Automated Business Process Discovery and Analysis for the International Higher Education Industry

Juan Gonzalez-Dominguez[1(✉)] and Peter Busch[2]

[1] AlphaSys Pty Ltd., Sydney, NSW 2000, Australia
gonzalezd.juan@gmail.com
[2] Department of Computing, Macquarie University,
Marsfield, NSW 2109, Australia
peter.busch@mq.edu.au

Abstract. The international education sector in Australia expects a rise in demand for higher education programs in the future. This increase drives universities to monitor and continuously improve their business processes. Automated Business Process Discovery or Process Mining helps organisations create a business process graphical representation or process model, leveraging data from their information systems. The lack of a business process model and inefficiency in harnessing process data deters fact-based process improvement decision-making, impacting a company's operations. Within the international education context, Process Mining can help discover, analyze, manage and improve processes efficiently and deal with increasing market demand of international student enrolments. Here we have used the Process Mining Methodology Framework for Discovery Analysis focusing on Control-flow and Case Data perspectives. We highlight challenges encountered in process data extraction and preparation as well as performance analysis to identify potential bottlenecks.

Keywords: Automated business process discovery · Process mining
Business process · Process modeling

1 Introduction

Business Process Management (BPM) uses different techniques to achieve organisational goals through well established, monitored and continuously improving processes [1]. One of these techniques is *Automated Business Process Discovery* or *Process Mining*. Process mining discovers business process representations (or process models) using data mining techniques [2]. Furthermore, process mining can help identify process deviances and support business process improvement initiatives [3]. The data for process mining is recorded in Process-Aware Information Systems (PAIS) owned by organisations to perform daily activities [4]. The data is extracted in the form of Event Logs [4]. Within the process mining body of knowledge, the Process Mining Manifesto covers guiding principles and challenges that need to be addressed [3]. Process mining could also be implemented to achieve different goals, for example,

basic process discovery, case data analysis or organisational analysis. In-depth analysis can be performed to achieve conformance and performance analysis [5]. Process mining approaches can be applied to different industries, for instance, government and financial institutions [5, 6]. The international education industry in Australia, more specifically the higher education sector could harness process mining approaches to cope with an increase in future enrolment demand predicted by *Deloitte Access Economics* [7] and continuously improve processes. Here we present knowledge obtained by applying process mining techniques to discover a sample *International Students Admission Process*.

2   Background

Business processes are a series of interdependent activities carried out at different levels. Processes can run internally in an organisation, either inside a department or among various departments, or externally through different organisations [8]. Several authors have approached the study of business processes from different perspectives. Papazoglou and Ribbers [9] describe business processes as part of the technical and management foundations for e-business solutions. Hammer [10] notes processes and process management as a whole discipline supports transformation strategies and organisational operations. Hung [11] notes process alignment and people involvement are variables impacting the successful implementation of business process management as a competitive advantage. Business process initiatives allow organisations to develop or implement methodologies to maximise resources by business process continuous improvement [12].

The use of data mining methodologies for automated business process discovery (ABPD) has increased in research [2]. The main goal is to leverage existing data recorded at the execution of activities within process-aware information systems (PAIS) to create a business process graphical representation or business process model [2]. Automated business process discovery is also known as Process Mining. In this area, Professor Will van der Aalst is recognised as one of the major contributors to the process mining body of knowledge. His publications cover a range of process mining areas, e.g. workflow mining [13], obtaining data [4], differing techniques to deal with event logs [14] and the Process Mining Manifesto [3].

Mans et al. [15] and Jutten [16] explain there is considerable academic literature about process mining, but research about adoption is infrequently found, for organisations do not foresee business benefits from this technique. There are some documented case studies, e.g. a provincial office of the Dutch National Public Works Department provided different standpoints for the use of process mining for process discovery and performance [6], and business alignment or conformance checking [17]. In De Weerdt et al. [5] a detailed case study in a financial institution shows the capabilities of process mining in real-life environments, but the authors conclude the need to focus on practical applicability to improve process mining techniques. But what is process mining?

3   Process Mining Overview

Since the publication of the value chain by Porter [18] organisations have shifted from function-centric to process-centric operations. Business processes are the collection of interdependent activities performed in an established order to reach organisational goals [8]. Several measurements can be attached to business processes, for instance, time, cost, performer and quality. Furthermore, business processes' scope can range from processes running within an individual business unit to processes running across different organisations [9], but they typically require some form of management.

3.1   Business Process Management

Business Process Management (BPM) is considered a mature discipline that "supports business processes using methods, techniques, and software to design, enact, control and analyse operational processes involving humans, organisations, applications, documents and other sources of information" [1]. BPM uses techniques such as business process modelling to help create representations of business processes (also known as business process models), for better visualization and to avoid misunderstandings [19]. BPM can use tools and techniques from different approaches. Methodologies based on creativity like Business Process Re-engineering (BPR), internal or external benchmarking or statistical analysis such as Six Sigma [20] are also relevant. For example van der Aalst et al. [21] highlight how Six Sigma methodologies for improving a business process by "statistically quantifying process performance" [21] could be not accurate as the data would be manually collected, as well as prove expensive and time-consuming. *Automated business process discovery* better known as Process Mining can overcome this issue.

3.2   Process Mining

Process mining is a technique depicting business process models through patterns in big data sets comprising data recorded at process run time in information systems; it originates from the adjustment and growth of data mining techniques applied to the field of business processes [2]. "Discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems" [3] is process mining's goal. Process mining makes a significant distinction between real processes and assumed processes. The former relates to the creation of process models based on tasks recorded in information systems – i.e. actions that really happened. The latter refers to the creation of process models based on human observations, expectations and assumptions [22, 23].

In 2009 an IEEE Process Mining Task Force was created, and in 2011 it released the final version of the Process Mining Manifesto [24]. The manifesto describes six guiding principles (GP) that explain the importance of data (event logs), extraction, how to treat the depicted process model as an abstraction of reality and process mining as a continuous activity [3]. These principles are: GP1 - *Event data should be treated as First-Class Citizens* [24]. GP2 - *Log Extraction should be driven by questions* [24]. GP3 - *Concurrency, choice and other basic control-flow constructs should be*

supported [24]. GP4 - *Event should be related to model elements* [24]. GP5 - *Models should be treated as a purposeful abstraction of reality* [24]. GP6 - *Process Mining should be a continuous process* [24].

3.3   Event Logs

Process data plays the most important role in process mining. This data is often extracted and converted into event logs, representing the starting point for process mining [24]. For process mining implementations as discussed in GP4, event logs shall only contain event data related to the process under analysis [4]. An event log can be deconstructed to the following elements: *Cases*, which represent a process instance, therefore the event log would contain several cases. *Events*, every case is formed by events, these could be understood as a task in the process and every event is part of one and only one case. Event *attributes* - any extra information related to the process. Common attributes are activity, timestamp, cost and resources [4]. Figure 1 represents an example event log:

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	Register request	Pete	50	...
	35654424	31-12-2010:10.06	Examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	Check ticket	Mike	100	...
	35654426	06-01-2011:11.18	Decide	Sara	200	...
	35654427	07-01-2011:14.24	Reject request	Pete	200	...

Fig. 1.   A sample event log (Source: van der Aalst, 2011: 99)

Challenges for event logs are covered by the Process Mining Manifesto. Ly et al. [25] noted four of them: incorrect/incomplete log data, data contribution through parallel branches, infrequent traces (planned exceptions) and ad-hoc contributed data (unplanned exceptions. Just as an aside, GP1 establishes the level of maturity for an event log. The maturity level considers four criteria for measurement Trustworthiness, Completeness, Well-defined Semantics and Security.

3.4   Types of Process Mining

Process Mining may be divided into three types. *Discovery* - here event logs are used to create a graphical representation of the business process. This graphical representation is also known as *process model*. The main output for *discovery* is a business process model. Discovery is the main application of process mining in organisations. *Conformance* - here an event log is compared to an existing business process model. The aim is to corroborate if the process recorded in the event log (real activities) runs as expected by the preliminary model. The primary output in *conformance* is process model diagnosis. *Enhancement* - the goal is to improve the current process model by using information in the event logs. In contrast with *conformance*, instead of comparing with

established measurements, in *enhancement* the idea is to change and improve the process model. The primary output in *enhancement* is a new business process model [3].

3.5   Process Mining Methodology Framework

The process mining follows methodologies according to the aim of the implementation, for instance, Bozkaya et al. [26] propose five phases for process diagnostic with process mining, log preparation, log inspection, control flow analysis, performance analysis and role analysis. Albeit this methodology has been applied to government and financial organisations [26, 27], it does not cover all the process mining perspectives.

De Weerdt et al. [5] developed the Process Mining Methodology Framework (PMMF). This methodology can be considered more flexible, including an analysis phase where different perspectives can be considered depending on the process mining aim. PMMF comprises five phases. The first phase is *preparation*. In this phase data is extracted from information systems and prepared as event logs. Next *extraction* takes into consideration the scope and timeframe that best fits the process mining implementation. An excess of information can result in extraction of activities not belonging to process scope. A timeframe not covering possible seasonal behavior, will not be represented in the process discovery [5]. The next phase - *exploration* is an iterative activity where scope and timeframe are analysed and re-configured based on multiple process visualisations with different algorithms. This phase holds a relationship with the activity of data extraction as exploratory outcomes can show the need for more data or different time frames [5]. In the *perspectivization* phase, the perspective the analysis will pursue is determined and if it is necessary to construct different event logs for each perspective [5]. The next phase is *analysis* and could be divided in two - *Discovery analysis* and *in-depth analysis*. Basic *discovery analysis* is where various dimensions recorded in event logs are analysed. *Control-flow analysis* refers to the sequence of activities. *Organisational analysis* relates to the users' relations and social networks created by the process run time. *Case data analysis* refers to other attributes included, for example cost and time [5]. With *in-depth analysis* more detailed examination takes place. *Conformance analysis* would help identify activity deviations from the expected process model based on the mined process model. *Performance analysis* supports organisations to find insights related to the process, for instance process execution times and waiting periods [5]. The final phase is *results* - in this phase all findings during the analysis phase are considered for decision making and continuous business process improvement [5]. Let us now examine a working scenario to place the above in to context.

4   Process Mining Knowledge Generation

With growing demand from international students, a university should have the continuous capability to monitor and improve its *International Students Admissions Process*. The importance of this process relies on the ability of a university to (a) ensure candidates' qualifications and English proficiency are appropriate, (b) comply with the Australian Education Legislation of providing written agreement [28], (c) facilitate the transition between the application stage and the student enrolment stages and

(d) articulate market development strategies for recruitment. To manage the process, a university may use an application system recording all daily transactions regarding admission activities and helping to keep one source of information. In general, the process is triggered when an application for study is received. The application can be made online through the application system or in hard copy (paper application). Data entry clerks can review the application and create a new student record in the student management system. If the application is paper-based, student details are entered in two systems. The application can then be turned over for assessment to the admissions assistant or admissions officers.

During assessment, admission officers confirm if the documentation is complete, if it complies with university policies and if the student meets the entry requirements for the desired program. Furthermore, faculty permission and recognition of prior learning advice are sought if necessary. After assessment, admissions officers either reject the application, ask for more information, or issue an offer letter with all the information about the program, tuition fees, commencement dates, insurance and other information requested by Australian legislation [28]. The offer letter can be a full offer letter, full offer with conditions or conditional. The next task is acceptance. A significant time gap can be expected between issuing the offer letter and acceptance. The acceptance is processed by the admissions officer with assistance in coordination from the finance department. When the finance department confirms the student's payment, a confirmation of enrolment (CoE) is issued and sent to the student with a welcome message and instructions for program enrollment. At this point, the *International Students Admissions Process* can be considered finished. A theoretical example of the *International Students Admission Process* model can be *discovered* by extracting an event log from an application system. Remember what we seek here is process discovery, following the PMMF for control flow and case analysis.

4.1 Event Log Extraction

A system records all the activities related to the admissions process including application status, any action performed in the system, the name of the person executing activities, documents, a student's demographic information and more. The system's outputs are reports in comma separated value format extracted by the reporting team. Every day an internal student report with all students in the system is extracted, this report only represents a static picture of a student's status at the moment of reporting. The system is a stand-alone solution with no implementation of an Application Programming Interface (API) for extraction of student information. Due to the aforementioned reason, extraction must be made manually from each student activity log, taking into consideration the scope and timeframe.

As a working example, we focus only on complete process instances [5]. Based on a generated daily report, a sample of applications with "Acceptance", "Withdrawn" and "Not Qualified" was created. These statuses are considered termination points for the admissions process. A sample calculation was made with 95% confidence, 5% of margin error, 50% distribution and a finite population correction applied [29]. Randomly generated samples by status were: 198 over 825 "not qualified", 243 over 1306 "withdrawn" and 294 over 2486 "acceptance" for a total of 735 cases.

4.2   Event Log Preparation

After extraction of data, an event log has to be explored to understand the data that is on it and verify if it will help to achieve the project goals [30]. That is to say, either eliminate data not useful for analysis or include data from other sources to complement the analysis [30]. Furthermore, any additional transformation such as de-identification of personal data was applied in this phase.

The event log extracted was enhanced with extra attributes that will help with the *Discovery Analysis for Case Perspective.* Data such as timestamp, person starting the event (*originator*), country of citizenship, application made by an agent or directly, and type of program, are some of the attributes added to the event log. An event log is stored in .csv format. To be used, it should be transformed to a MXML or XES (eXtensible Event Stream) format. XES was chosen as being the most current standard, is also more flexible to handle extra attributes in the event log and a plug-in is already available in the ProM [31].

4.3   Automated Business Process Discovery

To create a business process model based on the data provided by a sample event log, process mining makes use of algorithms, mainly data mining clustering algorithms [32]. For this experiment a heuristic miner was used. The algorithm analyses an event' "direct dependency, concurrency and not-directly-connectedness" [33] frequency, can handle short loops and is recommended for semi-structured processes [33]. The heuristic algorithm supports control-flow constructors like sequence, parallelism, choices, loops, invisible tasks and in some degree non-free-choice [34]. It also provides a middle point between underfitting and overfitting process models allowing fitness, simplicity, precision and generalisation criteria mentioned in the *Process Mining Challenge VI* [34]. In contrast, an α-algorithm requires structured processes, complete logs with no noise and no loops [33]. Albeit there have been some improvements in this algorithm such as α+ and α++, they still lack flexibility [33]. According to De Weerdt, De Backer, Vanthienen and Baesens [33] a Fuzzy Algorithm is preferred for event logs with a high level of noise and unstructured processes.

A generated event log for *discovery* contained 726 cases, 7,401 events, and 26 classes (status). By using a heuristic algorithm, there was potential to mine the *International Student Admission Process*. Figure 2 depicts the discovered process model. The discovered process shows certain tasks with higher frequency such as new application (starting task), submitted, pending assessment, assessing and application edited. Notable are also a start activity and multiple end activities, as well as several loops found in the event log.

The mined process shows the most common flow - new application - submitted - pending assessment - assessing – qualified - acceptance pre-processing - acceptance, achieving the fitness criteria. It also depicts possible paths the process can follow achieving precision. Note that simplicity and generalisation can be improved. One option can be reducing the process scope or by creating concept hierarchies for assessing or agent assignment activities.

The process illustrated in Fig. 2 comprises 26 different possible tasks performed during execution. The top five activities with more occurrences are assessing,

Fig. 2. A sample data-mined *International Student Admission Process*

application edited, submitted, pending assessment, and incomplete application with 1152, 909, 727, 711, 689 occurrences respectively. The mined process considers "new application" as the main starting activity (94.5% of the instances). However, some cases start with "submitted" and "incomplete application" both 5.5% of the cases. Regarding the end events, eight different events were found - acceptance, withdrawn, not qualified making 97.38% of the total cases and identified from the process scope. Application edited, course change, agent assignment – request by student, agent assignment – administrative error are the other end activities shown in the discovered process model. On average a process case can change status 9 times and include 10 activities from beginning to end. The maximum number of activities is 29 and the minimum is 2. The control flow analysis discovered 476 different process paths. Figure 3 included the ten most common paths:
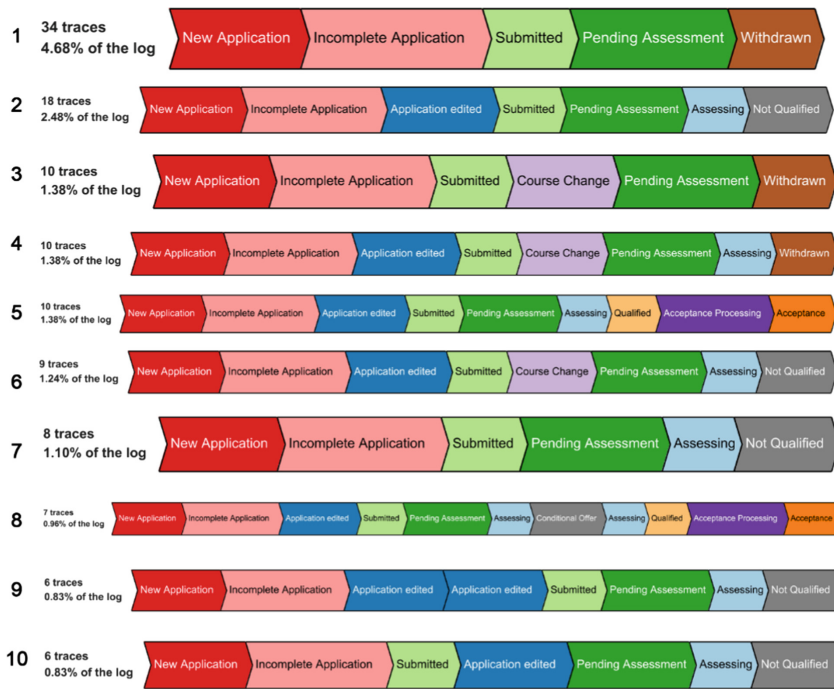

Fig. 3. Ten most common traces in the *International Student Admission Process*

These 10 paths represent only 16.26% of the entire paths found in the event log. Many of them end either with withdrawn or not qualified. Paths 1 and 3 were withdrawn without any assessment, which can suggest these applications do not fulfil the department's application processing policies. The mined process in Fig. 2 shows two activities that do not accord with the assumed process flow. "Application edited" is an activity that can happen in any part of the process but it happens mainly before the application is submitted. In the process model, this activity leads to "assessing" -

skipping two activities "submitted" and "pending assessment". The other activity is "not qualified" - considered as an ending activity. However one case in the event log continues with "agent assignment request by student" which allows the algorithm to link the activities - this case can be an exception within the process.

The goal of the *performance analysis* is to identify activities that impact the process flow negatively. Performance analysis within ProM was achieved by mining a Petri Net. To include all the possible task and the time calculations, the process model was overfitted for the 726 cases. A performance and conformance checking plugin was then applied. Table 1 summarises the main events and processing times per sequence pattern

Table 1. Summary events transitions processing times in days

Event from	Event to	Avg.	St. Dev.
New application	Submitted	3.2	18.09
Submitted	Pending assessment	0.9	1.01
Pending assessment	Assessing	4.2	7.16
Assessing	Assessing - more info required	1.7	7.42
Assessing	Assessing - referred for RPL	9.5	12.37
Assessing	Assessing - referred to faculty	6.3	12.92
Assessing - referred to faculty	Assessing - returned from faculty	21.5	14.94
Assessing - referred for RPL	Assessing - returned from RPL Assessment	6.8	10.64
Assessing	Qualified	11.3	18.58
Assessing - more info required	Qualified	23.0	20.81
Assessing - returned from faculty	Qualified	8.3	9.32
Assessing - returned from RPL assessment	Qualified	9.7	10.25
Assessing	Conditional offer	8.6	14.18
Assessing - more info required	Conditional offer	18.5	16.67
Assessing - returned from faculty	Conditional offer	2.3	1.78
Assessing - returned from RPL assessment	Conditional offer	6.2	4.36
Assessing	Not qualified	17.2	24.32
Assessing - more info required	Not qualified	32.8	24.72
Assessing - returned from faculty	Not qualified	11.3	22.88
Submitted	Withdrawn	20.8	27.10
Conditional offer	Withdrawn	32.9	26.75
Full offer with conditions	Withdrawn	82.0	27.80
Qualified	Withdrawn	33.2	27.88
Assessing	Withdrawn	26.2	29.28
Pending assessment	Withdrawn	20.2	26.70
Conditional offer	Qualified	21.2	18.36
Qualified	Acceptance processing	19.0	13.92
Acceptance processing	Acceptance	1.0	3.28
Full offer with conditions	Conditional acceptance processing	20.0	18.61
Conditional acceptance processing	Conditional acceptance	1.4	4.08

in days. The table shows average times and standard deviation. The latter is only indicative of the occurrences distribution (the range where most of the observations fall). It can be assumed the data has a positively skewed distribution, as the goal is to maintain the response times as low as possible. ProM does not provide median and quartile data to make a more accurate evaluation [35].

Overall results for the complete sample data set mentioned an average of 1.06 months to complete a case with a standard deviation of 1.02 months. The maximum time of processing is 9.61 months and the observation period was 11.69 months. Applications can be reviewed and entered into the student management system in less than one day and then hold a pending assessment status for 4.2 days. If advice is requested from faculties or recognition of prior learning, this could take 21.5 and 6.8 days respectively. Issuing an offer or rejection letters when documentation is complete and no further faculty or recognition of prior learning is sought, could be done in an average of 12.4 days (including qualified, conditional offers and not qualified status).

It was determined four activities increased processing time - *pending assessment, refer to faculty*, *refer for RPL and assessment* and *more information required. Pending assessment* relies completely on the admissions department and has a waiting time of 4.2 days. The latter three activities are delegated to external actors, meaning the admissions department cannot control the waiting time. Waiting time for faculty response is determined to 21.5 days. When the application required more information from the student, a total of 24.8 days to reply could elapse. The RPL team replied in 6.8 days on average.

5   Discussion

Process mining applications have been documented in academic research in different domains, such as government and financial services [5, 6]. Here process mining approaches were used in a university domain to obtain a theoretical process model for the *International Students Admission Process*. Process mining leverages data recorded in an application system to yield a business process model and information about the process performance. Some challenges and limitations were observed. The mined process model failed to achieve the *simplicity* and *generalisation* quality criteria as application processes can be very flexible and activities within the flow can be repeated at any given time, creating short loops. Due to this reason, the heuristic algorithm mapped all possible paths determined by the dependencies and occurrences, resulting in a complex process model with multiple flows – i.e. a process not simple to explain and general enough to cover all possible process flows.

The discovery analysis also detected some process instances that did not conform to what was assumed. *Starting* and *ending* events deviated from expected events. These process instances can be considered exceptions, but further analysis can be pursued to determine the original cause and authorizations. During data exploration, it was realised manual activities or activities not performed within the application system would not be included in the analysis. For instance, lead time between paper applications received and data entry into the system and issue escalation for special approvals, is occasionally not documented in the system. Some specific flows ended in "withdrawn" applications

without any assessment. Perhaps causes for these cases could be reviewed and evaluated if the system can help reduce these occurrences or put in place business rules, business process or policies to reduce the workload generated from these instances.

Performance analysis identified potential activities that could be reviewed further. Most of these involve collaboration with external entities. More detailed assessment of these activities is recommended to address the reduction in processing time in a collaborative fashion. Regarding *assessing – more information required*, this is an activity relating directly to the student submitting incomplete applications. Understanding the root causes can help to elaborate solutions to solve potential issues, for instance information accessibility. The performance analysis also provided a general overview of processing times. Although processing times appeared within normal processing times, goals and strategies need to be put in place to consider the increasing demand leading to bottleneck activities described previously. Further work is required to achieve additional meaningful insights about the process. Mining specific flows separately, such as flows ending in acceptance, not qualified or withdrawn could offer more detailed information about performance and an applicants' characteristics to implement business process improvement or data mining initiatives. Finally, ProM is a powerful tool for process mining; it contains plugins providing functionality for process discovery, analysis and transformation [36]. While the software includes comprehensible user interfaces, it requires a certain type of process, data types, document formats, data mining and process mining understanding; hence non-expert users could find difficulties in using this tool. ProM could improve workspace management, particularly how results can be stored to avoid repetition of tasks for result gathering.

6   Conclusion

Automated Business Process Discovery or Process Mining is a technique for business process discovery, analysis and improvement [24]. This project implemented process mining approaches for control-flow and performance analysis [5] via a theoretical enrollment case. A business process model was discovered achieving fitness and precision quality criteria [3]. Non-standard activities for process start and end were also uncovered. Performance analysis identified process bottlenecks in four tasks - *refer to faculty*, *refer for RPL, assessing – more information required* and *pending assessment*. Further analysis will almost certainly produce better process insights. We recommend detailed process mining discovery and performance be applied to detailed flows. Other case perspective variables can be included for analysis as well.

References

1. van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M.: Business process management: a survey. In: van der Aalst, W.M.P., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44895-0_1
2. Tiwari, A., Turner, C.J., Majeed, B.: A review of business process mining: state-of-the-art and future trends. Bus. Process Manag. J. **14**, 5–22 (2008)

 3. van der Aalst, W., et al.: Process mining manifesto. In: IEEE Task Force on Process Mining, pp. 1–19 (2012)
 4. van der Aalst, W.M.P.: Getting the data. In: van der Aalst, W.M.P. (ed.) Process Mining: Discovery, Conformance and Enhancement of Business Processes, pp. 95–123. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19345-3_4
 5. De Weerdt, J., Schupp, A., Vanderloock, A., Baesens, B.: Process Mining for the multi-faceted analysis of business processes—a case study in a financial services organization. Comput. Ind. **64**, 57–67 (2013)
 6. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business process mining: an industrial application. Inf. Syst. **32**, 713–732 (2007)
 7. Deloitte Access Economics, Growth and Opportunity in Australian International Education, pp. 1–95. Austrade, Melbourne (2015)
 8. Davenport, T.H., Short, J.E.: The new industrial engineering: information technology and business process redesign. Sloan Manag. Rev. **31**, 11 (1990)
 9. Papazoglou, M.P., Ribbers, P.: e-Business: Organizational and Technical Foundations. Wiley, Chichester (2006)
10. Hammer, M.: What is business process management? In: Vom Brocke, J., Rosemann, M. (eds.) Handbook on Business Process Management, pp. 3–16. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-00416-2_1
11. Hung, R.Y.-Y.: Business process management as competitive advantage: a review and empirical study. Total Qual. Manag. Bus. Excell. **17**, 21–40 (2006)
12. Harrington, H.J.: Business Process Improvement: The Breakthrough Strategy for Total Quality, Productivity, and Competitiveness. McGraw-Hill, New York (1991)
13. van der Aalst, W.M.P., Weijters, T., Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Trans. Knowl. Data Eng. **16**, 1128–1142 (2004)
14. van der Aalst, W.M.P.: Extracting event data from databases to unleash process mining. In: vom Brocke, J., Schmiedel, T. (eds.) BPM - Driving Innovation in a Digital World, pp. 105–128. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14430-6_8
15. Mans, R., Reijers, H., Berends, H., Bandara, W., Rogier, P.: Business process mining success. In: Proceedings of the 21st European Conference on Information Systems, AIS Electronic Library (AISeL), Utrecht University, The Netherlands (2013)
16. Jutten, M.G.: The Fit Between Business Processes and Process Mining Related Activities: A Process Mining Success Model (2015)
17. van der Aalst, W.: Business alignment: using process mining as a tool for Delta analysis and conformance testing. Requir. Eng. **10**, 198–211 (2005)
18. Porter, M.E.: The Value Chain and Competitive Advantage (Extract). Free Press, New York (1985)
19. Burattin, A.: Introduction to business processes, BPM, and BPM systems. In: Burattin, A. (ed.) Process Mining Techniques in Business Environments: Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining. LNBIP, vol. 207, pp. 11–21. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-17482-2_2
20. Siha, S.M., Saad, G.H.: Business process improvement: empirical assessment and extensions. Bus. Process Manag. J. **14**, 778–802 (2008)
21. van der Aalst, W.M.P., La Rosa, M., Santoro, F.M.: Business process management. Bus. Inf. Syst. Eng. **58**, 1–6 (2016)

22. van der Aalst, W.M.P.: Trends in business process analysis: from verification to process mining. In: Cardoso, J., Cordeiro, J., Filipe, J. (eds.) Proceedings of the 9th International Conference on Enterprise Information Systems (ICEIS 2007), pp. 12–22. Institute for Systems and Technologies of Information, Control and Communication, INSTICC, Medeira, Portugal (2007)

23. Medeiros, A.K.A.: Process mining: extending the α-algorithm to mine short loops. In: Beta, Research School for Operations Management and Logistics (2004)

24. van der Aalst, W.M.P.: Process mining: overview and opportunities. ACM Trans. Manag. Inf. Syst. (TMIS) **3**, 1–17 (2012)

25. Ly, L.T., Indiono, C., Mangler, J., Rinderle-Ma, S.: Data transformation and semantic log purging for process mining. In: Ralyté, J., Franch, X., Brinkkemper, S., Wrycza, S. (eds.) CAiSE 2012. LNCS, vol. 7328, pp. 238–253. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31095-9_16

26. Bozkaya, M., Gabriels, J., van der Werf, J.M.: Process diagnostics: a method based on process mining. In: 2009 International Conference on Information, Process, and Knowledge Management, pp. 22–27 (2009)

27. Jans, M., van der Werf, J.M., Lybaert, N., Vanhoof, K.: A business process mining application for internal transaction fraud mitigation. Expert Syst. Appl. **38**, 13351–13359 (2011)

28. Department of Education Employment and Workplace Relations, National Code of Practice for Registration Authorities and Providers of Education and Training to Overseas Students 2007. In: Australian Government (ed.) ACT, pp. 1–34 (2007)

29. Agresti, A., Franklin, C.A. (eds.): Statistics: The Art and Science of Learning from Data, 2nd edn. Pearson Prentice Hall, Upper Saddle River (2009)

30. Medeiros, A.K.A., Weijters, A.J.M.M.: ProM Framework Tutorial, p. 41. Technische Universiteit Eindhoven, Eindhoven, The Netherlands (2009)

31. Verbeek, H.M.W., Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, Wil M.P.: XES, XESame, and ProM 6. In: Soffer, P., Proper, E. (eds.) CAiSE Forum 2010. LNBIP, vol. 72, pp. 60–75. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-17722-4_5

32. Burattin, A.: data mining for information system data. In: Burattin, A. (ed.) Process Mining Techniques in Business Environments: Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining. LNBIP, vol. 207, pp. 27–32. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-17482-2_4

33. De Weerdt, J., De Backer, M., Vanthienen, J., Baesens, B.: A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. Inf. Syst. **37**, 654–676 (2012)

34. van Dongen, B.F., Alves de Medeiros, A.K., Wen, L.: Process mining: overview and outlook of petri net discovery algorithms. In: Jensen, K., van der Aalst, W.M.P. (eds.) Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems. LNCS, vol. 5460, pp. 225–242. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00899-3_13

35. Han, J., Kamber, M. (eds.): Data Mining: Concepts and Techniques, 2nd edn. Elsevier/Morgan Kaufmann, Amsterdam, Boston/San Francisco (2006)

36. van der Aalst, W.M., Van Dongen, B.F., Günther, C.W., Mans, R.S., de Medeiros, A.A., Rozinat, A., Song, M., Verbeek, H.M.W., Weijters, A.J.M.M.: Process mining with ProM. In: Proceedings of the 19th Belgium-Netherlands Conference on Artificial Intelligence, Belgium-Netherlands (2007)