Trace Clustering for an Online Freight Exchange at TIMOCOM

Jens Brunk[1], Caroline Methner[1], Ann-Kristin Cordes[1], Sven Thiermann[2], and Tobias Goerke[3]

[1] University of Münster - ERCIS, Münster, Germany
{jens.brunk,ann-kristin.cordes}@ercis.uni-muenster.de
carolinemethner@gmx.de
[2] TIMOCOM GmbH, Erkrath, Germany sthiermann@timocom.com
[3] viadee, Münster, Germany Tobias.Goerke@viadee.de

Abstract. To guarantee a high level of customer satisfaction, providers of highly frequented platforms strive to continuously improve their application design. One promising potential to discover issues regarding the application design and at the same time consider the heterogeneity of different users lies in the clustering of traces and the examination of the corresponding process models. In order to leverage this potential, this study investigates how large-scale event logs can be preprocessed and clustered in order to enable cluster-specific application design by examining an event log of an online freight exchange. Using the Action Design Research approach, which allows a joint shaping of a solution with practitioners, a pipeline is developed, which enables the preprocessing of large-scale event logs and the execution of three trace clustering approaches. The investigation of the resulting clusters in the form of process models delivers valuable insights into customer behavior, helps to identify weaknesses and improvement possibilities of the platform and thereby provides a basis to support application design.

Keywords: Trace Clustering · Application Design · Freight Exchange.

1 Motivation

When using web applications customers sometimes experience undesirable surprises i.e. they feel lost or fail in finding needed information. In order to avoid that customers are dissatisfied with the service of the provider, it is of great importance for companies to adapt their application design to user needs [7]. When interacting with (web) applications, users leave traces in the form of event logs. Especially in the case of highly frequented applications, the collected amount of data can quickly become tremendously large and difficult to manage [1]. Nevertheless, it is worthwhile to face the challenge of handling large amounts of data because of its great potential to be used for process mining [4, 10]. Process mining enables to analyze users' behavior on web applications by generating process models of event logs [11]. The investigation of the models helps to derive insights of the user behavior and detect weaknesses and improvement possibilities of the platform, which can be regarded as a basis for adapting the application design.

Users do not behave homogeneously on applications and thus also choose different features and encounter different problems during use [5]. For this reason, an undifferentiated analysis of the entire event data may not deliver insights into user-specific behavior. Besides, when analyzing the behavior of all users at the same time with the help of process models, these models may become very large and difficult to comprehend [3]. Through the accurate grouping of users, the process models of the usage behavior become less complex and it becomes easier to examine specific customer groups [3, 12]. Companies can benefit from this by identifying improvement possibilities and weaknesses of the platform for specific customer groups. On this basis, they can adapt their application design to customer needs and thereby improve customer satisfaction.

2 Situation Faced

A company that faces the above-described situation is TIMOCOM. It offers Europe's biggest online freight exchange platform. With the help of their platform, TIMOCOM mediates the supply and demand of transport services via an online marketplace. The core business of the company is that customers can publish and search freight and cargo hold offers on the platform. Each day up to 750,000 international freight and cargo hold offers are published online. For this reason, the online freight exchange is a highly frequented platform. TIMOCOM has about 130,000 users coming from 43,000 different companies[4].

The TIMOCOM offers a "Smart Logistic System" including the smart app Freight Exchange, where customers can offer freight. Therefore, they must provide information in the freight input mask about the freight type, size, weight, point of departure and destination. In addition to offering freight, customers can also search for freight offered by other customers. To do this, they must enter their freight search criteria and then they will receive a list of all available freight offers. If they want to contact a provider, they can either use the TIMOCOM instant messenger or retrieve other contact information such as e-mail and telephone. The vehicle exchange works in a similar manner to the freight exchange. Customers can offer their vehicle by providing information such as the vehicle type, volume, date or point of departure and destination.

The performed activities of a user are all recorded in the form of event logs.

TIMOCOM professionals are convinced that the usage behavior on the platform differs between the customers. They assume, for example, that some customers use certain functionalities more intensively than others, choose different ways to navigate on the platform, and differ regarding their degree of activity and their speed of executing activities. Accordingly, user behavior that is usual and desirable for one user group cannot be evaluated equally for another user group. In addition, a user group might exhibit unwanted behavior, which would remain

---

[4] https://www.timocom.de/Unternehmen

hidden when not analyzing this group individually. Therefore, the TIMOCOM's goal is to extend the learned and experienced understanding of their customers by a data based analysis of their application's event logs.

3 Actions Taken

3.1 Research Method

TIMOCOM faced the challenge how to process and cluster large-scale event logs in order to enable customer group specific application analysis and design. Answering this challenge was done in two parts. First, we developed a pipeline to preprocess large-scale event logs and cluster users by means of trace clustering approaches. We tested this pipeline using the event log of the highly frequented TIMOCOM platform. Second, we analyzed the resulting clusters to verify whether they provide a basis to support application design.

For the development of the pipeline and the investigation of TIMOCOM's trace clustering results, we followed the Action Design Research (ADR) approach to ensure that the solution meets the practitioners requirements [8]. Following the ADR approach, several feedback iterations with pratictioners of TIMOCOM and end-users of the platform took place in order to develop the artifact.

For the implementation of the pipeline, we needed to identify appropriate trace clustering methods. As TIMOCOM wants to explore the most suitable approach for their analysis, the pipeline should cover as many different trace clustering methods as possible. In the related literature, various different methods are applied for clustering traces. Based on the research goal and the situation faced by TIMOCOM, we consider only methods that meet the following requirements. (1) The trace clustering method should either be already implemented and publicly available or it should be manageable to implement the method as part of this study. (2) The selected methods must be able to handle very large event logs. (3) The trace clustering analysis should be performed to reduce the complexity and increase the understandability of the process models.

To select suitable trace clustering methods, we used the literature review of [9] as a starting point. It provides an overview of trace clustering methods published until 2015. All methods that meet the previously defined requirements are initially collected. To consider trace clustering methods after 2015, we carried out an additional literature search for the time interval from 2015 until today.

3.2 Implementation

We checked the trace clustering methods listed in the literature review of [9] for the three previously defined requirements and unsuitable methods were rejected. Of the 20 methods presented in the literature review, eight methods then remained. Based on our own literature research, we identified six additional methods fitting our requirements. Finally, we examined all collected methods to determine whether very similar or even identical methods are applied. After removing these, eight methods remained. The literature on trace clustering distinguishes between three trace clustering approaches: trace-sequence-based, feature-vector-based and model-based clustering. In order to cover the entire range, we decided to implement at least one method from each of the three approaches. Respectively, we selected the methods from which we expected the best results.

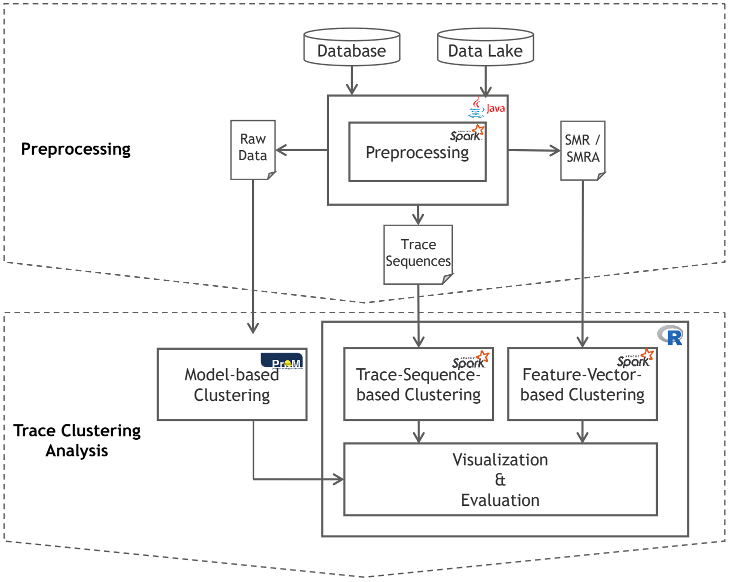
The first selected method is a trace-sequence-based clustering approach, which calculates the similarity of traces based on Levenshtein distances [2]. Since both partitioning and hierarchical clustering algorithms are used in the literature, the approach is tested in this work with both techniques. Second, we implemented the feature-vector-based clustering approach proposed by [3]. Since they determined in their article that the feature sets Super Maximal Repeat (SMR) and Super Maximal Repeat Alphabet (SMRA) feature set perform best in terms of process model complexity, we implemented these two methods in the scope of this study. Furthermore with respect to this approach, we decided to realize a partitioning and a hierarchical clustering algorithm, since both methods are considered in the literature. The last approach is the model-based clustering approach Active Trace Clustering (ActiTraC) presented by [12]. Thus, we selected seven different trace clustering methods, which represent the three types of approaches, for implementation as part of the pipeline.

As already mentioned, the architecture of the envisioned pipeline consists of two components. The first component is dedicated to the general preprocessing of the data. The data can be extracted either from a data lake or from a database. During preprocessing, three new data sets are created based on this data. First, the data is structured, transformed into the required data format and stored on disk (data set: raw data). Then, for each account, the corresponding traces represented as a sequence of characters, where each character stands for a specific activity and each trace is assigned to one user, are created and archived (data set: trace sequences). Finally, the feature sets SMR and the SMRA are generated and stored, which are required for the feature-vector-based clustering approach (data sets: SMR and SMRA). After preprocessing, the trace clustering analysis - the second component - follows based on the previously identified methods. To enable the investigation of the tremendously large event logs, we apply Apache Spark, where possible. Additionaly, some methods were already available as implementations for the Process Mining Workbench (ProM). The entire architecture of the implementation is depicted in Figure 1.

4 Results

As part of the prototypical application within the ADR cycle, we selected one day of the event log as the time interval for the analysis. Since different users are active on the platform at different times of the day, a smaller time interval would not be sufficient to represent all existing user groups. It is assumed that larger amounts of data are not necessary, since users behave similarly on different days. This assumption is not statistically proven in the scope of this study, but it is based on statements of TIMOCOM experts. One day included up to 20

Fig. 1. Pipeline Architecture



million events, which still poses a significant amount of data. For this reason, the choice of one day was a compromise between a sufficiently large amount of data representing users as good as possible and a manageable small amount of data allowing to perform an efficient data and process map analysis with many prototypical iterations.

4.1 Trace Cluster Analysis

We performed each of the seven trace clustering methods on the preprocessed TIMOCOM event log for the number of clusters $k = 2$ to $k = 10$. Three objectives are considered when selecting an appropriate clustering result, i.e. the right number of clusters. Firstly, the process models of the resulting clusters should have as little complexity as possible. Secondly, the observations of the resulting clusters should be as homogeneous as possible and thirdly the clusters between each other as heterogeneous as possible. These two objectives can be evaluated with the help of different trace cluster validity indices (CVI) [6]. First, as a difference-like CVI for the hierarchical cluster analysis, a dendrogram is considered and for the clustering analysis with k-means the With-In-Sum-Of-Squares-plot (WSS) is examined. Second, as an optimization-like CVI, the average silhouette index is investigated. The three process model complexity indices and the cluster stability index support the choice of the number of clusters for all three trace clustering approaches. In the end, the different indices provided ambiguous results. With the help of the indices, however, several candidate number of clusters could be determined for each method. Through a more detailed examination of the process models of the results of the candidate number of clusters, we then determined which cluster results were best suited for defining differentiable and characteristic TIMOCOM user groups. As a final result of this process, the partitioning trace-sequence based clustering method with five clusters was chosen as the best result.

4.2 Evaluation of Clusters

An overview of the resulting five clusters and the distribution of the activities (denoted as anonymized activities a to t) is shown in Fig. 2. Together with a business analyst and two members of TIMOCOM's Applied Technology Research (ATR) team, we matched each of the clusters to a specific user type.

The users grouped in the first cluster represent the *occasional users*. Their activity is considerably lower compared to the users of all other clusters. The reason for this is that they are only active on the platform for relatively short periods per day and therefore carry out comparatively few activities. This may be the case for small forwarding companies that own only a few trucks and therefore only occasionally look for offers.

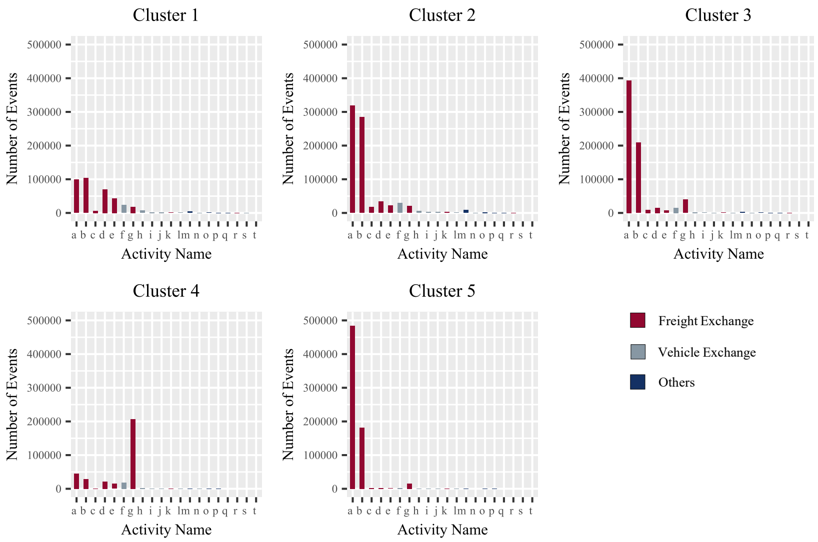
The users combined in the second cluster represent the *full-time allrounders*. First of all, these users are much more active than those in the first cluster. They use the platform for a longer period of time per day and therefore produce many activities. In contrast to the users of cluster 1, working with the TIMOCOM platform is the main task of a *full-time allrounder's* working day. His main activity is to update freight offers, take a closer look at offers of interest and to acquire contracts. Thereby, the user proceeds in a very targeted manner. This is derived from the fact that most of his searches arouse his interest and he takes a closer look at the offers displayed. The users in this cluster are "no specialized power users" (ATR team employee), but rather use many different functionalities of the platform and therefore take different paths.

The users of the third cluster, who are termed *hard-working users*, are characterized by their very strong activity constant over the course of the day. Their main activity is to intensively search for freight offers. In comparison to the users from clusters 1, 2 and 4, the rate of offers that are examined more closely after searching is far lower for the *hard-working users*. A reason for this might be that they have chosen the search criteria in such a way that only a few results are returned, wherefore they have to refresh their search repeatedly before an offer can be selected. It is also possible that their selected search criteria are too "broad", wherefore the results displayed do not meet their expectations. The placement of own freight offers as well as the search for cargo hold are secondary activities of the users in this cluster.

The fourth cluster includes the *passive users*. In contrast to the other users, they show a comparatively low level of activity. Especially, if one disregards the auto-refresh functionality, which is not triggered every time by the user himself, but only has to be activated once, the activity of the *passive users* is comparatively low. The users stand out for the fact that they "simply let the auto-refresh run alongside and check from time to time whether a new interesting offer has been published" (ATR team employee). If this is the case, the offer is checked in detail and in addition manual freight searches are carried out.

The fifth cluster contains the *specialized power users*. In this cluster, it is conspicuous that users perform by far the most activities on the platform during the day. One could assume that this cluster contains "bots" (ATR team employee). However, it is well known at TIMOCOM that certain human users exhibit robot-like behavior on the platform. For this reason, it can be assumed that these users are represented in this cluster.

Fig. 2. Number of Performed Activities per Cluster



5 Lessons Learned

The results of this study give TIMOCOM a better understanding of their user base and the different types of customer groups. Trace Cluster Analysis was found to be a helpful technique in this regard due to the reduction of the complexity of the process models of the clusters in comparison to the process model of the entire event log. The models are easier to analyze and comprehend. Additionally, some behavior is only shown by certain user groups.

The process models resulting from the clusters and the gathered information about the characteristics of the users grouped in the clusters were investigated to find weaknesses and improvement possibilities of the application design of the platform. For example, one issue identified is that a number of functions included a small delay until they were loaded due to implementation limitations and the process model of the occasional users revealed that they often click on these functions repeatedly. Another exemplary issue is that a published freight offer is not directly listed in the list of freight offers, which resulted in a repeated refresh of the list after publication of which TIMOCOM was not aware.

From an implementation perspective, the trace-sequence-based clustering method and the model-based clustering approach implemented in our study are not infinitely scalable, as they load a lot of data into the RAM. Nevertheless, we were able to process the selected 2.9 million events, which can be sufficient to generate valuable insights into a process. In contrast, the feature vector-based clustering approach is completely supported by Spark and therefore scalable to arbitrarily large amounts of data, which can be clustered in a reasonable time.

In retrospect, implementing the trace clustering pipeline was a complex and time-consuming endeavor. Nevertheless, the experiences gained from the initial application prove to be well worth the effort and the implemented pipeline as well as the experiences gained throughout the project can be used in future iterations of the application design improvement process at TIMOCOM.

References

1. Bose, R.J.C., van der Aalst, W.M.P.: Abstractions in process mining: A taxonomy of patterns. In: Dayal, U., Eder, J., Koehler, J., Reijers, H.A. (eds.) Business process management, vol. 5701, pp. 159–175. Springer, Berlin (2009)
2. Bose, R.J.C., van der Aalst, W.M.P.: Context aware trace clustering: Towards improving process mining results. In: Park, H. (ed.) Proceedings of the Ninth SIAM International Conference on Data Mining. pp. 401–412. SIAM, Philadelphia (2009)
3. Bose, R.J.C., van der Aalst, W.M.P.: Trace clustering based on conserved patterns: Towards achieving better process models. In: Rinderle-Ma, S. (ed.) Business Process Management Workshops, vol. 43, pp. 170–181. Springer, Berlin, Heidelberg (2010)
4. Bose, R.J.C., van der Aalst, W.M.P.: Process diagnostics using trace alignment: Opportunities, issues, and challenges. Information Systems **37**(2), 117–141 (2012)
5. Diamantini, C., Genga, L., Potena, D.: Behavioral process mining for unstructured processes. Journal of Intelligent Information Systems **47**(1), 5–32 (2016)
6. Kryszczuk, K., Hurley, P.: Estimation of the number of clusters using multiple clustering validity indices. In: Gayar, N.E., Kittler, J., Roli, F. (eds.) Multiple classifier systems, vol. 5997, pp. 114–123. Springer, Berlin (2010)
7. Lee, Y.L., Hwang, S.L., Min-Yang Wang, E.: An integrated framework for continuous improvement on user satisfaction of information systems. Industrial Management & Data Systems **106**(4), 581–595 (2006)
8. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS quarterly pp. 37–56 (2011)
9. Thaler, T., Ternis, S.F., Fettke, P., Loos, P.: A comparative analysis of process instance cluster techniques. In: Proceedings of the 12th International Conference on Wirtschaftsinformatik. pp. 423–437 (2015)
10. van der Aalst, W.M.P.: Service mining: Using process mining to discover, check, and improve service behavior. IEEE Transactions on Services Computing **6**(4), 525–535 (2013)
11. van der Aalst, W.M.P., Weijters, T., Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering **16**(9), 1128–1142 (2004)
12. de Weerdt, J., vanden Broucke, S., Vanthienen, J., Baesens, B.: Active trace clustering for improved process discovery. IEEE Transactions on Knowledge and Data Engineering **25**(12), 2708–2720 (2013)

Process Mining for Production Processes in the Automotive Industry

Merih Seran Uysal[1], Sebastiaan J. van Zelst[1,2], Tobias Brockhoff[1], Anahita Farhang Ghahfarokhi[1], Mahsa Pourbafrani[1], Ruben Schumacher[4], Sebastian Junglas[4], Günther Schuh[3,4], and Wil M. P. van der Aalst[1,2]

[1] Process and Data Science Chair, RWTH Aachen University, Aachen, Germany
[2] Fraunhofer Institute for Applied Information Technology, Sankt Augustin, Germany
[3] Lab. for Machine Tools and Production Engineering, RWTH Aachen University, Aachen, Germany
[4] e.GO Mobile AG, Aachen, Germany
`uysal@cs.rwth-aachen.de`

Abstract. The increasing digitization of organizations leads to unprecedented amounts of data capturing the behavior of operational processes. On the basis of such data, process mining techniques allow us to obtain a holistic picture of the execution of a company's processes, and their related events. In particular, production companies aiming at reducing the production cycle time and ensuring a high product quality show an increased interest in utilizing process mining in order to identify deviations and bottlenecks in their production processes. In this paper, we present a use case study in which we rigorously investigate how process mining techniques can successfully be applied to real-world data of the car production company e.GO Mobile AG. Furthermore, we present our results facilitating more transparency and valuable insights into the real processes of the company.

Keywords: Process Mining · Internet of Production · Operations Management · Operational Processes · Automotive Industry.

1 Introduction

Process mining [1] is an emerging scientific discipline that allows for extracting knowledge from *event logs*, i.e., collections of historical execution data, available in modern business information systems. Process mining is primarily used to discover, monitor, and improve processes by applying various techniques to event logs generated by the execution of processes [10]. Process discovery, conformance checking, and process enhancement form the three main tasks in process mining, which have been extensively applied to business processes in numerous application fields, such as finance, logistics, and health care. In the literature, very few authors report on the application of process mining in production processes [3,5], e.g., when compared to other application domains such as sales, procurement, banking, and insurance. In addition, existing works mostly focus on use cases at an abstract level [4].

Manufacturing companies aim at establishing automated production processes in order to sustain high production quality and decrease the overall costs of their manufacturing processes. Founded in 2015, e.GO Mobile AG is a young manufacturer of cost-effective and customer-oriented electric vehicles for short-distance traffic. The company was primarily established as a semi-automated manufacturer of electric vehicles for which the involvement of the human resources in the production processes comprises the levels of supervisors and operators at the stations of the factory. As such, e.GO adopted an *agile manufacturing model* with Autonomous Guided Vehicles (AGVs) navigating each car through its production process autonomously. Furthermore, every step within the production process is recorded, i.e., creating a basis for a *digital shadow* of the production process [2]. Within the Internet of Production (IoP) researchers from RWTH Aachen University are developing a reference architecture supporting digital shadows in production. In this context, e.GO is a highly relevant use case for IoP.

On the one hand, e.GO typifies the application of the digital shadow of its production processes by elaborately taking technical and organizational changes into account. On the other hand, the fact that the human operators are an essential part of the production line results in various challenges to deal with. Both the notion of a digital shadow and the inherent variability and flexibility caused by manual operations justify the use of process mining. Hence, it is interesting to examine this semi-automated car manufacturer as a use case and derive insights into its real production processes using process mining.

This paper presents a thorough application and examination of process mining techniques on event logs of semi-automated production processes of e.GO where human resources serve as an active, indispensable part of the production processes. Furthermore, we shed light on the arising challenges in operational processes for which process mining can provide tangible results and support by discovering performance and compliance problems. Consequently, the company is able to leverage its data to gather full transparency about how processes are executed. Our paper is structured as follows: Section 2 introduces the business challenges faced and Section 3 shows the solution approach to deal with these challenges. Section 4 discusses the results and benefits from the applied process mining techniques and Section 5 elaborates on the lessons learned.

2 Business Challenge Faced

In this section, we explain the basic manufacturing process at e.GO, the structure of the obtained process data, as well as the main problems and challenges observed.

2.1 As-Is Manufacturing Process

As indicated before, an autonomous manufacturing model is adopted at e.GO. After basic mounting of a to-be produced car's chassis, i.e., including its electronic engine, etc., AGV navigates the car through the production process. Consider Figure 1 for a schematic overview of the general manufacturing process. AGV navigates the car to a number of *General Assembly Stations* (GAS's), i.e., the horizontal blue chevrons in Figure 1 represent these stations (GA0, ..., GA28). In each GAS a specific (sub)-part of

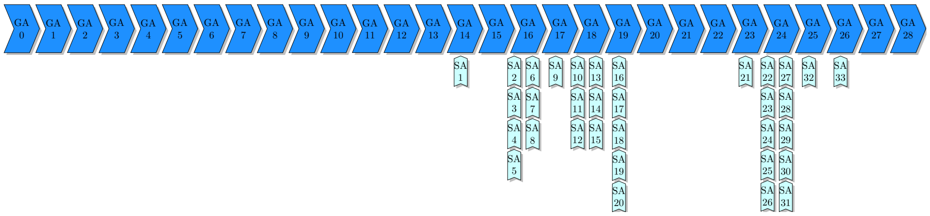


Fig. 1. An overview of the process model generated by the company. The general assembly line consists of *general assembly stations* (e.g., GA0, GA1) and sub-assembly lines include *sub-assembly stations* (e.g., SA1, SA2), which are numbered consecutively.

Table 1. Basic information on the event data used in the case study.

Time range	Two months in 2019
Cars (i.e., cases)	116
Events	12.062
Case attributes	6 (e.g., car ID, color, release version)
Event attributes	10 (e.g., event name, timestamps)

the car assembly is performed, e.g., mounting the doors of the cars. The control-flow of the manufacturing process is strictly sequential, i.e., first GA0 is completed, then GA1, etc. In some cases, preliminary work, i.e., not part of the "general assembly", needs to be performed prior to performing work at a GAS. Such work is executed in a *Sub-Assembly Line* (SAL) comprising *Sub-Assembly Stations* (SAS's), i.e., represented by vertical blue chevrons in Figure 1. Hence, for some GAS, the completion of activities depends on the completion of activities at one or more SAS's.

In its current form, e.GO does not apply any form of *buffering*, i.e., allowing cars to temporarily park somewhere after completion at a GAS. As such, a car awaits completion of the car at the subsequent station in order to proceed.

2.2 Dataset Description

In Table 1, we provide some information on the data which are used in this case study[1]. The event data capture the production of a specific release of the basic low-cost electric car produced by e.GO, i.e., the *e.GO Life*[2]. Within the provided event log, the production of a total of 116 cars is captured. The selected dataset spans a total of 12.062 different events. These events are captured on the GAS/SAL level, i.e., the duration at which the car is present at a GAS, as well as the duration of SAL-based activities is captured. Note that some properties of the car are logged, e.g., the color of a car, as well as some properties specific to the captured events, e.g., which operator was responsible

---

[1] In agreement with the confidentiality policy of e.GO, we apply anonymization and do not reveal station names, process descriptions, release numbers, and properties of cars and events.

[2] A release of a car is a specific configuration of the car, i.e., in terms of sub-assemblies, software, etc. The data reported here are the data considered after preprocessing, i.e., the data obtained after the removal of earlier car releases.

for a specific production step. During the provided period, the average production rate of cars of the specific release considered is ~ 2.7 cars per working day.[3]

2.3 Problems Faced

While preprocessing the data several challenges were encountered. Here we mention two: *(i) Improper Logging*: Timestamps of entering and leaving stations were automatically recorded by AGVs. However, start and end time of activities at the station level were not always recorded properly by the operators. *(ii) Data Quality Issues*: In the information system, the business analyst rarely modified some data related to the rework operations, which also affected the timestamps of some operations in the production line. Thus, the actual time an event takes place and its timestamp in the log differ from each other. As a result, the order of such events seems to be unreliable in the information system, which results in difficulties in attaining reliable process mining results [1,8].

From the productivity management perspective, the company had further challenges to tackle. First, it was of high interest to reveal which stations and activities often cause a delay in production. Second, they were interested in perceiving whether the observed patterns change over time. Third, the question arose about whether the execution of the sub-assembly lines conform with their normative models. The following section elaborates on the methodology considered to overcome these challenges.

3 Solution Approach

The case study performed together with e.GO followed the general steps of the *process mining project methodology* (PM^2) [7]. The preprocessing concerned the estimation of the true service times, since the automatically recorded end timestamps by AGV include both the service times and waiting times. Thus, assuming that the operator confirms the completion of all tasks at a GAS before starting to wait for the next car, we considered the timestamp of his last operation to estimate the service time. Furthermore, as mentioned before, we filtered the dataset for a specific release `RX`.

For an initial overview of the data, we first visualize production activities over time by using the dotted chart implemented in ProM. Second, for the identification of the bottlenecks, we use PM4Py (open source process mining platform in Python) and Pandas (open source data analysis tool built in Python) to generate the results. Since commercial tools, such as Celonis or Disco, based on directly-follows graph cannot deal with the concurrency, they cannot visualize the production line properly. Therefore, we use a custom visualization to visualize the obtained statistics. Third, we analyze the control-flow behavior of particular GA stations indicating bottlenecks and their linked SA lines by using ProM. Since, as expected, the high-level control-flow of the car production is in line with the reference model (see Figure 1)[4], we do not focus on presenting results of conformance checking applied to GA line. Last, to gain deeper insights in process evolution, we examine the process performance on a weekly basis.

---

[3] The true production rate is higher, i.e., only cars of a specific release are included, however, at the beginning of the dataset, cars of an earlier release type were still in production, cf. Figure 2.

[4] These results were obtained by applying conformance checking using car production data and the reference model as presented in Figure 1, i.e., solely using the dark blue chevrons.

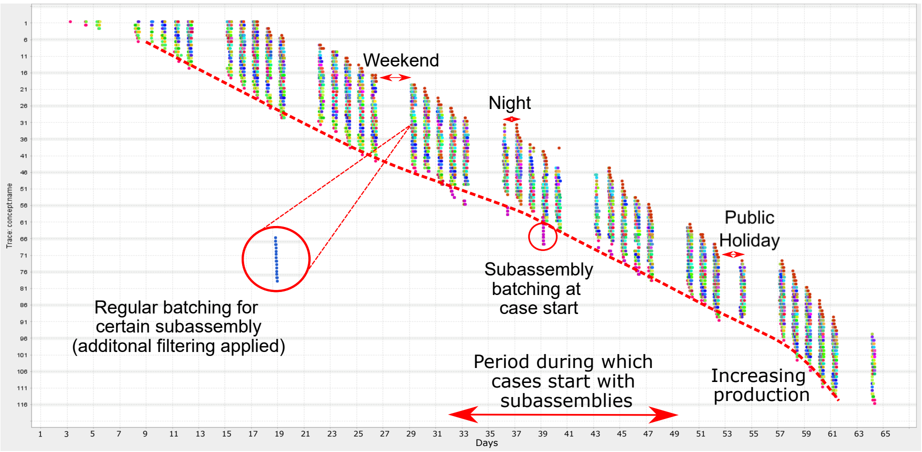


Fig. 2. Dotted Chart Analysis: Visualization of the production of e.GO Life `RX` over time. Each line on the vertical axis represents a car, each dot represents a production activity performed for a car. Production activities either represent leaving a GA station, or the completion of an SA line. The data is sorted on the first production event logged for a car.

4 Benefits

In this section, we present the main results and benefits of the conducted process mining case study, based on the provided e.GO production data.

4.1 Basic Bottleneck Identification

To gather a basic overview of the production of e.GO Life `RX`, consider Figure 2, in which we visualize a *Dotted Chart* [9] analysis of the data for which the temporal range spans two months. Despite its rather simple nature, the dotted chart provides various interesting insights w.r.t. the general behavior of the process, as highlighted in the figure. When considering the production activity captured on the first few days of the production (the number of dots per vertical line in the graph), we observe a rather low activity. However, this is explained by the fact that the data only describe the production of e.GO Life `RX`, i.e., on these days the production of earlier versions is still performed.

Furthermore, we can see typical patterns in the event data. First, we clearly observe weekends and nights in the data, i.e., there is no weekend shift or night shift which continues production. Second, for some sub-assembly lines, we observe *batching behavior* due to improper logging, i.e., a large amount of sub-assemblies are completed and/or executed at the same time. This is illustrated by several similarly colored dots appearing at the same vertical line in the dotted chart. Third, no production activity is recorded on one particular day which coincides with a public holiday in Germany.

To gain better insights in potential bottlenecks, consider Figure 3, in which we present relative performance information on top of the production line. In the figure, the waiting times are visualized in-between stations, using a grey-purple-blue color

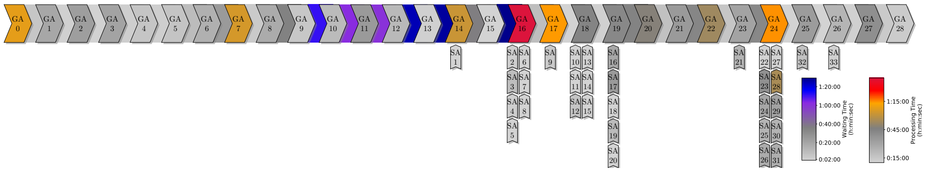


Fig. 3. Visualization of median waiting time (blue color scale) and service time (red color scale) on the GA line and SA stations. The major bottleneck in the process is formed by the general assembly station GA16. Due to the sequential nature of the production process, waiting times, and to some degree service times, of prior stations (GA9,...,GA15) are significantly high.

scale. A light gray color reflects a low waiting time, a purple color reflects an average waiting time, a (dark) blue color reflects a long waiting time. Analogously, the service times of the stations are colored in a gray-orange-red color scale, and are visualized on top of the stations in the production process. Clearly, station *GA 16*, i.e., roughly in the middle of the production process takes the longest. However, note that, this is the first step in the production line which depends on the sub-assembly lines including many SA stations. The idle time, i.e., cars' waiting to proceed to the next station, is highest in-between GA15 and GA16, GA13 and GA14, GA12 and GA13, and GA9 and GA10. Interestingly, the idle time in-between GA14 and GA15 is relatively low, which can be elucidated by the slightly increased service time at station GA14. Finally, note that, out of the 9 general assembly stations that require input from sub-assembly lines, only 4 show a significantly reduced performance. However, all stations depending on sub-assembly lines, perform worse than those without such a dependency.

4.2 Performance Evolution

To better explain the observed performance, we examine the *influx* (aka *arrival rate*) of cars in the production line. Within the dotted chart analysis (Figure 2), the arrival rate is schematically depicted by means of a dashed red line. We observe that an increase in the steepness of said red line corresponds to an increase in influx. Correspondingly, we expect such an increase to be in line with an increased production rate due to no buffering possibilities in the production line. Given the sequential nature of the production line, and, the lack of buffering possibilities, it is likely that the overall production rate, i.e., the average number of produced cars per day, has increased.

Consider Figure 4, in which we depict the process performance for six subsequent weeks in the production ramp-up phase (weeks are presented from top to bottom, e.g., the first week is presented at the top). This further inspection indicates that the production rate almost doubled from $\sim$2.5 cars per day in the first week, to $\sim$4.5 cars per day in the last week. Interestingly, the first week shows a large amount of waiting time and several assembly stations with slow service. In the second week, we observe that certain stations performed better compared to the first week, however, many stations performed worse, and waiting time in-between stations was reduced. In the subsequent weeks, we observe that service times and waiting times were gradually reduced, finally leading to the peak performance observed in the last week captured within the data.

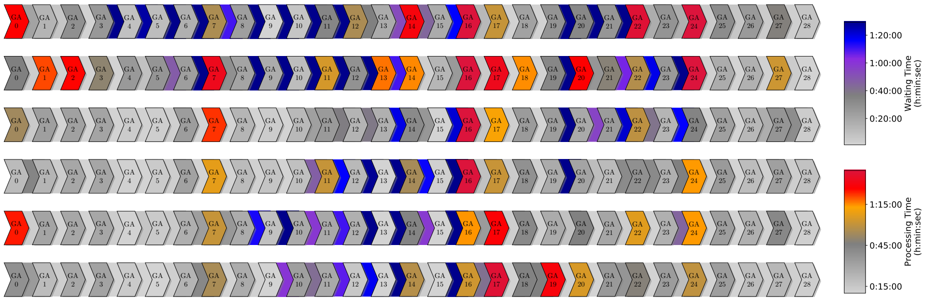


Fig. 4. Visualization of the process performance on a weekly basis, including six consecutive weeks. The most significant waiting times within the production process change over time.

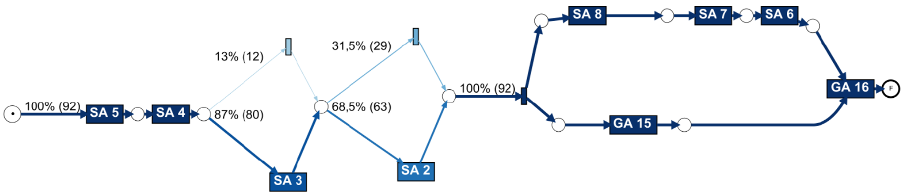


Fig. 5. Discovered control-flow result [6] of the interaction between sub-assembly stations SA2, ..., SA8 and general assembly stations GA15 and GA16. Colors indicate frequency of execution.

4.3 Control-Flow Analysis

Due to the sequential nature of the production process, the application of automated process discovery techniques and/or conformance checking techniques adds little to no value. However, it is worth examining the scheduling of the sub-assembly stations. For example, in Figure 1, we consider the sub-assembly stations which need to be executed, prior to GA16 (i.e., the main bottleneck in the process). There is a clear order between stations SA5, SA4, SA3, SA2, and, also at the same time, between stations SA8, SA7, and SA6. However, these two SA lines are allowed to be executed independently and/or concurrently. We discover a process model for stations SA2, ..., SA8 and GA15 and GA16, as given in Figure 5. Interestingly, the model describes that SA5, SA4, SA3 and SA2 are executed, prior to SA8, SA7 and SA6. Furthermore, we observe that SA stations are indeed executed in the correct order. Note that, in the data, SA3 is skipped in 13% of the cases, and, SA2 is skipped in 31.5% of the cases, which can be attributed to improper logging and data quality issues. Finally, we observe that the full concurrency potential of the sub-assemblies is not achieved within this batch of production.

5 Lessons Learned

We presented a case study using a range of process mining techniques on event logs of a semi-automated automotive production company where human resources conduct assembly sets. In order to attain tangible results and tackle performance and compliance problems, we applied the PM^2 process mining project methodology and analyzed the process execution in the production line comprising general assembly and sub-assembly lines. Furthermore, we observed that automotive production processes are often sequential by nature, allowing for reliable, and, equally important, understandable performance statistics and visualizations.

From both empirical and methodological points of view, an important lesson to learn is that the improvement of the data quality and enabling correct logging will result in more reliable process mining analyses. Consequently, the results will facilitate more transparency and valuable insights into the real processes, which overall serves as a basis for the opportunity to increase the real production rate. Moreover, we will be able to gather more insights into the execution of sub-assembly stations in terms of concurrency, which shows the potential for reducing the production cycle time.

As expected, another important finding emerging during the study is that a bottleneck general assembly station influences the throughput at preceding stations, since no buffering is applied in the production line. In order to circumvent this situation and increase the productivity rate, it is worth considering and simulating bypassing such stations causing substantial delays, which constitutes an interesting direction for future work. Unlike traditional backward-looking approaches, forward-looking techniques, e.g. simulation approaches can be used to answer what-if questions regarding the whole production line and explore various process design alternatives.

Acknowledgments This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC-2023 Internet of Production – 390621612. We also thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

References

1. W. van der Aalst. *Process Mining: Data Science in Action*. Springer-Verlag, Berlin, 2016.
2. T. Bauernhansl et al. The digital shadow of production - a concept for the effective and efficient information supply in dynamic industrial environments. *Pr. CIRP*, 72:69 – 74, 2018.
3. K. Demtroeder and P. Lechner. Agilitaet durch Process Mining, Manager Magazin, Manager Wissen, June 2019.
4. L. Reinkemeyer. *Process Mining in Action*. Springer-Verlag, Berlin, 2020.
5. P. Lechner. *BMW: Process Mining @ Production*, pages 65–73. Springer, Cham, 2020.
6. F. Mannhardt, M. de Leoni, and H. A. Reijers. The multi-perspective process explorer. In *Proc. of BPM Demo Session co-located with BPM 2015*, pages 130–134, 2015.
7. M.L. van Eck et al. PM^2: A Process Mining Project Methodology. In *CAISE 2015*, volume 9097 of *Lecture Notes in Computer Science*, pages 297–313. Springer-Verlag, Berlin, 2015.
8. R. Andrews et al. Leveraging Data Quality to Better Prepare for Process Mining: An Approach Illustrated Through Analysing Road Trauma Pre-Hospital Retrieval and Transport Processes in Queensland. *Int J Environ Res Public Health*, 16(7):1138, 2019.
9. M. Song and W.M.P. van der Aalst. Supporting Process Mining by Showing Events at a Glance. In *WITS 2007*, pages 139–145, 2007.
10. W. van der Aalst et al. Process mining manifesto. In *Business Process Management Workshops*, pages 169–194, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

Geo-Enabled Business Process Modeling

Behshid Behkamal[1], Asef Pourmasoumi[1], Mehdi Akbarian Rastaghi[1], Mohsen Kahani[1], Hamid Reza Motahari-Nezhad[2], Mohammad Allahbakhsh[3,1], and Issa Najafi[4]

[1] WTLab, Computer Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran
`{behkamal,asef.pms,kahani,allahbakhsh}@um.ac.ir`
`mehdi.akbarian@mail.um.ac.ir`
[2] Ernst & Young AI Lab, Palo Alto, USA
`hamid.motahari@ey.com`
[3] University of Zabol, Zabol, Iran
`allahbakhsh@uoz.ac.ir`
[4] National Post Company of Iran, Tehran, Iran
`issa.najafi@post.ir`

Abstract. Recent advancements in location-aware analytics have created novel opportunities in different domains. In the area of process mining, enriching process models with geolocation helps to gain a better understanding of how the process activities are executed in practice. In this paper, we introduce our idea of geo-enabled process modeling and report on our industrial experience. To this end, we present a real-world case study to describe the importance of considering the location in process mining. Then we discuss the shortcomings of currently available process mining tools and propose our novel approach for modeling geo-enabled processes focusing on 1) increasing process interpretability through geo-visualization, 2) incorporating location-related metadata into process analysis, and 3) using location-based measures for the assessment of process performance. Finally, we conclude the paper by future research directions.

Keywords: Location · Process Mining · Geo-Enabled Process Model.

1 Introduction

Process mining is an emerging research area, which helps business analysts and managers gain more insight into their organization's processes. Business process mining research has focused on both intra-organizational as well as cross-organizational processes [1]. There are two main types of cross-organizations process mining. In the first, different organizations work with each other on the same instance of a process to complete it, so they act like puzzle pieces; e.g., the process of 'obtaining a building permit'. The latter group includes the processes in which several organizations use a common infrastructure and execute the same process model with some minor customizations; e.g., the process of 'parcel pickup and delivery'. In both groups, the activities of a business process may be performed in different geographic locations. Thus, the location where the process activities are performed can provide helpful information for analyzing process performance. There are a very limited number of research initiatives that investigate geospatial information for process analysis. The concept of location-dependent task was firstly introduced by Zhu et al. [2]. They presented five location-dependent process model patterns to show how location impacts business process modeling. Zhu et al. also proposed a Petri net modeling mechanism, which incorporates location aspects and ways to constrain the execution of activities [3, 2]. In another work of the same authors [4], they investigated how location impacts the primary logical relationships in a process control-flow. Also, [5] introduced a process mining technique to understand similar movement patterns in an indoor environment.

Our paper reports on our industrial experience in working with business process mining technology while explicitly considering geospatial information. We demonstrate that when business process event logs are enriched with geospatial information of process activities, other peripheral information, which are only obtainable through geolocation can be accessed for performing more advanced business process mining tasks. In the rest of the paper, we first present a motivating scenario in Section 2. Then, we discuss the shortcomings of available process mining tools and propose our solution in Section 3. Our developed tool suite is presented in Section 4. Finally, we conclude the paper and present future research directions in Section 5.

2 Motivation

The work reported in this paper is a result of our industrial investigation of business process mining techniques in collaboration with National Post Company of Iran (NPCI). We present a running example based on data provided by our project partner. One of the most recent challenges in NPCI is to identify the best route for parcel delivery across the country so that they can improve their current distribution process. As such, the process that has been reported throughout this paper is a business process named 'parcel pickup and delivery'. There are four types of activities in this process, each of which is performed at one of the post offices placed in different geographic locations across the country. These activities are: 1) parcel pickup, 2) parcel check-in, when the parcel arrives at a middle station, 3) parcel check-out, when the parcel exits from a middle station, and 4) parcel delivery. All activities are stored in the event logs of the web-based system as presented in Table 1.

In this log file, the *Case_id* refers to *Parcel#* which is a unique number used for tracking, and contains a sequence of activities that are performed to transfer a parcel from source to destination. Besides *Case_id* and *Event_id* which are mandatory components of a process instance, there are supplementary information recorded for each case, including *Timestamp, Resource, City*, and*Location*. Notably, the location of each activity was not recorded in the original log file,

Table 1. A sample of an event log of our experiment

Case_id	Event_id	Properties				
		Timestamp	Activity	Resource	City	Location
1986638	245	25-05-2017: 11.50	Parcel pickup	P.O. 123	Mashhad	37.75888900,45.97833300
	246	25-05-2017:14.01	Parcel check-out	P.O. 123	Mashhad	37.75888900,45.97833300
	247	26-05-2017: 08.12	Parcel check in	P.O. 240	Tehran	37.55527800,45.07250000
	248	26-05-2017: 15.20	Parcel check-out	P.O. 240	Tehran	37.55527800,45.07250000
	249	27-05-2017: 09.27	Parcel check in	P.O. 285	Shiraz	35.84001880,50.93909060
	250	27-05-2017: 10.02	Parcel check-out	P.O. 285	Shiraz	35.84001880,50.93909060
	251	27-05-2017: 14.38	Parcel Delivery	Postman 12	Shiraz	35.12001440,49.93909060

but we have extracted it from Google Maps using the name of the city. In this experiment, we used a small portion of the event log containing 1,137,643 cases. We first draw a process model using BPMN notation, in which the *Case_id* is *Parcel#* and the activities are *parcel pickup*, *parcel check-in*, *parcel check-out* and *parcel delivery*, as shown in Figure 1. In this process model, the activity column was selected as *activity*. This provides a view on the flow of the different process steps; however, it cannot present how different process instances are executed.

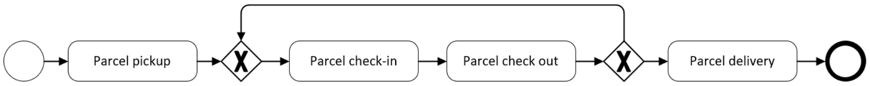
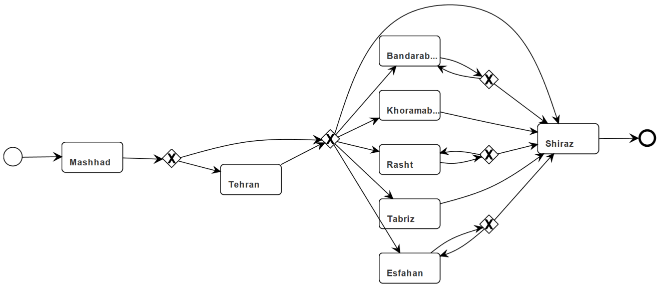


Fig. 1. The process model using activity column as ’activity’

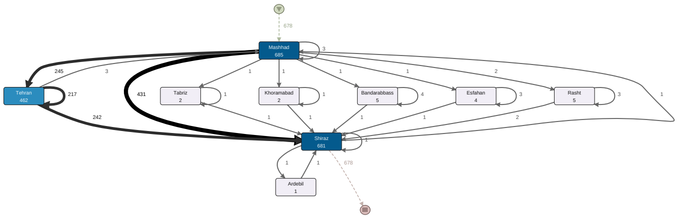
Given such a dataset, we can look at the process in a different way by setting another column as *activity*. In this experiment, we consider city as *activity* and automatically discover the process map using two well-known process mining tools, Disco [6] and ProM [7]. Figure 2 shows a filtered path of the process map demonstrating only those events which have parcels sent from ‘Mashhad’ to ‘Shiraz’. Although these process models provide valuable insight to analysts and process owners, however, they cannot help with providing a geospatial perspective on the process and answering any geo-related question about the process, such as how can the process of ‘parcel pickup and delivery’ be visually rendered on the map? or what are the paths with the highest degree of parcel traffic across the country? In the next section, we investigate the problems of currently available process mining techniques and introduce our solution for geo-enabled process modeling.

3 Proposed Approach for Geo-Enabled Process Modeling

Based on our practical experience in the analysis of cross-organizational business processes [8], we have realized that currently available techniques do not allow including the geospatial information about process execution into the process models, e.g., the location in which the activities are executed. Here, we



(a) ProM tool



(b) Disco tool

Fig. 2. The process map discovered by ProM and Disco tool using city as 'activity'

are going to briefly present the problems we have practically encountered, and introduce our proposed approach for the geo-enabled visualization of the process.

Challenge 1. Complex process models cannot be rendered in an understandable manner. A business process can be the visualized by many well-known standards (e.g., Petri Nets [9], Business Process Model and Notation [10], Workflow Nets [11]) most of which render process models based on control flow in event logs. There are some methods for designing process models in a way that increases their understandability. For instance, in [12], the authors investigate the effect of layout features of process models on model understandability.

Challenge 2. There are valuable location-related metadata that are not considered in process analysis. Geospatial information provides access to additional peripheral metadata such as weather and location hierarchy, which are not accessible if geospatial information is not stored in event logs. For example, if we have some information related to the routes such as ground paths, airlines, and railways, or even more detailed information like maximum speed and the number of lanes in a highway, we can effectively mine and analyze a business process log by taking this additional information into account.

Challenge 3. Location-based metrics are not currently applied for performance assessment of business processes. Typically, three dimensions of performance are identified: time, cost and quality. Time-related information is used to detect deviations and to improve the processes. Also, cost is taken into consideration for performing process analysis [13]. However, addition of geospatial information to event logs makes it possible to consider new metrics for the assessment and enhancement of business processes. For example, in Figure 2, if there are two routes between 'Mashhad' and 'Shiraz' with different path lengths, current tools cannot explain why it takes more time to deliver the parcel via the longer path from 'Mashhad' and 'Shiraz'. In other words, it is not possible to use the geospatial information as a metric to assess business process performance.

To address these problems, we propose a novel modeling approach for geo-enabled processes. We define a geo-enabled process model consisting of four perspectives: 1) process perspective focusing on the control flow of geo-located activities, 2) organizational perspective, which depicts the organizational structure to represent geo-visiulized inter-organizational activities, 3) case perspective, showing the characteristics, including geospatial features, of a process instance, and 4) event perspective, which represents the timing, the order, and the location of the events. In other words, for geo-enabled processes, we add the geospatial perspective of business processes focusing on the geographic location of process activities. In a geo-enabled process model, the nodes of the model represent real-world geographic locations and the transitions between nodes represent the activities taken place between different nodes. Given different scales of locations (from coarse to fine-grained), the process model may be defined hierarchically, with regard to the location, and include hierarchical locations in multiple scales. We may overlay the geo-enabled process model with additional layers on the map to represent location-related data from external data sources to allow for better analysis of the process. We have implemented these abstractions and illustrate the use of such additional layers in the next section. Figure 3 illustrates one realization of a geo-enabled business process model on a map.

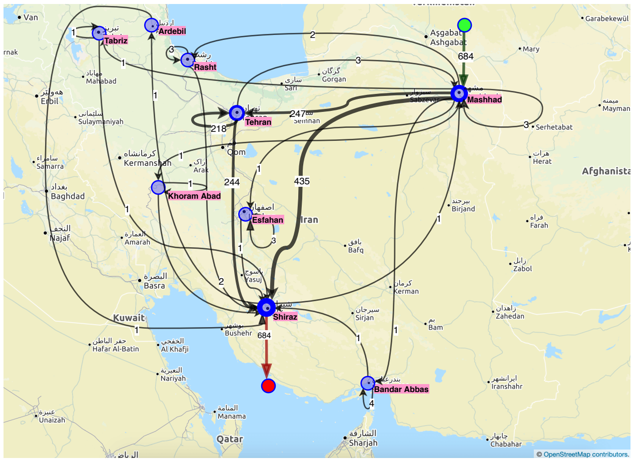


Fig. 3. The geo-enabled business process model

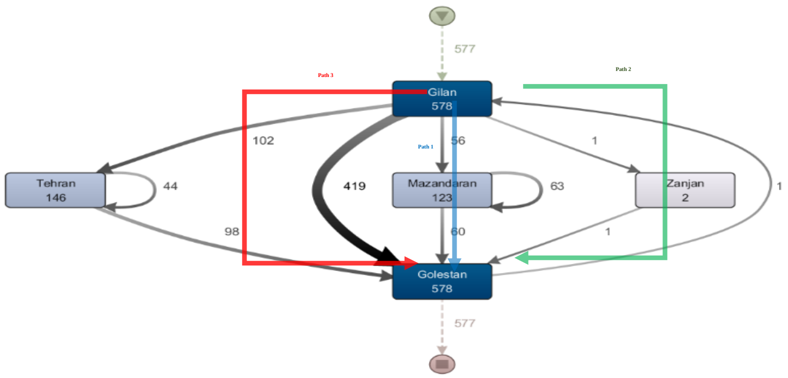
4 Geo-enabled Process Modeling in Practice

To bridge the gap between the process execution in the real world and the process model discovered from event, we have developed a tool suite, called MEEM, using the 'Open Street Map API' platform with an open-content license. MEEM stands for 'Meeting of Events and Evidence on the Map', which integrates two different aspects of data, i.e., '*Events*' and '*Evidence*'. On the '*Events*' perspective, the information related to the process activities is extracted from the geo-tagged event log presented in Table 1. Two layers are implemented in the '*Events*' perspective to show the overview of parcel delivery across the country in both static and dynamic fashions. We have also, implemented the '*Evidence*' perspective in which valuable location-related data available in external data sources are collected and visualized in different layers, such as main roads, airlines and railways. The online demonstration of the tool, repositories, and other dissemination materials are publicly available at *https://github.com/makbn/meem.* Here, we present the advantages of geo-enabled process modeling and address the challenges mentioned in the previous section.

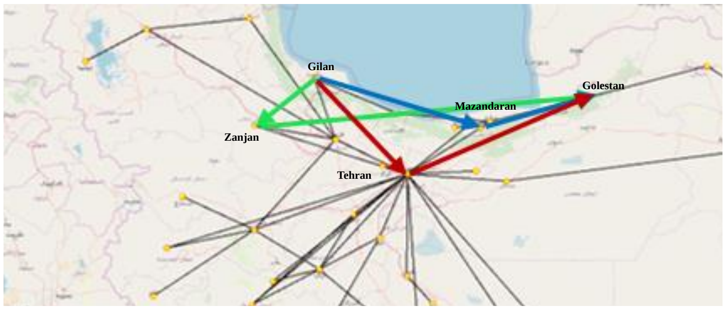
S1. Increasing process interpretability. One of the challenges of current process mining tools is to understandably draw the workflow of a process. Given geospatial information, we are able to address this issue and demonstrate how a process model would be more understandable through geovisualization. To this end, we first draw the workflow of the parcel delivery process without considering geo-tags using Disco [6]. Then, we depict the same process model on the map and explain its main advantages compared to the former. To visualize the workflow of the process, we selected two Iranian provinces, namely 'Gilan' and 'Golestan', as the source and destination, respectively. Then, we filtered for only those events which have parcels sent from 'Gilan' to 'Golestan' using 'Endpoints Filter' option of Disco shown in Figure 4(a). We then depict the process graph discovered from geo-tagged event logs using our tool which is shown in Figure 4(b). It is clear that when the process graph is visualized on a map, it is more understandable as the activities (nodes of the graph) are placed on real-world geographic locations.

S2. Incorporating location-related properties into process models. We believe that process mining will benefit from location hierarchy. For example, we are able to filter the events at different levels of post offices, cities, and countries. So, we have used location hierarchy as metadata to address the second challenge.

S3. Using location-based measures for the assessment of process performance. Considering geospatial information will enable process analysts to improve the efficiency of process execution. For example, the distance between the locations of two consecutive activities can be used as a measure for the assessment of process performance along with the execution time. As shown in Figure 4(a)., there are three routes for parcel delivery from 'Gilan' to 'Golestan', which are drawn in different colors. Based on the thickness of the arrows (that shows the number of parcels which have traveled through that path) we can see that most of the parcels are transferred via Path 1. However, by this representation of the process graph, we cannot detect which path is optimal since the distances between nodes are not presented in this model. When the process



(a) The discovered process map from event log using Disco.



(b) The discovered process map from geo-tagged event log using our tool.

Fig. 4. The filtered paths from 'Gilan' to 'Golestan'

graph is drawn using geo-tagged data, it is possible to place the activities on the exact location they are executed. As such, not only the execution time of a process instance is considered for the assessment of process performance, but also the distance between activity locations can be applied as a complementary metric to better analyze the process execution. Thus, we can simply infer that among all three existing paths from 'Gilan' to 'Golestan', Path 1, which is drawn in blue, is the optimal path in terms of distance between source and destination.

5 Conclusion and Future Research Directions

The goal of this paper is to contribute to process mining by introducing 'geo-enabled process modeling'. We have first motivated our work by presenting a real world scenario and enumerating some challenges faced by available tools. Then, we have presented our developed tool and demonstrated how process mining can benefit from adding the geospatial dimension in practice. There are two promising research directions: 1) visualizing the formal process model on the map. Here, we have just confined to drawing a simple representation of the process graph on the map. To geovisualize the formal process model, the existing plug-ins such as Alpha Miner and the Inductive Visual Miner will be used to create new plugins; 2) Proposing a novel location-based metric. Most of the process mining techniques calculate performance based on execution time. Adding the new location dimensions to event logs, we can better analyze the performance of business processes using a distance-based metric.

References

1. Wil Van Der Aalst, Arya Adriansyah, Ana Karla Alves De Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter Van Den Brand, Ronald Brandtjen, et al. Process mining manifesto. In *International Conference on Business Process Management*, pages 169–194. Springer, 2011.
2. Xinwei Zhu, Jan Recker, Guobin Zhu, and Flávia Maria Santoro. Exploring location-dependency in process modeling. *Business Process Management Journal*, 2014.
3. Xinwei Zhu, Guobin Zhu, Seppe KLM vanden Broucke, Jan Vanthienen, and Bart Baesens. Towards location-aware process modeling and execution. In *International Conference on Business Process Management*, pages 186–197. Springer, 2014.
4. Xinwei Zhu, Seppe Vanden Broucke, Guobin Zhu, Jan Vanthienen, and Bart Baesens. Enabling flexible location-aware business process modeling and execution. *Decision Support Systems*, 83:1–9, 2016.
5. Peter Blank, Michael Maurer, Melanie Siebenhofer, Andreas Rogge-Solti, and Stefan Schonig. Location-aware path alignment in process mining. In *2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 1–8. IEEE, 2016.
6. Christian W Günther and Anne Rozinat. Disco: Discover your processes. *BPM (Demos)*, 940:40–44, 2012.
7. Boudewijn F Van Dongen, Ana Karla A de Medeiros, HMW Verbeek, AJMM Weijters, and Wil MP van Der Aalst. The prom framework: A new era in process mining tool support. In *International conference on application and theory of petri nets*, pages 444–454. Springer, 2005.
8. Asef Pourmasoumi, Mohsen Kahani, and Ebrahim Bagheri. The evolutionary composition of desirable execution traces from event logs. *Future Generation Computer Systems*, 98:78–103, 2019.
9. Wolfgang Reisig. *Petri nets: an introduction*, volume 4. Springer Science & Business Media, 2012.
10. Remco Dijkman, Jorg Hofstetter, and Jana Koehler. *Business Process Model and Notation*. Springer, 2011.
11. Wil MP Van der Aalst. Verification of workflow nets. In *International Conference on Application and Theory of Petri Nets*, pages 407–426. Springer, 1997.
12. Vered Bernstein and Pnina Soffer. Identifying and quantifying visual layout features of business process models. In *Enterprise, Business-Process and Information Systems Modeling*, pages 200–213. Springer, 2015.
13. Bart FA Hompes, Abderrahmane Maaradji, Marcello La Rosa, Marlon Dumas, Joos CAM Buijs, and Wil MP van der Aalst. Discovering causal factors explaining business process performance variation. In *International Conference on Advanced Information Systems Engineering*, pages 177–192. Springer, 2017.

Energy-KPI tracking in building construction processes using Adaptive Case Management

Antonio Manuel Gutiérrez Fernández[1][0000-0002-2669-5942], Christoph Ruhsam[1], Simon Fallmann[2], Anna Shadrina[3], Gerhard Zucker[3]

[1] ISIS Papyrus, Brunn and Gebirge, Austria
{antonio.gutierrez, christoph.ruhsam}@isis-papyrus.com
[2] CES Clean Energy Solutions, Vienna, Austria
s.fallmann@ic-ces.at
[3] Austrian Institute of Technology, Vienna, Austria
{anna.shadrina,gerhard.zucker}@ait.ac.at

Abstract. Clean Energy Solutions (CES) is a competence center for energy efficiency, renewable energy and sustainable development and is cooperating with the Austrian Institute of Technology (AIT) on innovations in building planning and construction. Because of the heterogeneity of building projects, the number of involved parties and the building regulations, a solution founded on the principles of Adaptive Case Management (ACM) is applied to support a holistic workflow. The Papyrus ACM solution Papyrus Converse is used to implement flexible business processes fulfilling business goals rather than following predefined workflows. The planning processes are described as value streams by means of a domain specific ontology and actions such as "upload model version" performed by users (architects, building physicists, etc.) are constrained through formalized business rules to guarantee business compliance. Energy specific goals are defined by evaluating the performance of energy key performance indicators (KPIs), which support the decision-making during the whole planning phase. Permanent energy efficiency tracking is done by calculating these KPIs for each version of the evolving building models. Their evaluation requires formalization in the domain specific business ontology and define how they are assessed like comparison with a target value, trend checking: increasing or decreasing, etc. The specification of such assessments for each KPI is a tedious and error-prone task. In this work, we show with practical use cases how the generalization of the business concepts related to KPI management is practically achieved to provide a continuous observation of the KPIs by applying energy calculation tools to the building models during the value stream execution. These generic concepts can be reused for all the energy indicators of a building project, so that additional KPIs can be easily defined by extending the specific ontology.

Keywords: energy efficiency, planning and construction, Building Information Modeling (BIM), Adaptive Case Management (ACM), Business Ontology

1 Scenario

Clean Energy Solutions (CES)[1] is a competence center for energy efficiency, renewable energy and sustainable development and offers complete solutions out of one hand to its clients taking responsibility for the energy design and management in the building sector. CES cooperates with the Austrian Institute of Technology (AIT)[2] on innovations in planning and construction. The building projects involve a number of different parties with their own procedures and tools. Building Information Modelling (BIM) has been extended in the last years to cover the whole building project life cycle, including the geometric modelling, collaboration mechanisms and building physics information.

A solution based on the principles of Adaptive Case Management (ACM) has been designed to describe flexible business processes based on the achievement of business goals as the heterogeneity of processes, projects and parties involved in a construction project makes it difficult to maintain a solution based on predefined processes with BPM [1]. This solution was implemented with Papyrus Converse [2] to define a building ontology which is used as the vocabulary to formalize the building goals and rules that support the user actions. The business goals are (i) the creation of the building models required to calculate energy indicators, and (ii) the implementation of simulations on the models to derive the energy KPIs. The business cases defined in Papyrus Converse as Value Streams (VSs) follow the business architecture proposal [3] [4] of providing an end-to-end value for the stake holder. For example, the creation of a new architecture model shall be provided by an external contractor such as a mechanical, electrical and plumbing (MEP) engineering company. The VSs defined for the building planning are depicted in Fig. 1.

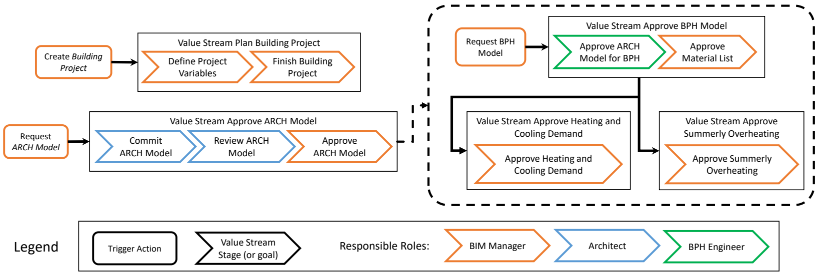


Fig. 1. Value Stream definitions

Three roles were identified: BIM Manager, Architect and BPH Engineer, each of them using their own authorized software, such as Revit for building modelling or CYPETherm for the definition of thermal properties. The Architect provides the architecture model (ARCH model) and does all the necessary changes to the building geometry; the BPH Engineer creates a building physics model (BPH model) based on the ARCH model and designs the material list based on the targeted thermal properties; the BIM

---

[1] https://ic-ces.at/

[2] https://www.ait.ac.at

Manager makes sure the data provided is valid and coordinates all involved parties deliver the information in the requested form. The BIM Manager handles the creation of the building project with its properties, including KPI targets in the VS “Plan Building Project”.

The creation of a new building model starts with the VS “Approve ARCH Model” which aims to create a spatial model of the building by the architect. When an ARCH Model is approved by the BIM Manager, it is extended with the thermal properties of the building elements, described by the BPH model, which is done by the BPH Engineer in the VS “Approve BPH Model”. After the BPH Model was approved by the BIM Manager, the BPH Engineer performs the simulations for heating and cooling demand and summerly overheating in the VSs “Approve Heating and Cooling Demand” and “Approve Summerly Overheating”, respectively, which have also to be approved by the BIM Manager.

The energy efficiency is calculated throughout the planning stages of a building project and assessed regarding specific goals such as energy regulations (e.g.: heating demand is less than 25 kWh/m^{2}a), or to support project internal decisions like that the heating demand for each new model iteration must not be more than 10% of the previous model. This assessment guides users in a compliant way through the building process to take decisions about which are the next actions to fulfill the business goals. The energy-related properties (heating demand or summerly overheating) are used as operational indicators to support the decision-making process towards the optimization of the various building components such as materials or heating, ventilation and air conditioning (HVAC) infrastructure. Therefore, they are considered key performance indicators (KPI) for effective measurement of the achievement of energy goals.

In this work, we track the evolution of energy efficiency KPIs, such as heating and cooling demand and summerly overheating of one of the recent projects from CES, a residential tower project, which we will call Tower throughout this paper.

2 Problem

A number of solutions attempt to manage the evaluation of KPIs for structured business process with BPMN [5] [6] where the predefined workflow of events supports the processing and evaluation of the KPIs. In Papyrus Converse, the first step to manage the energy KPIs is creating the corresponding concepts and properties in a business ontology to support the building information requirements. In Fig. 2, we depict a set of concept instances involved in the tracing of heating demand throughout the Tower project (the names of concepts and properties are in italics).

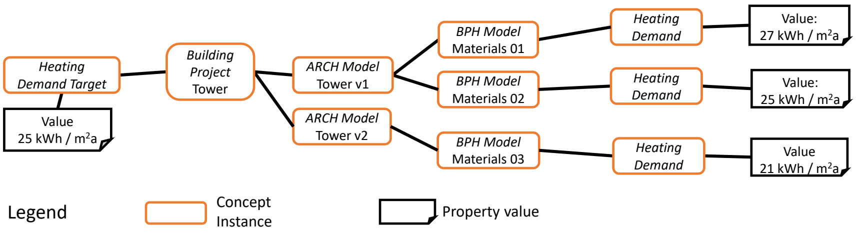


Fig. 2. Business instances of concepts related to the KPI *Heating Demand*

First, the *Building Project* Tower aims at having a heating demand lower than the *Heating Demand Target* 25 kWh/m^{2}a (which is defined according to country regulations, building purposes, etc.). The *Building Project* has two *ARCH Models*, Tower v1 and Tower v2, which are extended with the *BPH Models* Materials 01 and Materials 02 (for Tower v1) and Materials 03 (for Tower v2). The BPH Engineers calculate the *Heating Demand* value for each *BPH Model* as it also depicted in Fig. 2 (27, 25 and 21 kWh / m^{2}a, respectively). The assessment of each *Heating Demand Value* regarding their target requires checking the related concepts at runtime (i.e.: *BPH Model*, *ARCH Model* and *Building Project).* Note: other KPIs involve different VS concepts and relationships.

The implementation of the assessment of such KPIs for given ontology concepts and relationships requires defining the relationship between each instance of the concept for the measured KPI value and its corresponding instance of the concept for the KPI target (e.g. instance of *Heating Demand* and instance of *Heating Demand Target).* In other construction solutions [7], each KPI is ad-hoc managed and evaluated. As it is stated before, the concepts and relationships involved in each KPI calculation are different. However, certain operations to support the assessment of such KPIs (e.g. historic evolution of the value, comparison of current values with the target value) are independent of the KPI. As the creation of the relations between the measured KPI value and the KPI target is a repetitive and error-prone task and increases the complexity of the business application, we propose to generalize such assessments. This problem is addressed in the next section.

3 Actions taken

In order to support the management of KPIs, we extended the business ontology with generic concepts and properties for KPI management independent of the domain, as it is depicted in Fig. 3: *KPI Measurement* with property *Value*, and *KPI Target* with properties *Value* and *Operator*. The *Operator* is used to express the mathematical constrain to assess the achievement of the measured value, such as *Heating Demand* has to be ***less than*** *Heating Demand Target*. The required energy indicators are identified as KPIs with the relation between the specific concept *Heating Demand* with the generic concept *KPI Measurement*. Similarly, other business concepts can be defined as KPI concepts. Summerly Overheating and Cooling Demand are examples of KPIs also related to the BPH Model. We depict in Fig. 3 the concept *Envelope Quality*, which is related to the ARCH Model. The *Envelope Quality* is not directly an energy measure but it is evaluated before creating a BPH Model to extend the ARCH Model. The figure shows how the business concepts from Fig. 2 are extended: "*Heating Demand* is a *KPI Measurement"*, "*Heating Demand Target* is a *KPI Target*" so the generic KPI properties are inherited. Instances of these concepts can be created in the system with the proper property values, such as *Heating Demand Target* has a *Value* of 30 kWh/m^{2}a and the *Operator* is ***less than***. Based on the *KPI Target* and *KPI Measurement* concepts together with the business ontology, the connection between corresponding concept instances are automatically traced in the generic implementations of Papyrus Converse to create KPI reports (i.e. reports assessing the KPI value regarding the KPI target value). For example, for the *Heating Demand*, each instance of the *Heating Demand Target* belongs to a *Building Project* instance, which is connected to a set of *ARCH Model* instances, which are also connected to a set of *BPH Model* instances with their own *Heating Demand*. An extract of a set of connected instances is listed in Table 1.

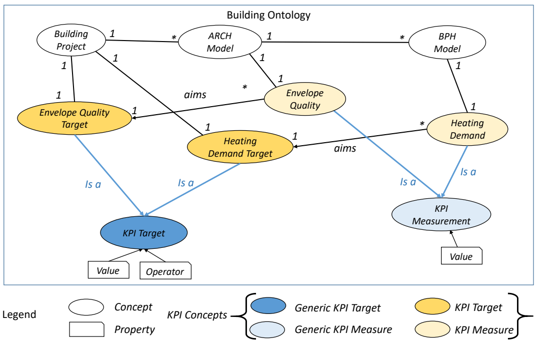
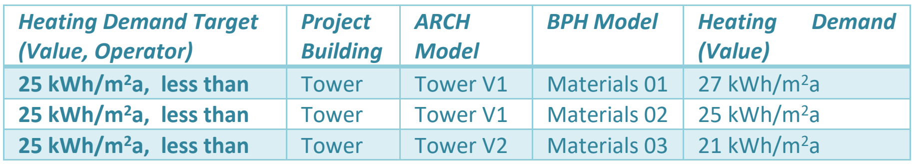


Fig. 3. Ontology enhanced with KPI concepts

Such a list of instances is extracted from building project data, starting from the instances of *Heating Demand Target*, which is the KPI target for the concept *Heating Demand*. This map results of exploring the connected instances of *Heating Demand Target* and *Heating Demand* for the current case. For instance, in the case for the instance of *BPH Model* Materials 01 with *Heating Demand* of 27 kWh/m^{2}a, the connected instance to the *Heating Demand Target* is the *Building Project* Tower. Then, all the instances of *BPH Model* for the *Building Project* Tower are extracted. The Tower (with a *Heating Demand Target Value* of 25 kWh/m^{2}a and an *Operator* ***less than***) has two *ARCH Models*. The first one, Tower v1, has two *BPH Models*, Materials 01 and Materials 02, with their corresponding *Values* for *Heating Demand*, 27 kWh/m^{2}a and 25 kWh/m^{2}a, respectively. The second *ARCH Model*, Tower v2, has only one *BPH Model*, Materials 03, with a *Heating Demand* of 21 kWh/m^{2}a. Therefore, the generation of the report for *Heating Demand* will include three different *Heating Demand Values* for the Tower and only the third of them, 21 kWh/m^{2}a, fulfills the assessment "***less than*** 25 kWh/m^{2}a" (*Heating Demand Target*).



Heading Demand Target (Value, Operator)	*Project Building*	*ARCH Model*	*BPH Model*	*Heating Demand (Value)*
25 kWh/m²a, less than	Tower	Tower V1	Materials 01	27 kWh/m²a
25 kWh/m²a, less than	Tower	Tower V1	Materials 02	25 kWh/m²a
25 kWh/m²a, less than	Tower	Tower V2	Materials 03	21 kWh/m²a

Table 1. Property values to evaluate *Heating Demand*

These reports are automatically provided in Papyrus Converse after the concepts are properly defined as KPIs so the BPH Engineer can uses them to analyze the evolution of the *Heating Demand Value* and its assessment regarding the *Heating Demand Target*, as it is depicted in Fig. 4. which includes a larger sample of *BPH Model* instances than Table 1.

Furthermore, rules based on the evaluation of KPI measurements regarding the KPI target can be added in Papyrus Converse to constrain the possible actions for the user. In Papyrus Converse, these rules for actions have the following syntax:

- In order to [Action], [Constraint over KPI].

For example, the BIM Manager can define the following rule:

"In order to *approve a BPH Model*, its *Heating Demand* must be *less than Heating Demand Target Value*". In the current state of the project, these rules are evaluated for actions modifying the involved concepts (e.g.: Action "Provide *Heating Demand*" for rules evaluating *Heating Demand*) to guarantee the business compliance when users go along with their project work. Alternatively such business rules could also be defined as goals which have to be fulfilled only at the end of a certain value stream.

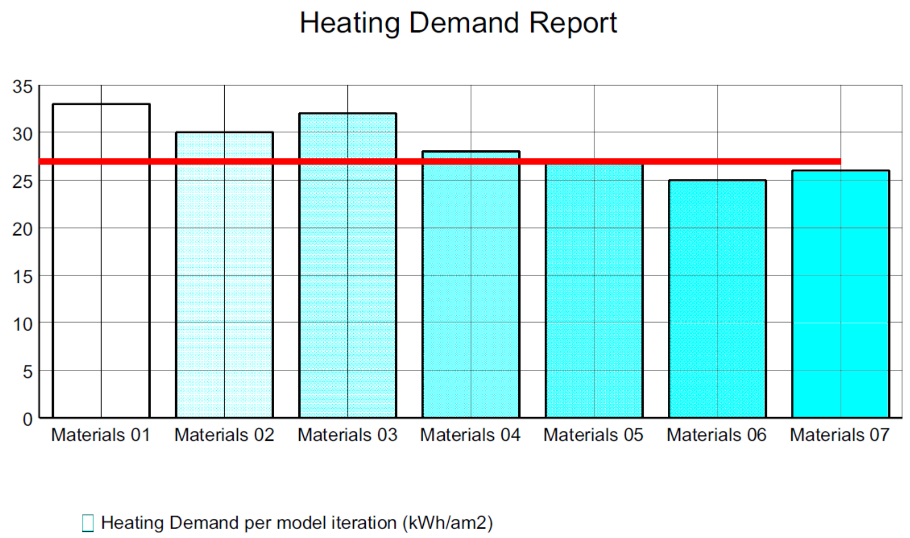


Fig. 4. *Heating Demand* Report

4 Results achieved

As described above, Papyrus Converse [2] has been used to implement the ACM-based continuous tracking and evaluation of the KPIs *Heating* and *Cooling Demand* and *Summerly Overheating* for the Tower. The energy simulation has been performed with products from the CYPE ecosystem (CYPETherm, IFC Builder, etc.). Additionally, CES makes use of the ACM flexibility to execute additional manual and/or automatic steps in the workflow, as it is very important for the participants in the construction industry because the software might vary greatly from one project to another and a fully integration and automated implementation becomes inconvenient and inefficient.

The KPI's are calculated in the VS “Approve BPH Model” using the available *ARCH Model* and the appropriate energy calculation software. As the ontology is used by all the Value Streams in the application, the definition of new Value Streams can also make use of the existing KPI values. Typically, several iterations of an *ARCH Model* as well as of a *BPH Model* are created and tested for compliance with the heating demand regulations and goals resulting in multiple design iterations (e.g. by changing the material, changes in the geometry, etc.). Each new result is stored in the Papyrus Converse value stream instance, so that the KPI evolution can be clearly traced as shown in Fig. 4. The resulting heating demand report offers valuable information to support the future design iterations or further decisions made within the project. It also provides an easy access to the project decision-making history as the changes to the material list and resulting energy demand KPIs are documented in the value stream, offering the necessary tracking and for any auditing purpose like in liability issues. Furthermore, although this solution addresses the energy KPIs in construction projects, the presented approach can be applied in other business domains as well, as it is purely based on the business ontology and thus, its KPIs are freely configurable.

5 Lessons learned

Owing to the state-of-the-art in building simulation tools and certain gaps in the data transfer between ARCH and BPH models, it is technically difficult to incorporate building energy definitions in the early stages of the architectural design. This influences the achievable quality of the building and might cause unwanted last-minute optimizations: when the floor plans and the overall design of the building are already finalized, not much can be done in terms of optimization of the energy use. A clear definition of KPI targets for the energy efficiency in buildings and the ability to track the impact of the changes in the ARCH model on the BPH model fosters the incorporation of the energy design into the early stages of planning.

A flexible, constantly adaptable system of ontology definitions, the relationships between them and finally, the value streams provide a solid structure for the relationships and communication conventions. The need for such systems has been identified by the construction industry, and communication mechanisms are adopted but the use of supporting platforms to formalize them is not sufficiently yet. With the herein described novel ACM-based solution, the important files such as the model files of the ARCH model versions and associated material lists for BPH model are not only kept organized, but also each version associated with a certain decision and a target KPI, is clearly documented. This is a high benefit compared to the current way of storing all this information as unstructured e-mails or project protocols.

The KPIs evaluated in this paper use similar assessments but more complex assessments of KPIs, such as “Cooling Demand has to be ***20% lower than previous*** Cooling Demand”, can limit the development of general mechanisms. Furthermore, the BIM manager wants to define scopes for their evaluation, which the current implementation at CES does not yet consider but will be implemented in the future.

Acknowledgements

This work is funded by the Austrian Research Promotion Agency (FFG) within project BIMSavesEnergy (grant agreement number 861710) where ISIS Papyrus delivers the business architecture enhanced ACM methodology and CES and AIT the construction industry and energy domain specific aspects, respectively.

References

1. Gutiérrez Fernández, A. M., Van Rijswijk, F., Ruhsam, C., Krofak, I., Kogler, K., Shadrina, A., Zucker, G.: Applying adaptive case management to enable energy efficiency performance. *Proceedings of the Industry Forum at BPM 2019 co-located with 17th International Conference on Business Process Management BPM 2019*, Vienna, 2019a.
2. ISIS Papyrus: Papyrus Converse 2019. [Online]. Available: https://www.isis-papyrus.com/e15/ pages/business-apps/papyrus-converse.html.
3. Open Group, Value Streams. The Open Group Architecture Forum Business Architecture Work Stream 2017. [Online]. Available: https://publications.opengroup.org/g170.
4. Gutiérrez Fernández, A. M., Van Rijswijk, F., Ruhsam, C., Krofak, I., Kogler, K., Shadrina, A., Zucker, G.: Applying Business Architecture Principles with Domain-Specific Ontology for ACM modelling: A Building Construction Project Example,» de *Business Process Management Workshops - BPM 2019 International Workshops*, Vienna, 2019b.
5. Han, K. H., Kang, J. G., Song, M.: Two-stage process analysis using the process-based performance measurement framework and business process simulation. *Expert Syst. Appl.*, vol. 36, p. 7080–7086, 2009.
6. del–Río–Ortega, A., Gutiérrez, A. M., Durán, A., Resinas, M., Ruiz–Cortés, A.: Modelling Service Level Agreements for Business Process Outsourcing Services. *Advanced Information Systems Engineering*, Cham, 2015.
7. Chan, A. P. C., Chan, A. P. L.: Key performance indicators for measuring construction success. *Benchmarking*, pp. 203--221, 2004.

Process Mining Applied to Process Redesign in a Domestic Navigation Business Company

Wesley da Silva Santos[1,3], Flavia Maria Santoro[2], Vinicius Arpini Abreu[3], Leandro de Mattos Barral[3]

[1] Federal University of the State of Rio de Janeiro, Rio de Janeiro, Brazil
[2] University of the State of Rio de Janeiro, Rio de Janeiro, Brazil
[3] Log-In Logística Intermodal, Rio de Janeiro, Brazil
wesley.santos@uniriotec.br, flavia@ime.uerj.br, {wesley.silva vinicius.abreu, leandro.barral}@loginlogistica.com.br

Abstract. This paper presents a case study of the adoption of process mining in a domestic navigation business company in the redesign of the "invoice payment request by service suppliers" process. The challenge faced was to understand the process and how the data changed during its evolution along time to elicit the necessary requirements to implement new services. The company followed the Process Mining Project Methodology (PM²).

Keywords: Process Mining, Process Redesign, Domestic Navigation Business.

1 Business Challenge Faced

Log-In[1] is a Brazilian company that offers full door-to-door logistic solutions: planning, management and operation for cargo movement by means of domestic navigation, complemented by roads. The company operates an integrated network which allows for a geographical range in Brazil and Mercosur[2] (The Southern Common Market formed by Argentina, Brazil, Paraguay, Uruguay and Venezuela).

The intermodal transportation is characterized by two or more transportation modals (maritime, road, air and rail) to reduce the logistics costs. Since containers can be easily transferred from one modal to another, the result is the optimization of the cargo displacement and reduction of risks involving accidents or losses. Each container is an independent cargo unit, with dimensions that follow an international standard. The measure adopted to indicate the capacity is known by the acronym TEU, which stands for Twenty feet Equivalent Unit. Domestic navigation takes place between ports in the same country. It is considered a promising modal in countries such as Brazil which has an extensive navigable coast, and the main cities, industrial locations and big consuming centers are concentrated at the coastline or near cities. Compared to the road and railroad transportation, in terms of cost, cargo capacity and environmental impact, the domestic navigation turns to be a viable alternative to comprise the various sectors´

---

[1] https://www.loginlogistica.com.br/

[2] https://www.mercosur.int/en/

supplier chain. Log-In has invested more than 200 million dollars in domestic navigation, with total capacity of 16,700 TEUs. It has more than 1,500 clients in Mercosur and operates with its own and chartered container ships.

The implementation of the process "invoice payment request by service suppliers" emerged as a demand from the Operational Cost Management area at Log-In. This process was considered poorly structured and with gaps in the interactions between the areas involved. Many emails were lost in the inbox of the process participants, and, as a result, updates on some requests to suppliers' status were not sent. There was also a backlog of uncharged invoices due to the lack of an integration with the Enterprise Resource Planning (ERP). In 2017, the Financial department started a project to implement this process on a Business Process Management System (BPMS) aiming to digitalize the process. The areas involved were Procurement, Operational Cost Management and Information Technology.

The initial scope was to create a portal through which the Log-In suppliers could start and follow the process. However, the process was not properly documented, and some typical implementation problems showed up: nobody knew the real end-to-end process; knowledge silos in each area; weak integration with other systems; different types of cases addressed the same way; and low efficiency. When the IT member of the team involved in the digitalization of this process left the company and demands for implementation of new services were requested, it was necessary to go over the process and associated subprocesses from the start. The challenge faced was to understand the process and how the data changed during its evolution along time to elicit the necessary requirements to implement the new services.

The application of process mining [2] seemed to be an alternative to understand the as-is situation and data/control behavior, as well as to help conducting more accurate interviews with process participants and validate hypotheses faster. During this project, many unpredicted outcomes about the process were discovered and in-depth discussions about them made it possible to suggest further improvements.

2 Solution Approach

The company followed the methodology PM2 proposed by [3] which aims to guide process mining projects. The methodology comprises six stages with several different input and output objects of the following types: goal-related objects, data objects, and models. The stages are: (1) planning; (2) extraction, (3) data processing, (4) mining & analysis; (5) evaluation; and (6) process improvement & support. In the stages (1) and (2), the initial research questions are defined, and event data are extracted. The stages (3), (4) and (5) can comprise one or more analysis iterations, possibly in parallel. Each iteration focusses on answering a specific research question by applying process mining and evaluating the discovered process models as well as other findings. If the conclusions are satisfactory then they can be applied for improvement (6).

Planning. The goal of this step is to set up the project and to determine the research questions. Three activities take place: identifying research questions, selecting business processes, and composing project team. Despite some problems related to the quality of data available, the selected process was “invoice payment request by service suppliers”. In our case, the initial goal was to understand the process because the documentation did not correspond to the way the process was actually running, and then improve it by adding the new services requested. The team was composed by 1 process analyst, 3 IT members with low or no expertise about the process and 3 business experts. During the project execution, we noticed that the experts only knew the process at a high level, but process mining helped finding other people that understand the real process operation to answer detailed questions.

Extraction. In an initial data assessment, we observed that part of the data was missing due to non-mandatory fields in the system and many variants were addressed together. Besides, the timestamp of some activities was not precise, which could introduce errors in the process tasks ordering. According to an internal consultant, there was another attribute for the last time an activity was performed, so we decided to order the execution of activities from that attribute. The current process model version was considered the reference as-is model (Figure 1). We renamed the labels of the activities to make the semantics clearer. The process log was extracted from January to December 2019. We used ProM3 to perform the initial evaluation of this log [7]. The log was composed of 4460 cases, 80468 events and 16 event classes. More details about the log could be found in https://bit.ly/2xSpPfZ and https://bit.ly/2U9a7pC. From the log, we found out that there are more cases in the end/beginning of the months and we also discovered the participants responsible for each type of service provided. All data consumed by the process were also extracted.

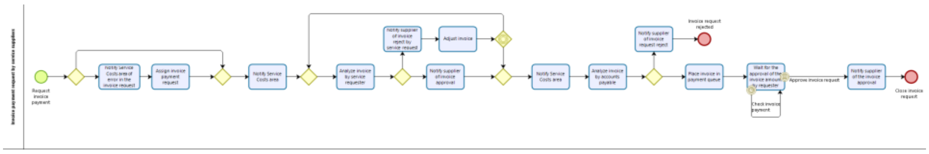


Fig. 1. The AS-IS process model: https://bit.ly/2MdXSTW

Data Processing. In this step, we applied filters to remove the open cases and to include the processing time attribute for each case. On the as-is model events were converted into activities because the ProM does not provide a good support for converting from BPMN to Petri net, which is the format used by most plug-ins.

Mining & Analysis and Evaluation. With the log converted and filtered, Apromore4 was used to discover a model to allow quickly examining the behavior of the process within its dashboard functions [6]. This stage was conducted in 3 iterations. The Evaluation step was performed concurrently with each iteration .

Iteration 1: After the log processing and data assessment, we used the inductive visual miner plug-in available on Apromore and ProM. Inductive Visual Miner aims to bridge a gap between commercial and academic tools by supporting the steps of process exploration and improving evaluation by a new notation and animation [14]. The ID

---

[3] http://www.promtools.org/

[4] https://apromore.org/

attribute of the process instance in the BPMS was used to identify the cases. We could visually detect the absence of activities and paths in the discovered model as opposed to the BPMS model. Thus, we raised some questions based on the model discovered, visually comparing it with the as-is reference model. For example, understanding the lack of certain activities in the model discovered with process mining and why there are more active cases during the end and beginning of the month.

At this point, it was not feasible to run a conformance analysis because we still did not have proper knowledge about the process activities, so, we decided to carry out some interviews with process users, and started working on improving semantics for a better understanding of the flow. The biggest benefit reaped here from process mining was knowing through the model and resources discovered how to come up with more focused questions. The questions were asked concentrating on each type of service. With the application of filters, it was possible to realize which paths one specific type of service took in the process and the average time for its resolution and which attributes were used or not on a type of service. As a result of these interviews, we created another reference model to be used in a future conformance analysis. In this new model, we improved the semantic notation of the process with the knowledge of the users acquired through the interviews. Moreover, we realized that many activities in the process take place outside the BPMS. When asked about this issue, the users said that they decided to focus a priori on digitalizing the process partially on the most important integration (with the ERP) in the approval part of the invoice payment release.

Iteration 2: After renaming the events, we applied the discovery algorithms again. The final model discovered are depicted in Figure 2 and Figure 3. In ProM, the first analysis performed was the visualization of the dotted chart (available at: https://bit.ly/2UbI0WX). We noticed that most cases start and end quickly, but there is a considerable number of outliers that should be analyzed in more detail to understand the context which the instance were. Another important insight gained during analysis was that few requests are opened during the weekends, which in our case was a surprise that some types of service are provided for the ship during weekends or in exchange for a stopover. The ProM "Explore event log" plug-in discovered 20 process variants (available at: https://bit.ly/3dfM8vT). The first three variants show the "happy path" of the process, accounting for 84.7% of the log. The other variants indicate paths that contains rework (requests rejected, re-analyzed and approved or rejected). The main outcome of this analysis was the elaboration of questions to the participants to help understanding the causes that can result in a rejected, re-analyzed or approved instances. The questions were based on common attributes of each case variant such as type of service, the supplier of the case and even the resource performing the internal analysis of such cases. Furthermore, it was also important to specifically understand why there was a variant resulting in a rejection since there was a payment request made in the ERP.

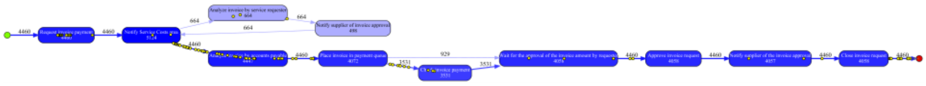


Fig. 2. Process model extracted with ProM: https://bit.ly/36Hk8yZ

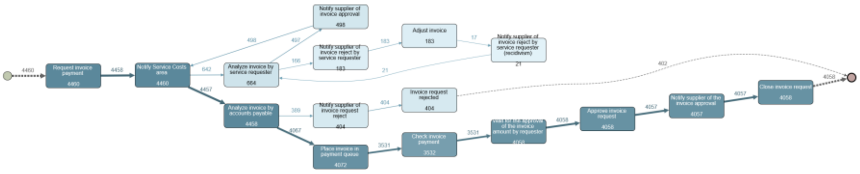


Fig. 3. Process model extracted with Apromore: https://bit.ly/2TRpG4u

Another technique applied in the second iteration was the Discovery of Process Dataflow, using the Decision-Tree Miner plug-in in ProM. This plug-in falls into the category of decision mining algorithms. The goal of decision mining is to detect data dependencies that might affect the routing of a case. Machine learning techniques can be used to discover how data attributes influence the choices made in the process based on past process executions [5]. The partial result is shown in Figure 4. The application of this technique allowed to validate some decisions that occur within the process (such as, the container services following a particular flow with two approvals). Additionally, we could also realize that some suppliers fit into a group where all requests were initially rejected, and therefore, they had to correct the errors in the request.

Fig. 4. Extract of the Process Data-Flow model discovered: https://bit.ly/2MauT38

Iteration 3: We decided to apply conformance checking to analyze how the process cases behaved (details at: https://bit.ly/3djMlhU and https://bit.ly/2Bg2qXI). We used the plug-in "Replay a log on petri net for conformance analysis" of ProM. The replay technique forces a test log to be run through the process model, even if the model and test log are not compatible, so the activities that are missing or left over during the execution are counted and inserted in the fitness calculation in order to obtain a compliance index [1]. This plug-in aims to compare a test log with a model, so that a petri net can be generated with various information about how a process is being performed in relation to its reference model, such as: the fitness between the log and the model, quantification and evidence of alternative paths to the model, points of resource overload and analysis of the individual performance of the modeled activities.

Our hypothesis was that the fitness would be high as the slider on the inductive miner only showed infrequent paths when it came to represent 99% of the log data. As we assumed, there were just a few non-conforming cases when aligned with the model. A more in-depth analysis showed the interaction of the former IT member and the BPMS administrator at some points of time in certain activities of the related cases. We could infer that the former IT member and the BPMS administrator changed manually the outputs of some activities producing non-conforming behavior in the log.

We also made an analysis of precedence relations between activities of the process with bupaR[5] (available at: https://bit.ly/2MbnYXB) [8]. We observed the absolute frequency of the occurrences of each activity that followed another one and used to check compliance, because it synthesizes all the behavior of the process activities. One example is the analysis of successors to the activity “Approve invoice request” that is usually executed before the activity “Notify supplier of the invoice approval”. We identified a case that skipped this activity directly to the final activity “Close invoice request”.

There were some cases in which the events related to messages about the status of the payment release did not occurred, and a more in-depth analysis together with the system's supply team has been carried out. We discovered that in fact the event did not happen even with the condition setup for it.

Another problem was in the activity aimed to analyze the costs of container services. At some point, the system should inform the users about the recurrence of disapproval of the request; but we observed that they were not receiving such notification. Thus, in their perception, as soon as a request is rejected, they should start a re-analysis, even before receiving the communication that the supplier changed the invoice request data.

There was also a specific case where a timer event that checks with the ERP if an invoice request was approved occurred even before the request was made. In this specific case, the former IT member again altered manually the trigger of this event and its execution was canceled and the process proceeded normally. So, in summary, we realized that the process has been evolving over time, but the system was not properly update. All these problems were checked using filters with the conditions that generated the model move only in ProM.

Process Improvement & Support. After the analyzes, a document suggesting improvements was prepared to provide subsidies to the redesign of the process. In addition to the two new services that would be added to address the initial demand, some opportunities for improvement were obtained with the use of process mining. Among them, the integration between the BPMS and the taxes launch system for a better audit assessment after the launch of the invoice in the ERP, and moreover, the integration with the systems related to the service request opening stage are noteworthy. This improvement would ensure that a payment request is automatically opened after the service is provided. Furthermore, there were many errors in requesting payment on account of the supplier's input, causing the rework of opening them or correcting data.

We also found that the digitalized process treats macro activities related to several cases as one. This led to a very general process in which many activities specific to each service are carried out outside the BPMS. These specific activities would be important for obtaining more valuable insights from the process. In some types of services, the cost analysis activity is performed by comparing data from the process instance with data reported by other systems. For these cases, an integration with the BPMS was suggested to carry out automatic and faster checking of data. This is one of the most time-consuming activities in the system with an average of 44 hours, almost two days.

---

[5] https://www.bupar.net/

3 Benefits

The company observed several benefits from this case. First, the most obvious benefits emerged in this case. The team could reach to a fast and objective analysis of the process, avoiding long and tiring interviews or workshops that can take days until full understanding to analyze requirements for new features in a process, and it was possible to start auditing the process from a more transparent and complete perspective, looking at the digitized parts of the process and the resources that perform the activities.

Moreover, we can also highlight the following gains: the discovery and Monitoring of more complex KPIs and the possibility to perform applicability assessment of other automation technologies. Process mining enabled the monitoring of integrations between systems related to processes, providing the possibility of examining more complete KPIs. We can mention the creation of KPIs related to the time between the provision of the service and the opening of the request invoice payment by the supplier, the time to approve invoice request in the ERP, and the charging on the tax system.

In our case, some activities in the digital process that are performed manually could be integrated via BPMS REST API to perform data entry in other internal systems and even make comparisons of the invoice with another system to validate automate decisions. Process mining also made it possible to visualize activities where such integrations would not be a possible approach due to technological limitations of third parties (in our case, some suppliers); however, the orchestration of activities with a BPMS and the use of Robotic Process Automation (RPA) appears as a possibility to not only automate an activity, but also keep it visible in logs for future analysis via process mining.

4 Lessons Learned

We found out that Log-In had more than one information system supporting only this process (ERP, BPMS, and others). The process mining approach identified the lack of integration at some parts of the process and moreover based on the insights provided visually by its outcomes, we could infer how to solve the problem. Grisold et al. [4] explain that some algorithms allow inductive and deductive theorizing. Thus, we can state that process mining (specifically inductive visual mining) was useful for the analysis and monitoring of IT resources integration. BPMS and ERP should integrate the different parts of the requesting invoice payment by supplier process, specifically the request and the consolidation stages.

The process analyzed in this case study is crucial to the domestic navigation business since the services provided by the suppliers could not be neglected. Some delays or reworks could mean big losses. The adoption of process mining in this context impacted not only the identification of improvements but also the reorganization of the whole macro-process. Thus, process mining (specifically variants identification) was valuable for the re-organization of sub-processes and different types of services should be addressed by different process models. The analysis led to discovering the issues discussed in this case and moreover to the suggestions of improvement.

Moreover, we can infer some strategic implications of process mining usage in our case: people cannot prevent from registering manual activities and data collected from suppliers need to be checked and integrated within systems. In this sense, process mining (specifically conformance checking) is useful for the identification of activities performed out of the information systems. Notification of suppliers should be based on data shared within different types of services to guarantee correctness of payment.

It is important to notice that all the findings about the process and related systems came not only from the data but also from the interviews with participants. It confirms that the insights from process mining alone are not able to explain everything about the context of a process or organizational changes as Grisold et al [4] affirm. In this sense, a superficial resource analysis was carried in the very beginning of the project to conduct focused interviews; however, future work includes depth resources analysis such as social network, handover of work between areas and resource performance.

We also have some considerations about the method adopted. Following a systematic approach was very beneficial since the team could organize the work and more than that, it was possible to demonstrate the value added much easier step by step. However, in practice, besides the mining & analysis phase, which is planed through iterations, with the results of each iteration stimulating the next one, it would be important to provide feedback also to the other phases. For example, in our case, with the discovery of resources in the activities, we were able to better compose the team (phase 1) and gain more valuable insights from the process actors. Besides, the evaluation step was performed together with the iterations in the mining & analysis phase; we felt it difficult to separate them. Finally, we missed a distinguished approach for structured vs. unstructured process, i.e., specific guidelines to deal with different types of findings and moreover templates (document, spreadsheet, etc.) to represent the findings and suggestions that could be implemented in the improvement stage.

References

1. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. WIREs Data Mining Knowl. Discov. 2(2), 182–192 (2012).
2. van der Aalst, W.M.P. Process mining: data science in action. 2nd ed. Springer (2016).
3. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P. PM2: A Process Mining Project Methodology. In: Zdravkovic J., Kirikova M., Johannesson P. (eds) Advanced Information Systems Engineering. CAiSE 2015, pp 297-313. Springer, Cham (2015).
4. Grisold T, Wurm B, Mendling J, vom Brocke J (2020) Using process mining to support theorizing about change in organizations. In: Proceedings HICSS 2020.
5. Rozinat A., van der Aalst W.M.P. Decision Mining in ProM. In: Dustdar S., Fiadeiro J.L., Sheth A.P. (eds). LNCS, vol 4102. Springer, Berlin, Heidelber (2006).
6. Conforti, R., Dumas, M., La Rosa, M., Maaradji, A., Nguyen, H., Ostovar, A., Raboczi, S. Analysis of Business Process Variants in Apromore. In Proceedings of the Demo Track of the 13th International Conference on Business Process Management (BPM'15) (2015).
7. De Leoni, M., van der Aalst, W.M.P & Dees, M. A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems, 56*, 235-2 (2016).
8. Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., & Vanhoof, K. bupaR: Enabling reproducible business process analysis. Knowledge-Based Systems, 163, 927-930 (2019).

‘Means’ for achieving process modeling success: Insights from a large Australian bank

Wasana Bandara[1] (0000-0003-2785-7859), John C. Merideth[1, 2], Elizabeth M. Holmes[2],
[1]Queensland University of Technology, School of Information Systems, 2 George Street, Brisbane, Australia
`{w.bandara,j.merideth}@qut.edu.au`
[2] Commonwealth Bank, Sydney, Australia
`Liza.Holmes@cba.com.au`

Abstract. Process modeling plays a critical role in Business Process Management (BPM). There is a plethora of resources on process modeling in general and process modeling success as well. However, practical guidelines about how to obtain common process modeling success factors are lacking. This case comprehensively documents the actions taken to conduct successful process modeling efforts by one of the largest banks in Australia. It demonstrates diverse means that the bank applied for avoiding common process modeling pitfalls and operationalizing critical aspects. The learnings from this case study can be applied by other organizations when designing and executing their process modeling efforts.

Keywords: Process modeling, Critical success factors, illustrative case examples, leadership, standardization, governance

1 Introduction

Business process modeling is an approach used for visually depicting how businesses conduct their operations, defining and depicting entities, activities, enablers, events, states and relationships between them (Dumas, La Rosa, Mendling and Reijers, 2018). Process modeling is extensively employed across diverse organizations and sectors with its effects potentially far-reaching, often resulting in the implementation of new processes/products and services, business models, organizational structures, and IT systems (Dumas, La Rosa, Mendling and Reijers, 2018). Growing reliance on process modeling has resulted in many larger organizations having hundreds of staff (in various roles) involved in designing and maintaining sometimes thousands of models (Bandara, Gable, Tate and Rosemann, 2020).

Guidance on how to maximize outcomes of process modeling is scarce. While literature on process modeling success and failure does exist, their value to practice is limited. For example, Rosemann (2006a); Rosemann (2006b) explains 22 common process modeling pitfalls that should be avoided. Bandara, Gable and Rosemann (2005) and Bandara, Gable, Tate and Rosemann (2020) identify antecedent factors critical to process modeling project success. While they present inventories of factors to consider, the real gap is understanding '*how to'* achieve these. Specific means for realizing those identified factors and better managing them has been called for (e.g. Bandara, Gable and Rosemann, 2005).

This paper presents a rich narrative of different mechanisms deployed by the case organization to achieve successful and sustainable process modeling outcomes. Section 2 introduces the case study. Section 3 presents the means they applied. Section 4 concludes with a summary and lessons for other practitioners. Additional information (e.g. about the modeling tool used, the structure of the process architecture, governance procedures applied and details of the integrated views resulting from modeling) to supplement this case study is provided as ancillary material available at https://drive.google.com/open?id=1ro8ZMQ2VftEyUJ3rofBEhs2fNmgcXnKj.

2 Introducing the case study and Business challenges faced

Commonwealth Bank of Australia (abbreviated to 'CBA') is one of Australia's leading providers of integrated financial services, headquartered in Sydney, Australia with global operations. Underpinned by the three strategic pillars of; 'simplifying the business', 'lead in retail and commercial banking', and being the 'best in digital', CBA positions the customer at its 'core' and aims to provide customers with the best possible products and services. Recognizing that business processes are at the nucleus of deriving and sustaining this goal, CBA places great value in an enterprise-wide process management paradigm, which extensively applies process modeling as a core mechanism in diverse contexts, including; to document existing processes and deconstruct CBA's complexity, support in CBA's business transformation and continuous improvement efforts, and identify process-based software requirements.

Process modeling is not a new concept at CBA. However, the more systematic tool supported, enterprise-wide standardized process modeling efforts that are described here commenced in 2015, with the establishment of a Process Office, which was led by a (then newly appointed) Chief Process Officer. The teams within the Process Office work with the Business units to; develop process modeling capabilities across the bank, help design and maintain the models to predefined standards, and to see that the process models are appropriately applied. Process modeling capabilities (across the bank) was a key. And despite a multitude of training programs (both internal and with external training providers), this took time. While process modeling brought many benefits to CBA, it has not been a trivial task. The overall efforts have consumed much resources and occurred within complex human/ systems/ process-dynamics that had to be carefully managed.

3 The Solution approach: means applied to achieve process modeling success at CBA

At CBA the topmost critical success factors for process modeling were; (i) having a clear 'process of process modeling', (ii) getting the stakeholders to see the value in process modeling, and (ii) obtaining top-down leadership support. Means of obtaining these are discussed below.

3.1 How to establish a 'process of process modeling'

At CBA a clear difference is made with 'Process modeling'[1] Vs process mapping[2]. CBA's process-of-process-modeling is powered by tool-support. CBA's Enterprise Services Leadership team chose a process modeling tool (internally referred to as the process knowledge warehouse - 'PKW') as the Group-wide tool of choice for both Enterprise Architecture and Business Processes. This was designed to; (i) allow for a Group-wide standard tool for process modeling – with one repository of models, (ii) maintain a 'true' end-to-end view of the business processes across the organisation, and (iii) have a 'single source of truth' for process models (i.e. to have a single resource from which to have all process related data and analytics available from). The PKW and its setup also enabled CBA to (iv) have a clear framework for their modeling methodology which allowed for proper modeling governance. (v) PKW has the capacity to be integrated with other systems, which can help CBA's business analysis and reporting capabilities. We describe below how the tool and modeling methodology was set up, with essential governance standards which enabled the effective process of process modeling at CBA.

Effective setup of the modeling tool: The PKW inherently had the ability to integrate different 'views' (see A.1 and A.4 of ancillary material) within its process modeling repository. Thus, different views such as a 'Risk view', 'Data view', 'Product and Service view' and a 'Function/ Application view' were designed to be integrated with the 'Process view' at inception to assist the multi-dimensional reporting deemed critical. The modeling tool's (PKW's) architecture is built with these embedded views (see A.1 of ancillary material for a visual overview). A core of minimum components necessary for process modeling within PKW was designed at inception to lift the PKW from being viewed as 'just a more complicated version of Visio', to being a 'proper' Business Process Management tool. The goal was to have information on risks, systems, controls etc., pertaining to each process available in real-time, enabling a '**process view of 'X' (PVo'X'**) for all processes maintained in PKW (see A.4 of ancillary material for further insights). For example, a *Process View of Risk* (*PVoR*) enabled to identify each potential risk and any mitigating control linked to the level 4 (L4) models of each process (further described below). The integration of the CMDB[3] application into the PKW supported the development of a *PVOT, or Process View of Technology*. The aim was to build an integrated architecture where the activities of a full end-to-end process could be linked to all of the application and support systems that are utilised to perform

---

[1] Process modelling is capturing an ordered sequence of business activities within a business process modelling tool with extended information (such as systems, business rules, process time etc.) so that a process can be analysed, simulated and /or executed).

[2] Process modelling is mere two dimensional drawings of a process that is not linked to any data.

[3] This is CBA's computers and configuration data base.

or validate those activities. A comprehensive integration of process and technology relationships is underway at CBA, with a first release planned in October 2020.

Effective setup of the Modeling Methodology: CBA chose BPMN as the group-wide process modeling language. BPMN is the international standard for process modeling (Object Management Group, 2011), thus enabling a unified and standardized way that everyone can understand. In addition to choosing BPMN, CBA developed a layered approach to define the different levels of granularity for the process models. This resulted in a 'Process Activity Library' (PAL) that consisted of 5 levels: L0 – L4 (see Figure 1). These also provided a consistent process taxonomy.

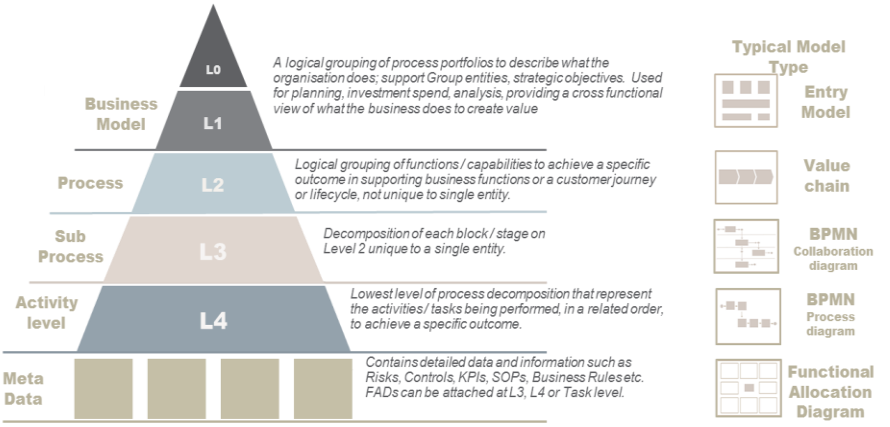


Figure 1: CBA's vertical process-modeling layering (PAL)

Levels L0-L1 were developed by the Process Architecture team to set the vertical process scopes (defining the core and support processes and their boundaries – see ancillary material part A.2 for an overview). Detailed process modeling took place from L2 onwards to elaborate on the processes defined in L0-L1. A L2 process is a high-level value chain diagram (VACD) that can be broken down further, into 2 more layers. A L3 process is a sub-process, or activity block, of the value chain described by L2. A L4 process is a high-level task that can be broken down further into 1 more layer.

All L3 processes and L4 process objects were designed to have a Functional Allocation Diagram (FAD) attached. The completion of the FAD data (also referred to as meta-data) is critical to achieve the process modeling related goals set at CBA. This is because FADs captured a detailed view of a process object that defines the relationship it has with attached information that can be used for analysis. These views show information such as application systems, risks, business requirements, business rules and KPIs. The purpose for capturing this information in a separate model is to ensure the process view of the model remains clear, linear and unobstructed.

CBA created both simple and detailed process modeling guidelines (see A.5 of ancillary material). The simple ones were 'unwritten' rules for the novice modelers from the business areas, which were embedded with the modeling trainings that took place for business areas. The detailed ones had comprehensive modeling conventions, supplemented with model quality assurance, tool usage and modeling governance details, and was primarily designed for the experienced modelers from the Process Office.

The governance of process modeling: All process models are first completed in a separate 'staging area' assigned to each individual PKW user, and once a model is completed within PKW, it went through a quality assurance process before being sent to the relevant Process Owner[4] for approval after which it would be placed in the Production (i.e. Public) area of the PKW. A further level of governance surrounded the addition, modification, or deletion of any elements within the Process Activity Library (PAL) (see ancillary material part A.3). However, the main difference here is that whilst a change to a L3 'PAL modeling block' may be approved by the same centralized team that performed the model quality assurance tests, anything L2 – L0 would require the approval of senior managers accountable for the process.

3.2 How to get the stakeholders to see the value of process modeling

Process modelers depend on process stakeholder participation for the design, review and approval of the models which has a direct impact on the model quality and model application (Dumas, La Rosa, Mendling and Reijers, 2018; Bandara, Gable and Rosemann, 2005). At CBA, a point of constant struggle was to get business stakeholders to spend the time and resources necessary to 'properly' model a full end-to-end process in an appropriate BPM tool, versus simply mapping it (e.g. a flowchart in Microsoft Visio). This challenge was finally overcome with the convergence of two events. (1) The creation of the Chief Process Office within CBA (in 2015), which gave the bank, for the first time, an internal group of process management professionals dedicated to the task of bringing objective, data driven process performance management to the bank's critical processes. And (2) the Hayne Royal Commission led examination of the Australian banking sector (in 2017), which requested banks to review all of the critical processes, risks, mitigating controls, and supporting technology; and to test for regulatory compliance and overall 'health and robustness'. These two events effectively decided the argument in favor of modeling vs. mapping, as only the modeling inherent in a BPM tool was capable of the level of aggregation, interrogation, and reporting required by these types of in-depth operational reviews.

Process mapping (i.e. Visio maps) representations are "one off's", not standardized and unable to be pieced together to give a broader view of the organization as they did not use common methodology and was not in any single repository. The demonstrations of the process view of 'X' (e.g. PVoR and PVoT etc., as explained above) during the commission review created 'penny-drop' moments amongst stakeholders where they collectively realized that the additional time and effort required for modeling was far less than the additional Opportunity, Risk in Change, Technology Discovery and Change, and Operational Risk costs, that could not be lessened or mitigated by simple

---

[4] See Bandara et al., 2019 for a detailed account of CBA's overall BPM governance, which details the role of Process owners and how they were determined and managed.

mapping. To gain stakeholder support, it was critical that they see what the process models can do, and the underlying importance of it all.

3.3 How to obtain essential top-down leadership support

"*The involvement and participation of senior management, and their ongoing commitment and willingness to devote their time and necessary resources to oversee the process modeling efforts*" is an established process modeling critical success factor (Bandara, Gable, Tate and Rosemann, 2020, p. 11). The means of obtaining leadership support is by creating process models that are 'worthy' for leadership to see.

At CBA, in 2015 whilst the Process Office's strategy to model the organization's High Impact Processes was endorsed by the Executive Committee (Ex-Co), the engagement of lower level leaders was variable. The Ex-Co endorsement provided an entry point to discuss with senior leaders in the business the value of process modeling. In general, areas of the business facing the customer had a higher interest in modeling their processes than did the back-office support functions.

Support for BPM declined over subsequent years as organizational changes occurred. It diluted to a point that process modeling was considered a 'nice to have' rather than a necessity and an expectation. In 2018, the central process team was no longer a free resource and business units had to cover the costs for obtaining their services. And while the whole team had been fully engaged within the business it was limited to a few areas (that saw the value of process modeling and continued requesting specialized support).

Recent renaissance of strong executive support and engagement has created a ripple effect across the business resulting in a notable spike of interest in process modeling across the bank. A number of actions have had impact on the resurgence of process modeling in previously disinterested areas. These include: (i) continuous messaging by a business unit executive on the importance of understanding People (Customers and Employees pain points), Process, Risk and Technology layers "front to back" to ensure appropriate investment decision making. (ii) Challenging those with two-dimensional process maps where the information they had on the process was stored, asking for it to be entered into the process repository so that it was accessible and searchable by all employees. (iii) Using widely spread video conferencing to discuss the learnings from incidents to re-iterate the importance of having up to date documentation in a discoverable repository so that it makes it easier not only for new employees to be onboarded and understand processes and procedures but also as a starting point for impact analysis in the event of a service outage. (iv) Sponsorship and funding of an Enterprise-wide program to understand and model all the processes and IT Services that enable the Essential Functions of the bank[5]. (v) Uplifting governance authority to the architecture teams; to gain oversight on all changes requested by the business and to ensure investment decisions have considered end-to-end process impacts of proposed changes.

---

[5] See Banking Finance Sector Group Trusted Information Sharing Network at https://www.tisn.gov.au/the-tisn/banking-and-finance for further details.

4 Benefits and lessons learned

Process Modeling has bought many benefits for CBA. It has helped to; bridge the gap between different parts of the business (e.g. Technology experts and Business Experts, who often find it hard to communicate), get business managers to consider up and down stream consequences of processes and combine forces to collaboratively solve problems, and create an end-to-end ownership of processes across a matrix organization. The 'process view of X' was a highlight that enabled corporate wide value adding impact from process modeling (see further details in ancillary material part A.4). Reflections of CBA's modeling efforts highlight several things that must be practiced by all organizations engaged in large scale process modeling, they are outlined below:

Methodological and tool insights:

- Establish a standard process modeling methodology embedded within a modeling tool fit for purpose. This should be supported with templates to standardize the output of the process modeling exercises.
- A process taxonomy should be created and integrated into other key organizational policies such as those around Risk, Technology and Data as well as into investment funding and delivery frameworks.
- The standards and governance set for the tools and modeling should not be overdone, else they become a non-value added overhead. People need to be sold on the value of using the set tool and methods.
- Set up regular meetings with the process office and business teams to walk through progress. Gain the business teams' commitment to attend these meetings (by showing its worth their time), schedule the meetings from the beginning through to the scheduled end of the engagement. Act on disengagement (e.g. halt work on the project if the stakeholder starts to disengage or not show up to meetings even if they are paying for your time).
- Ensure the resulting model(s) are fully embedded in the business (i.e. will not gather dust beyond delivery). Be clear on roles & responsibilities to regularly review the information to ensure currency. Provide capability uplift where needed

Executive engagement

- Metrics to demonstrate value of process modeling (and process management) should be established.
- Modeling should become part of the organizational culture, so all executives (current and new) will automatically see process modeling value.

Process stakeholder engagement:

- Start with areas that will promise most feasible impact.
    - Choosing areas with urgent problems helps to gain advocacy. As the business will need to work on it anyway- so modeling will be helping and not seen an additional task.
    - Projects must have a clear aim and scope. For example, one type of meta data (e.g. Process views of Risk or Technology) can be picked up to be associated to the process, which when done well can provide impactful insights.

- Start with the end in mind, with a clear view on how the process models will be used; and design the modeling approach from the outset to address these.
- Have an established stakeholder communication plan and execute it (e.g. showcase good/ success stories, explicitly demonstrating ‘value added’ of modeling).
- Involve process stakeholders in showcasing the modeling outcomes to other interested leaders (e.g. what decisions they have been able to make, and the benefits to their business as a result).
- Conduct a proof of concept with one business area and use the results to gain the interest and commitment for modeling in others.
- Modeling capability building should be ongoing with on the job coaching and a champion network created across the organization. Capability building should cover standardized modeling methodology, the use of the tool and effective workshop facilitation across different stakeholders.
- Involve the Stakeholder from kick-off to launch. Have ceremonial “handover” of the models to the business (where the business process owner talks the executive through the final delivery). This provides the opportunity to formally and symbolically re-enforce process model ownership sits with the business.

References

1. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer, Berlin, Heidelberg (2018)
2. Bandara, W., Gable, G., Tate, M., Rosemann, M.: A validated business process modeling success factors model. Business Process Management Journal (2020)
3. Rosemann, M.: Potential pitfalls of process modeling: part A. Business Process Management Journal 12, 249–254 (2006a)
4. Rosemann, M.: Potential pitfalls of process modeling: part B. Business Process Management Journal 12, 377–384 (2006b)
5. Bandara, W., Gable, G.G., Rosemann, M.: Factors and measures of business process modeling: model building through a multiple case study. European Journal of Information Systems 14, 347-360 (2005)
6. Object Management Group: Business Process Model and Notation (BPMN): Version 2.0. Object Management Group (2011)

Flexibility, Adherence, and Guidance for Regulated Processes with Case Management

Johannes Tenschert[1,2] and Michelle Hormesch[1]

[1] Pertuniti GmbH, Martin-Behaim-Str. 10, 90489 Nuremberg, Germany
{johannes,michelle}@pertuniti.de - www.pertuniti.com
[2] Chair of Digital Industrial Service Systems, University of Erlangen-Nuremberg,
johannes.tenschert@fau.de

Abstract. Knowledge-intensive processes in regulated domains show high requirements in regard to documenting processes and their instances, but inherently as knowledge-intensive also in flexibility. This typically leads to a lack of automation, manual documentation, and the process being scattered across many support systems. At Pertuniti, we combine the flexibility of social software and adaptive case management with subprocess modeling for aspects that can be modeled and automated. This way, we can provide one transparent system of record that allows implicit documentation of structured and ad-hoc activities, managing and linking all case artifacts, and an iterative approach on which aspects are modeled. We cope with drastically varying requirements on process support with a flexible and adaptable data model, and by facilitating low-code implementations of changes and transparently deviating from models.

Keywords: adaptive case management · quality management · subprocess modeling · knowledge-intensive business process · process flexibility

1 Introduction

Organizations in regulated domains need to consider many norms, e. g. for quality management (QM), and a large set of laws, guidelines, and further regulations. In QM, core processes become transparent to provide traceable, consistent, and continuously improving quality of products and services. In knowledge-intensive processes, planning is part of the work. Consolidating legal and organizational requirements is challenging on its own. Given that regulation is often tackled with predefined process models and a high degree of instance documentation, combining knowledge-intensive work with regulation yields obviously conflicting goals. Still, not all parts of norms enforce the degree of structure applied in practice, but they require a part of the process to be known a priori, and concrete instances to be traceable appropriate to their risk.

Knowledge workers like engineers, lawyers, or managers often apply a broad range of tools. They emphasize on flexibility and communication, e. g. social software and groupware. As their processes typically show a high degree of variability, the whole process can rarely be modeled a priori. Core activities get handled

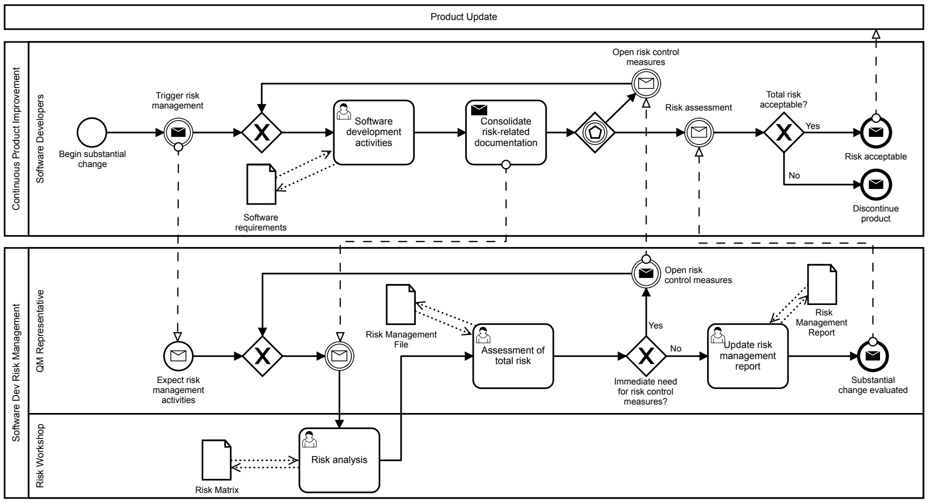


Fig. 1. Medical software product development and risk management (simplified)

outside the business process management system (BPMS), e. g. software development (project management) in Figure 1, as it only tracks performance without supporting it. Often, structured fragments could be modeled and automated.

At Pertuniti, we combine the flexibility of social software and adaptive case management (ACM) with subprocess modeling to facilitate traceability for the whole case and guide through activities that are actually enforced. Pertuniti targets knowledge-intensive processes emphasizing on ad-hoc processes, i. e. we implement these processes as project and knowledge management. Here, we outline challenges and our solution approach for extending it with flexible support of modeled, regulated processes in a consolidated system of record.

The following sections introduce the challenges we faced in supporting regulated domains, and the fundamentals for our solution approach. In Section 4, we outline our solution approach of combining adaptive case management with subprocess modeling and data management. We discuss these results and the lessons we learned. Finally, we conclude the paper and outline future work.

2 Challenge

Currently, we evolve the ACM platform Pertuniti for regulated domains that emphasize on laws, norms, and other regulations, but primarily are knowledge-intensive. With experience in medical quality management, we opted for this domain, but the nature of automotive, aviation, insurance, and finance is similar.

Our running example is a simplified process for updating existing medical device software. A slightly extended version of Figure 1 was successfully certified for ISO 13485 [4]. Companies in this domain have to consider many norms, e. g. also ISO 14971 (risk management) [5], ISO 62304 (software development lifecycle) [1], and ISO 62366-1 (usability) [2], as well as FDA regulation in the US, or the medical device regulation [8,9] and country-specific laws in the EU.

The challenge in developing medical devices is that engineering processes are *the* example for knowledge work: product development and the product itself unfold. Every development plan is performed only once, i. e. the purpose of the process is project management, and it will typically not be performed within a BPMS, as there, it is merely an abstract task. While creative, knowledge-intensive tasks rarely offer possibilities for automation, they do need a lot of documentation and templates, i. e. the typical goals of process support systems may differ in regulated domains. Scattered process data is especially expensive in regulated domains, as every system has to be evaluated. Consolidating information in one or few data source(s) is desirable for backup strategies and avoiding contradictory data. Transparency is mandatory in regulated domains.

Deciding for a system on a per-process basis may not yield the intended results, and just applying ad-hoc processes does not suffice: Often, some *aspects* are structured, while others are ad hoc. Hence, apart from engineering, this is also problematic in e. g. internal / external audits or CAPA. While creators of the processes know which parts are constrained by regulation, stakeholders applying them often do not. In summary, QM-related process support systems need to consider project management and ad-hoc activities to be documented, small automation (e. g. document templates), and structured *fragments* either for automation or as a guide through the process with a varying need for flexibility.

3 Fundamentals

Swenson introduced the term **adaptive case management** (ACM) in *Mastering the Unpredictable* [17]. There is not yet a clear definition that vendors and academia have agreed upon. This led to a wide range of "ACM" systems from a focus on modeling e. g. in CMMN or BPMN to configurable systems similar to groupware. We expect the following characteristics: ACM systems support emergent or unstructured processes [18]. An ACM system provides means to manage ad-hoc activities and master data, as *adaptive*, it is configurable and applicable to multiple domains, and it may provide modeling and automation capabilities.

Multiple paradigms for **process modeling** and automation are prevalent in case management. Since knowledge workers perform knowledge-intensive *and* routine work, many are necessary and useful in practice. Imperative languages like BPMN [13] are especially useful for processes with a high volume and low number of variants, i. e. routine work. Declarative languages like CMMN [14] and DCR graphs [12] describe dependencies between activities and allow a high number of variants and flexibility in the order of activities. Moreover, languages like BPMN-D [10] and accepting petri nets [6] show characteristics of both paradigms. Typically, the right notation depends on many parameters including the process, modeler, variability, external conditions (e. g. audits), and more.

Quality management systems are a set of documented core processes of an organization that influence the quality of their products and services.

They facilitate consistent and continuously improving quality. In regulated domains, applying a QM system is often mandatory. Common QM systems are ISO 9001 [3] and ISO 13485 [4]. During certification, the norm to certify for is typically not the only one to consider (cf. Section 2).

4 Solution Approach

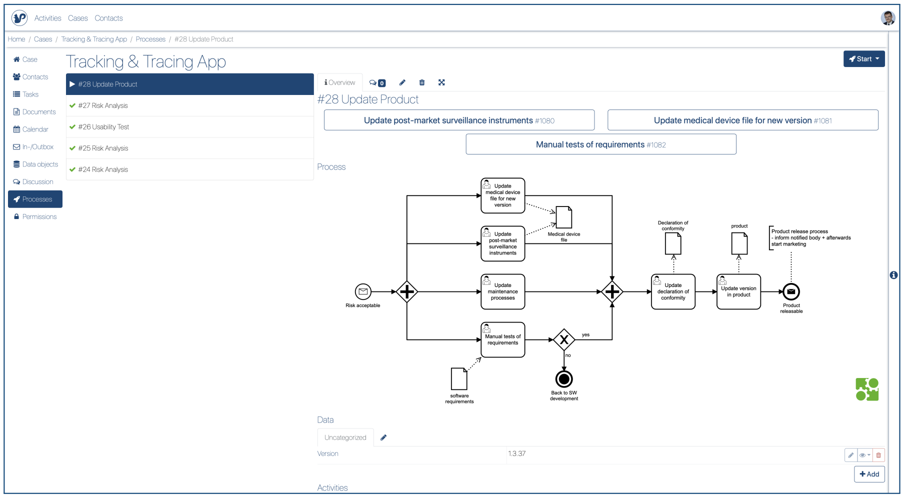
We address these challenges by covering requirements of knowledge-intensive processes, and then extending the approach with structure for automation and traceable compliance with regulations. Unlike traditional BPM, we see *deviations* from the process as an indicator for missing process model variants. This is a valid assumption as many regulated processes are not frequently performed, e. g. yearly internal audits or irregular risk analyses. Qualified team members know when to deviate from the plan, but they prefer being guided through regulation.

Pertuniti is an ACM solution resulting from the first author's PhD thesis [19]. The main intention is to support processes without process models. Hence, we implement process management as project and knowledge management. The basic idea is to apply ad hoc as the default, and make all activities of a case traceable - including activities on information. We started with basic automation like user-defined low-code document templates, and later added more options for structure. With ad hoc as the default, the approach allows to transparently deviate from models and supports processes that have not been modeled at all.

The running example depicts requirements in ad-hoc activities, modeling, data management, and traceability. Activities that influence quality of the product or risk for users need to be documented. Software development requires support in regard to project management, e. g. managing tasks, capturing requirements, identifying stakeholders, managing resulting files. Larger subprojects, Scrum sprints, milestones, or features could be managed as a subcase that shares master data with the overall case. Master data of different cases might deviate for other medical devices even in regard to the attributes stored, not only concrete values. In the example, there are completely ad-hoc tasks mostly for development activities, which should be documented to be re-used, and in risk analyses, where documentation is a protocol and results get revisited in later analyses.

4.1 Social adaptive case management

As we apply ad hoc as the default, Pertuniti seems similar to popular groupware, also in terms of integration with external calendar and contacts clients, but different in that all information that is typically scattered across many data silos is either captured or referenced from one single data source. We emphasize on collaboration by adopting activity streams that are prevalent in social software. These allow to track and trace what is happening in a case and within the organization. We provide activity streams on different levels: all visible cases, all case activities, all artifact activities. As all artifacts are stored or referenced within



Version 1.3.37

Fig. 2. Running multiple subprocess instances in Pertuniti

the same system of record, their relations can easily be made explicit, e. g. some requirements documents and stakeholders to a certain task. All artifacts provide an EAV schema for adding attributes on demand. To facilitate discussions and capture additional information, notes can be added on artifact and case level.

Currently, we provide structures for contacts, tasks, events, documents, correspondence, and arbitrary EAV data objects. The distinction enables client integration via CalDAV and CardDAV (calendar, tasks, contacts), and file synchronization. The flexibility allows users and teams to decide on and adapt their workflow themselves, and to capture information as they deem most appropriate.

4.2 Subprocess modeling

We apply modeled processes as subprocesses of a case. Subprocesses have a state and their execution is loosely isolated. They share the context of the parent case, and could use it to communicate, but each instance also has its own local context. Multiple instances of the same or of different models can be started for each case. One instance is always assigned to one case and process model version. We first applied the approach in Agora [20]. This approach captures our basic assumption that deviations from the model are always to be expected, e. g. additional tasks or changes can be applied ad hoc. It does not reduce capabilities of the model as a subprocess that runs as expected and does not include additional information results in a case with one process capturing the situation like a traditional BPMS.

As subprocess models and instances do not need to be connected, actual notations are only relevant in appropriateness for the case fragment. For example, structured risk management activities can be modeled in BPMN, while documentation activities for product updates with high variability are modeled as a DCR graph. Deviations can be performed manually, e. g. by managing corresponding tasks and artifacts in the case, referencing and commenting them for traceability, and by dividing processes into smaller fragments that are initiated as often as they are needed. In medical QM, we detected that most regulated processes benefit from being modeled as a subprocess. Figure 2 shows how different models are performed as subprocess instances within a flexible case.

4.3 Hierarchic data management

Hierarchic case structures are prevalent in BPM and in project management. In project management, this is typically a task-level hierarchy or hierarchically managed tasks. In BPM, data structures for subprocesses typically are message passing and in/outbound data objects, e. g. according to workflow data patterns [15]. If an ACMS should support gathering master data manually and automatically in a case hierarchy, ad-hoc data modeling becomes important.

Cases and artifacts contain EAV master data with tabbing, ordering, *inheritence* of attributes, ad-hoc parent case imports, lists, and blob attributes. Pertuniti captures case hierarchies. Next to predefined artifacts (documents, tasks, events, correspondence), cases may contain EAV data objects. These allow nesting via artifact references in attributes. Users may predefine data object schemata for visualization, translations, and forms, i. e. low-code entities.

5 Benefits

Our primary benefit is to facilitate **one system of record** for all relevant case and process data. All artifacts are managed or at least referenced in a case (hierarchy). Pertuniti contains and synchronizes groupware artifacts, but also allows creating new data schemata. As most artifacts are managed within the same system, relations can easily be made explicit and automation only considers one target. One system of record reduces the complexity of backup strategies.

We provide **activity streams** on case, process, and artifact level for traceability even in ad-hoc activities. Knowledge workers can track their own and team member's actions for coordination, but they can also pose as an audit trail. The case activity stream consolidates all ad-hoc activities and **implicit documentation** of process-level activity streams, i. e. performance of modeled activities. For manual work, stakeholders still need to adhere to regulation, but the activity stream can drastically reduce manual documentation.

The approach **combines regulated processes with project management**. The main aspects of the work do not need to be managed outside the BPMS. Automation within a flexible case allows to **transparently deviate** from modeled aspects, either in performing work manually, or in compensating or performing some work outside the process instance, but within the same case.

Finally, subprocesses can also **guide through regulation**. As the process models might already be provided for audits, stakeholders need to know how to apply them. “Executing” models without automation can be used as a guide or to simulate the process prior to actually performing it. Test runs can reduce fear of making mistakes, and remind which activities to perform and when.

6 Lessons Learned

Integrate data and processes. In regulated domains, each system influencing the quality has to be audited, and organizations need a backup strategy. Even if the number of tools is not reduced with it, a consolidated data model reduces backup efforts. In knowledge-intensive processes, artifacts are scattered across different data types and systems. Integration of data and processes allows using the best software for the task at hand, but a focus on standards and *one* system to integrate with can drastically reduce integration efforts. Typically, artifacts in different systems are more tightly coupled than their storage. Stakeholders take notes to relate tasks of one system with documents, events, or data in another. A consolidated data model can explicitly capture loosely coupled artifacts.

Requirements of ACM users vary. Even though manufacturers of medical devices need to apply the same norms, their requirements vary. They provide different products and services that directly influence the focus of their core processes. End users have different experience in BPM, data management, and automation. Requirements may change over time with experience. Hence, systems need to be configurable and adaptable to new situations.

Organizational development. Regulated organizations are typically required to switch between hierarchical and process views [11]. Transforming to a process-oriented view requires BPM competencies [7]. To develop a shared skill set, teams need *practice fields* or *microworlds* [16]. Practice fields are "computer-supported environments where team learning confronts the dynamics of complex business realities", and team members receive immediate feedback on consequences of their actions. We support iterative learning of BPM competencies by enabling users to utilize, see, and test their models. New processes directly become executable and organizational knowledge can evolve over time. Enhanced cross-functional communication facilitates implicit learning and organizational change by sharing information across levels. Facultative features enable end users to improve processes, e. g. via templates or automation. Processes can remain unchanged while a freedom to adapt them improves independence and acceptance.

7 Conclusion

Subprocess modeling and hierarchic data management can combine ad-hoc and structured activities in one system of record. This approach facilitates knowledge-intensive processes that are typically scattered across many data silos. As not the whole process has to be modeled a priori, we facilitate organizational learning: Organizations may apply an iterative approach for core processes, and capture more routine work later. In regulated domains, we enable processes that can easily be audited combined with project management, i. e. we are not restricted to support processes. In the future, we extend Pertuniti with additional notations, a larger subset of BPMN, and interfaces for BPMSs. Since guiding through regulation is often more important than automation, we want to further facilitate creating *interactive QM handbooks* and simulation. We believe our approach is not restricted to regulated domains, and want to scale to other use cases.

References

1. Medical device software - Software life cycle processes - Amendment 1. IEC Standard 62304:2006/AMD 1:2015 (Jun 2015)
2. Medical devices - Part 1: Application of usability engineering to medical devices. IEC Standard 62366-1:2015 (Feb 2015)
3. Quality management systems - Requirements. ISO Standard 9001:2015 (Sep 2015)
4. Medical devices - Quality management systems - Requirements for regulatory purposes. ISO Standard 13485:2016 (Mar 2016)
5. Medical devices - Application of risk management to medical devices. ISO Standard 14971:2019 (Dec 2019)
6. van der Aalst, W.M.P.: Everything You Always Wanted to Know About Petri Nets, but Were Afraid to Ask. In: Hildebrandt, T., van Dongen, B.F., Röglinger, M., Mendling, J. (eds.) Business Process Management. pp. 3–9. Springer International Publishing (2019)
7. Christiansson, M.T., Rentzhog, O.: Lessons from the "bpo journey" in a public housing company: toward a strategy for bpo. Business Process Management Journal (2019)
8. Council of the European Union: Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. Directive (Jun 1993)
9. European Parliament, Council of the European Union: Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Regulation (Apr 2017)
10. de Giacomo, G., Dumas, M., Maggi, F.M., Montali, M.: Declarative process modeling in BPMN. In: LNCS. vol. 9097, pp. 84–100 (2015)
11. Henk, A., Fallmyr, T.: Navigating through institutional complexity: adoption of a process view in functional organizations. Business Process Management Journal (2019)
12. Mukkamala, R.R.: A Formal Model For Declarative Workflows. Phd, IT University of Copenhagen (2012)
13. Object Management Group: Business Process Model and Notation (BPMN) (2011), `http://www.omg.org/spec/BPMN/2.0`
14. Object Management Group: Case Management Model and Notation (CMMN). Tech. rep. (2016), `https://www.omg.org/spec/CMMN/1.1/PDF`
15. Russell, N., ter Hofstede, A.H.M., Edmond, D., van der Aalst, W.M.P.: Workflow data patterns: Identification, representation and tool support. In: Delcambre, L., Kop, C., Mayr, H.C., Mylopoulos, J., Pastor, O. (eds.) Conceptual Modeling – ER 2005. pp. 353–368. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
16. Senge, P.M.: The fifth discipline: The art and practice of the learning organization. Broadway Business (2006)
17. Swenson, K.D. (ed.): Mastering the Unpredictable: How Adaptive Case Management Will Revolutionize the Way That Knowledge Workers Get Things Done. Meghan-Kiffer Press (2010)
18. Swenson, K.D.: Innovative organizations act like systems, not machines. In: Fischer, L. (ed.) Empowering Knowledge Workers: New Ways to Leverage Case Management, pp. 31–42. Future Strategies Inc. (2014)
19. Tenschert, J.: Speech-Act-Based Adaptive Case Management. Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen (2019)
20. Tenschert, J., Lenz, R.: Agora - speech-act-based adaptive case management. In: BPM Demo track, CEUR Workshop Proceedings (2016)

Enhancing communication of complex services through process modelling practices: Insights from an Australian Higher Education Institution

Felipe Masana1[0000-0001-8600-8609], Wasana Bandara1[0000-0003-2785-7859], Joanne Allbutt2

[1] Queensland University of Technology, School of Information Systems, 2 George Street,
Brisbane, Australia
[felipe.masanasepulveda,w.bandara]@qut.edu.au
[2] Queensland University of Technology, Learning and Teaching in the Science and Engineering Faculty, 2 George Street,
Brisbane, Australia
j.allbutt@qut.edu.au

Abstract. This paper presents a case study from the Learning and Teaching (L&T) portfolio within the Science and Engineering Faculty (SEF) of the Queensland University of Technology (QUT). The case study outlines and examines how the portfolio used Business Process Management (BPM) principles and tools to manage the complex processes that support academic and student experience with a focus on process communication. Key processes were mapped using 'user-friendly' process modelling guidelines. An overall Process Architecture was derived to catalogue and manage the plethora of L&T processes. Capability building activities to develop 'process model-building' competencies within the L&T team and 'model-reading and understanding' capabilities within the model users (the academics of the Faculty) took place. An overall 'process of process management' within the SEF L&T team was set up to enable the current BPM efforts to continue as an efficient and sustainable practice. This case describes how process modelling can create effective communication for complex processes in a novice BPM environment. This experience can be applied in other faculties and industries to build up and maintain BPM efforts.

Keywords: Business Process Management, tertiary education, communication, process modelling, process architecture.

1 Introduction

Business Process Management (BPM) has emerged as a discipline that provides agility to organizations to adapt to market dynamics by tuning their processes through methodologies and implementation of new technologies [1, 2]. One of the many facets of the BPM paradigm is increasing transparency and collaboration across functional areas in an organization by increasing the awareness of business processes [3]. Process models visually represents how the organizations perform their operations [4].

Universities play a vital role in the knowledge economy, which contributes to increasing the pace of technical and scientific advancements, on the one hand, and obsolescence on the other [5]. According to Alter and Nelson [6] in the last decades, academics have increased the evolution of learning and teaching activities. As tertiary education entities are evolving and becoming more complex, BPM has the increasing potential to provide agility and transparency to their business processes. However, evidence of positive experience with BPM practices in the higher education is scarce.

This case study narrates process modelling efforts (which were supported with governance and methods) designed to effectively communicate complex and ever-changing processes to a variety of stakeholders.

2 The case organization, processes and challenges

The Science and Engineering Faculty (SEF)[1] of the Queensland University of Technology (QUT)[2] is one of the largest faculties in Australia, serving approximately 11,000 students across 10 academic schools. All academic management processes for SEF are provided and managed by the SEF Learning and Teaching team (here after referred to as L&T team). L&T leads the processes that support academic and student management, which are subject to strict compliance and quality assurance processes. Their complexity lies in their position between functional business requirements, changing compliance requirements triggered by policy reforms and the needs of human beings interacting with several systems. Moreover, multiple stakeholders need to be appropriately informed of the changes in the processes. As with many workplaces, the provision of timely updates on process requirements is challenging. Within the Higher Education industry, this seems to prove particularly problematic. Academic staff are required to interact with a range of processes in support of course delivery; for example, course information, timetabling needs, examination requirements. This information is often characterized as administrative and trivial in nature but when the requirements are not met, their absence impacts on the student experience and can lead to compliance breaches.

To contain and contextualize the relevant information, the L&T team has organized all processes managed within the portfolio into three major groups: ‘Unit Lifecycle’, ‘Course Lifecycle’ and ‘Student Journey’. The processes across these groups interact and interface with multiple stakeholders; students, academics, other staff within the faculty and across the university along with the L&T staff leading them. All the activities within these groups were developed to comply with the Manual of Policies and Procedures (M.O.P.P.)[3] of QUT and include content regarding managing learning and teaching processes, governance, committee organization and support, and student-facing administration and support processes. These processes are supported in an online environment composed of different systems. The L&T team aims to communicate about processes through the Digital Workplace as the source of definitive information.

---

[1] See https://www.qut.edu.au/science-engineering for further details.

[2] See https://www.qut.edu.au/ for further details.

[3] QUT's rules and policy resource that guides the operations of the University.

The Digital Workplace is an intranet solution where academics, and in fact all university staff, can find all the information regarding teaching and learning activities and news about the university. Once academics join their faculty community, they have access to the specific faculty processes through the Digital Workplace.

As policies, procedures, people and technology are constantly changing, the L&T team is challenged in managing effective communication with the SEF academic body. Academic staff within most large faculties are expected to manage many learning and teaching processes at different times during a teaching semester. Many academics may only refer to a particular process for a few days every year and they are expected to interact across a range of systems in support of processes such as timetabling or unit delivery and design. The prevalence of email to describe and manage processes means that often processes are inconsistent in their delivery. The management of these processes is also often regarded as administrative work and so is not prioritized; however, the work itself is key in delivering services. The processes themselves are policy and compliance driven and may have defined roles and responsibilities. In an environment crowded with information, academic staff find it difficult to locate relevant information that is clear and concise. This necessitates a number of enquiries that L&T staff manage on a case by case basis which in turn creates inefficiency. As a result, there is a high frequency of information becoming out of date.

The L&T team saw an opportunity to use Business Process Management (BPM) tools to facilitate their critical communication and transaction needs. Influenced and supported by the Business Process Management (BPM) group situated within the Faculty, the L&T team started to use process models (documented by using BPMN 2.0) to communicate processes and subsequent changes in the Digital Workplace. They attempted to classify and group their processes by using lifecycle models that would describe all processes based around unit management, course management and support for better student experiences, creating three lifecycles, namely; the ‘Student Journey’, ‘Course Lifecycle’ and ‘Unit Lifecycle’. To support the lifecycles (process groups) within the intranet solution, the staff wanted to be able to illustrate the processes visually, provide supporting text that was depicted as a call to action and reference links to assist in providing rich information to users. The aim was to allow users to look at a process’s information and understand quickly and easily what was required of them.

In light of the above, a program of work was developed to support the communication effort. Sponsored by the L&T team manager and supported by the SEF L&T senior executives, a group of BPM specialists were engaged[4]. Out of this need, the process management project described herein emerged.

---

[4] This BPM specialist group was formed by Masters students who were participating in the QUT Master of Business Process Management program as part of their work-integrated learning component, and a senior BPM academic (with years of BPM executive training experience and specialised BPM know-how, especially process modelling).

3 Description of the BPM approach and artefacts implemented

The ultimate goal of the BPM project was to enhance communication and apply the process models as a way to describe what was required of an individual academic, where their responsibility is situated within a process, and who else is involved in the process. Besides process models, several other artefacts were created to support the modelling efforts. This included SIPOC diagrams to help derive the process models, a Process Architecture (PA) to list and prioritize all the different processes, and a Process Architecture Landscape (PAL) to visually present the processes associated with the project scope, together with clear modelling conventions. Training instances and support videos for both the L&T team as model designers and academics as model consumers were also developed. The project team also designed an overall approach referred to here as a 'BPM Roadmap', for providing guidelines about future BPM activities within the L&T team, including guidelines for updating the BPM artefacts. This was complemented by the development of a Communications Plan, and a preliminary evaluation plan been developed.

The creation of all these BPM artefacts alongside the process models enabled capability building in the L&T team. A key part of the project was to ensure that the team can sustainably maintain the BPM activities. The project developed a governance structure that included roles and responsibilities, detailed methods to support BPM activities and a quality assurance cycle.

Each of the above-mentioned artefacts are briefly described next. Samples of the artefacts along with detailed explanations on their development and how they were used are provided as ancillary material. Readers are strongly encouraged to review the ancillary material, to better understand and appreciate how these different parts each played a role and came together. The ancillary materials are available at: https://drive.google.com/file/d/1v-SDSzLm0gnglRWC9T7O0zKd9FQXOd7U/view?usp=sharing

SIPOC diagrams. The Supplier, Inputs, Process, Outputs and Customer diagram or SIPOC are often used as a preparatory tool for process modelling efforts. This tool was originally developed under the umbrella of the Six Sigma methodology [7]. It enables the practitioner to understand a process; its stakeholders, main triggers and outcomes, and aims to capture the main concepts of a process at a high-level, so it can be used to define the overall scope and 'skeleton' of the processes to be considered. In this project, SIPOC diagrams were used (across all the processes being considered) to derive the key information needed to build the overall Process Architecture (PA). Initially, SIPOC diagrams were applied to scan the 3 life-cycles' structures to identify which elements listed in each group were processes. This step allowed the team to discriminate between items as either part of a process, service or tasks, and reorganise the categories within each lifecycle accordingly. A sample is provided in the Ancillary Material, Part B.

Process Architecture (PA). A PA plays a major role, especially when many processes are looked at for modelling and/or improvements, and is an artefact derived by many organisations around the world when undergoing BPM initiatives [8]. A PA encapsulates all relevant processes for an organisation in one structure and identifies the relationships between them. The PA brings transparency to the organisation and a better understanding of the relationships between processes. It is important to recognise that in most of the cases, companies build their PA from the business (internal) point of view, considering external stakeholders as black pools where the activities they perform are not relevant in the design. However, as the purpose of this project was to communicate the processes to SEF academics, it was important to capture the activities relevant to the academic staff using the models. Therefore, the nature of the SEF L&T process architecture is different to the traditional internal approach as it is based on the processes behind the services that the L&T team supports in SEF. The PA was developed taking the collection of individual SIPOCs as an input to construct this artefact. A sample of the PA is provided in the Ancillary Material, Part C.

Process Architecture Landscape (PAL). The PAL is a graphical representation of the PA. It was designed to complement the PA, to make it user-friendly. There are no conventions for the design of this kind of process landscape, but the same principles related to the design of a PA apply here as well. The PAL is based on the services that SEF L&T support and provide and that requires communication to the SEF academic body. Most PALs follow the concept of Michael Porter's value chain [9], separating supporting processes and core process which are those that add value to the customer. As the PA of the L&T team is a service-based PA, all the relevant processes are core processes, and supporting processes are not part of this PA. Therefore, they were not considered in the forming of this PAL. The development of the PAL created a single structure making information easy to access and provide a visual representation that showed the relationship between processes. A sample is provided in the Ancillary Material, Part D.

Modelling conventions. The modelling conventions were based on principles of ease of communication, simplicity and consistency. The development of modelling conventions become important as the project progressed. The BPMN 2.0 was used to support the conventions and were developed based on the Seven Process Modelling Guidelines (7PMG) presented by Mendling, Reijers, van der Aalst [10] in favour of consistency and simplicity. The conventions provide modelling governance, as they not only standardise the way to model but also outline roles and responsibilities for quality checking the models. One milestone during the preparation of these modelling guidelines was determining the 'right' set of BPMN objects to use when modelling. Only a limited set of the core BPMN objects was selected. The idea was to keep the process models as simple as possible. This also supported modelling consistency as the process models are developed by different people whose modelling skills and notation knowledge differ. The original conventions are provided in Ancillary Material, Part E.

Modelling training. A number of training sessions were held to support the modelling capabilities of the L&T, including work-shops about applying modelling guidelines, conducting process identification to list the processes using the SIPOC diagram, drop-in sessions for L&T novice modelers, and a number of tutorial videos to support the training of the team. Further details provided in Ancillary Material, Part F.

BPMN process models for SEF L&T. At the start of this project, the L&T team had already begun using BPMN models to communicate processes to their target audience. BPMN is the standard process modelling language for process modelling, and the L&T team already had some exposure to it prior to this project getting started. During the project, the process models were mainly developed by the L&T team, the external BPM team assisted them in developing more complex models and by providing feedback during drop-in sessions about the models they were working on. The entire BPM project took place in a highly team-spirited environment. It was driven by the L&T team members who collectively 'owned and supported' the BPM efforts, strengthened by the team's genuine desire to address their day to day challenges and provide better services. Further details about the dynamics pertaining to the process model development, such as; roles and responsibilities, the nomination of models and modellers, and quality assurance cycles were also setup.

BPM Roadmap in SEF L&T. A detailed 'BPM Roadmap' was created to form a comprehensive guideline about all the Business Process Management (BPM) practices implemented in SEF L&T within this project. The roadmap is a knowledge tool for the sustainable development and maintenance of the BPM capabilities adopted by SEF L&T. The document consisted of two parts. The first one provided a governance structure (roles and responsibilities) and the second one the overall "process of process management" within the SEF L&T team. A model of conducting BPM process in provided in Ancillary Material, Part G.

Communication Plan. The communication plan plays a critical role in the successful deployment of any BPM project. Prior research has shown that a healthy communication plan can dissolve resistance, contribute to buy-in, and gain support, from key stakeholder groups at different levels [11]. The communication plan consisted of several diverse artefacts in written and multi-media form. A range of short video artefacts (that only lasted one minute each) targeting the academics to build basic awareness of where to find the details of the SEF L&T processes were amongst the more popular artefacts. It first instructed the L&T professionals to share the information internally with the team through different channels. Secondly, it enabled the L&T team to effectively communicate the processes underlying their services through some resources like tutorial videos in the Digital Workplace. Thirdly, it helped to promote the use of the intranet (i.e. QUT's Digital Workplace) as the 'source of truth' for the faculty processes.

Evaluation Plan. A preliminary evaluation plan was designed and executed (see the Ancillary Material for further details - including evaluation-results) to gain insights about; (i) how easy is it for users to find and locate process information via the digital workplace, (ii) if academics are able to understand the process more easily, and (iii) how the supporting text and other useful links provide assistance to the user? The results confirmed that the academics were struggling to navigate through the Digital Workplace to find the processes (which was more to do with the structure and searchability of the Digital Workplace, which was out of scope for this project). Nevertheless, the respondents agreed that the process models were a 'valuable resource' to understand the processes and that the models supported with overcoming administrative confusions. More specifically, the respondents were able to understand the flow of the process relatively easily and mentioned that it is quite simple and straight forward. Minor difficulties were admitted regarding the lanes and coloring of the models but after a brief introduction, they were comfortable with the models. Overall, the key goal of this project was achieved as per the evaluation results.

4 Benefits and lessons learnt

The L&T team required more efficient ways to communicate the relevant processes to the academic staff and decided to use business process modelling as a means to better communicate their processes. Though the L&T team initially lacked the required BPM capacities, this initiative equipped them with BPM tools, methods and governance techniques to enable them to become effective BPM practitioners within the business.

Given the number of BPM artefacts introduced to the L&T team, the main challenge was to create mechanisms that allowed the L&T team to maintain the artefacts by themselves after the life-span of this project. The modelling training (including tutorial videos) and the governance set-up were necessary but not enough. Hence, the BPM tools were designed in an integrated manner - to have common elements that can link to one another. For example, the SIPOC diagrams contain relevant questions to feed into sections of the PA, which had coloured columns in the stakeholder's section that match the swim-lanes of process models that were colour-coded for enhanced understanding and communication of specific roles.

The communication plan was a critical success factor, without which all the efforts would have crumbled – as the academics would not be aware of the many new resources, they have access to. External contextual factors (such as the generic structure of the digital workplace – which is governed by other areas of the university) played a role on the effectiveness and efficiency, demonstrating how BPM project members needs to reach outside the project scope to get things to work and/or work within constraints. Early evaluations pointed that the efforts are going in the right direction, with evidence that the process models were easy to understand, supporting the main goal of the project. The evaluation results reported here was preliminary, and periodic evaluations covering a wide range of processes and a larger sample are recommended.

More broadly, this case also demonstrates (with vivid explanations) some interesting aspects of BPM in general and more specifically process modelling practices. It shares a story that depicts the value of process modelling within the Higher education sector. BPM examples within the Higher education sector are rare yet it is a sector (globally) that can benefit significantly with BPM practices given the many challenges (such as funding cuts, evolving demands, and the need for better operational processes).

Furthermore, process modelling is often seen as a step towards other key BPM activities. Yet this case described a scenario where modelling was used as the main goal of the project; to simplify the communication of complex and important processes. Here process modelling contributed significantly towards information management and knowledge management goals. The case also depicted how a range of other artefacts including an overall 'process of process modelling' (referred to as the BPM Roadmap) crystallized and strengthened the process modelling efforts. It shared the mannerisms on how capability building, and appropriate governance structures were put in place to deploy sustainable BPM practices. The way process models themselves were deployed to articulate the meta-approaches set by the program of work as in the BPM roadmap is also an interesting observation of this case study.

To conclude, BPM capabilities were well developed within the L&T team along this project, enabling them to keep progressing in their communication efforts through process models in a sustainable way without relying on external BPM expertise. We also believe that the work described in this case can be easily adapted to other contexts to foster process modelling and BPM in other industries.

References

1. Fisher DM. The business process maturity model: a practical approach for identifying opportunities for optimization. Business Process Trends. 2004;9(4):11-5.
2. Rosemann M, vom Brocke J. The Six Core Elements of Business Process Management. In: vom Brocke J, Rosemann M, editors. Handbook on Business Process Management 1: Introduction, Methods, and Information Systems. Berlin, Heidelberg: Springer Berlin Heidelberg; 2015. p. 105-22.
3. McCormack K, Johnson WC. Business process orientation gaining the e-business competitive advantage. Johnson WC, editor. Boca Raton, FL: St. Lucie Press; 2001.
4. Curtis B, Kellner MI. Jim Over: Process Modelling. Communications of the ACM. 1992;35(9):75-90.
5. Powell WW, Snellman K. The knowledge economy. Annu Rev Sociol. 2004;30:199-220.
6. Alters BJ, Nelson CE. Perspective: Teaching evolution in higher education. Evolution. 2002;56(10):1891-901.
7. Marques PA, Requeijo JG, editors. SIPOC: A Six Sigma tool helping on ISO 9000 quality management systems. XIII Congreso de Ingeniería de Organización; 2009.
8. Harmon P. The state of business process management 2018 [Available from: https://www.bptrends.com/2018-state-of-business-process-management-lp/.
9. Porter ME. Value chain. The Value Chain and Competitive advantage: creating and sustaining superior performance. 1985.
10. Mendling J, Reijers HA, van der Aalst WM. Seven process modeling guidelines (7PMG). Information and Software Technology. 2010;52(2):127-36.
11. Chong S, Rosemann M, editors. Towards a Framework for BPM Communication. 7th itAIS Conference Naples; 2010.