiatives should focus.

## 6 CONCLUSIONS

Organisations rely on information systems to manage business processes in today's technology-driven industry. Thus techniques for ensuring that business process management is effective are important to ensure competitiveness. This paper presents a case study showing how an information system's event log can be used, together with the process mining technique, to model, analyse, and improve processes. The case study illustrates, firstly, how process deviations can be identified and used to gain insight into why deviations might occur. It also shows the time domain within which processes are executed, and how activities and deviations influence the duration of these processes. Lastly, the case study illustrates how suggestions can be made to improve process activities based on their impact on improving the overall process. Based on easily obtainable results from applying process mining to an information system's event log, as illustrated in this paper, it is concluded that process mining is a useful technique that should be part of any engineer's improvement toolkit.

## REFERENCES

- [1] Woodhouse, J. 2014. Briefing: Standards in asset management: PAS 55 to ISO 55000. *Infrastructure Asset Management*, 1, 57-59.
- [2] Amadi-Echendu, JoeE., Willett, R., Brown, K., Hope, T., Lee, J., Mathew, J., and Yang, B.-S. 2010. What is engineering asset management? In *Definitions, concepts and scope of engineering asset management*, Springer Heidelberg Dordrecht London New York, 2010, 3-16.
- [3] Wittwer, E., Bittner, J. & Switzer, A. 2002. The fourth national transportation asset management workshop. *International Journal of Transport Management*, 1, 87-99.
- [4] Van der Aalst, W.M. 2004. Business process management demystified: A tutorial on models, systems and standards for workflow management. In *Lectures on concurrency and Petri nets*, Springer Heidelberg Dordrecht London New York, 2004. (pp. 1-65).
- [5] Van der Aalst, W.M., Reijers, H.A., Weijters, A.J., van Dongen, B.F., De Medeiros, A.K., Song, M. & Verbeek, H.M. 2007. Business process mining: An industrial application. *Information Systems*, 32, 713-732.
- [6] Van der Aalst, W.M. 1998. The application of Petri nets to workflow management. *Journal of Circuits, Systems, and Computers*, 8, 21-66.
- [7] Van der Aalst, W.M. & van Dongen, B.F. 2002. Discovering workflow performance models from timed logs. In *Engineering and Deployment of Cooperative Information Systems* (pp. 45-63). Springer Heidelberg Dordrecht London New York, 2002.
- [8] Becker, J., Delfmann, P., Eggert, M. & Schwittay, S. 2012. Generalizability and applicability of model based business process compliance checking approaches: A state of the art analysis and research roadmap. *BuR-Business Research*, 5, 221-247.
- [9] Van der Aalst, W. 2011. *Process mining: Discovery, conformance and enhancement of business processes*, Springer Heidelberg Dordrecht London New York. ISBN 978-3-642-19344-6. (2011)
- [10] Dayal, U., Castellanos, M., Simitzis, A. & Wilkinson, K. 2009. Data integration flows for business intelligence. *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, 1-11.
- [11] Vassiliadis, P. & Simitzis, A. 2009. Near real time ETL. In *New Trends in Data Warehousing and Data Analysis*, 1-31. Gliwice Poland, Springer, LLC, 2009.
- [12] Günther, C.W. & Rozinat, A. 2012. Disco: Discover your processes. *BPM (Demos)*, 940, 40-44.
- [13] Tiwari, A., Turner, C.J. & Majeed, B. 2008. A review of business process mining: State-of-the-art and future trends. *Business Process Management Journal*, 14, 5-22.
- [14] Colaço, M.J. & Dulikravich, G.S. 2009. A survey of basic deterministic, heuristic and hybrid methods for single objective optimization and response surface generation. *Thermal Measurements and Inverse Techniques*, 360-372.
- [15] Winston, W.L. & Goldberg, J.B. 2004. *Operations research: applications and algorithms* (Vol. 3). Boston: Duxbury Press.
- [16] Holland, J.H. 1992. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, Cambridge, MA. 1992..
- [17] Konak, A., Coit, D.W. & Smith, A.E. 2006. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 91, 992-1007.
- [18] Leemans, S.J., Fahland, D. & Van der Aalst, W.M. 2014. Process and deviation exploration with inductive visual miner. In: Proceedings of the BPM Demo Sessions 2014. Volume 1295 of CEUR Workshop Proceedings., CEUR-WS.org.
- [19] Rozinat, A. & Van der Aalst, W.M. 2006. Conformance testing: Measuring the fit and appropriateness of event logs and process models. *Business Process Management Workshops*. In C. Bussler et al., editor, Business Process Management 2005 Workshops, volume 3812 of Lecture Notes in Computer Science, pages 163-176. Springer-Verlag, Berlin, 2006.