



The performance assessment framework (PPAFR) for RPA implementation in a loan application process using process mining

Roman Šperka¹ · Michal Halaška¹

Received: 2 June 2021 / Accepted: 2 November 2022 / Published online: 30 December 2022
© The Author(s) 2022

Abstract

When a company decides to automate its business processes by means of RPA (Robotic Process Automation), there are two fundamental questions that need to be answered. Firstly, what activities should the company automate and what characteristics make them suitable for RPA. The aim of the presented research is to design and demonstrate a data-driven performance framework assessing the impact of RPA implementation using process mining (PPAFR). Firstly, we comment on and summarise existing trends in process mining and RPA. Secondly, we describe research objectives and methods following the Design Science Research Methodology. Then, we identify critical factors for RPA implementation and design process stages of PPAFR. We demonstrate the design on real data from a loan application process. The demonstration consists of a process discovery using process mining methods, process analysis, and process simulation with assessment of RPA candidates. Based on the research results, a redesign of the process is proposed with emphasis on RPA implementation. Finally, we discuss the usefulness of PPAFR by helping companies to identify potentially suitable activities for RPA implementation and not overestimating potential gains. Obtained results show that within the loan application process, waiting times are the main causes of extended cases. If the waiting times are generated internally, it will be much easier for the company to address them. If the automation is focused mainly on processing times, the impact of automation on the overall performance of the process is insignificant or very low. Moreover, the research identified several characteristics which have to be considered when implementing RPA due to the impact on the overall performance of the process.

Keywords Process mining · RPA · Simulation · Loan application

✉ Michal Halaška
halaska@opf.slu.cz

¹ Department of Business Economics and Management, School of Business Administration in Karvina, Silesian University in Opava, Karviná, Czechia

1 Introduction

Alignment of information technologies and business processes is a key factor in ensuring business processes' performance as they are being enforced by information technologies. As such, information technologies intervene at both the strategic and operational level through e.g., business process goals, business process metrics, etc. (Tallon et al. 2016). This also applies to RPA, the implementation of which often does not require deeper intervention in the information infrastructure of the company or business process, or these are not as extensive and resource-intensive as in the case of traditional automation. Lower complexity and resource intensity of RPA implementation may lead to the choice of a lax approach when deciding on the deployment of RPA. Which in return may lead to an inadequate return on investment and inferior state of the process regarding efficiency, effectiveness, etc. RPA is a newly emerging technology that operates at an interface with other systems based on an outside-in approach in order to exchange human resources for software robots (Stople et al. 2017). Thus, it is possible to integrate RPA with any software system used by human workers (e.g., legacy systems, CRM, ERP, etc.). Madakam et al. (2019) postulated that companies that do not start implementing RPA in the near future will not be able to withstand the pressure of competition. This seems to correspond with adoption rates of RPA. According to Computer Economics Avasant Research (2021), 20% of all organisations have adopted RPA in 2021, which is up from 13% adoption rate in 2020 and 12% in 2019. Similarly, the percentage of organisations making new investments in 2021 was 26%, up from 24% in 2020 and 21% in 2019. Adoption rate of RPA is the highest in manufacturing industry (35%), followed by technology sector (31%), healthcare (10%), retail and consumer packaged goods (8%), finance (8%), healthcare (5%) and public sector (3%) (Acceleration Economy 2021).

However, when the company decides to automate their processes by means of RPA, there are two fundamental questions that need to be answered. Firstly, what activities should the company automate and what characteristics make them suitable for RPA (van der Aalst et al. 2018)? This issue is not new and relates to STP (Straight-Through Processing), which was popular in the mid-1990s, especially in the financial sector and limited to a very small number of processes. Having said that, technological advances in data science, machine learning, and artificial intelligence are crucial stimuli for revisiting this issue. Today, RPA offers far wider application possibilities for more complex, less structured and less routine processes, as shown by the ever-growing supply on the market. Secondly, why should companies automate their activities? Automation should not be done with purpose of simply automating the activity. Such projects typically do not provide the desired results. The automation of the activity should be based on a simulation, which will also allow for the evaluation of the success of the deployment of RPA within the process. The selection of inappropriate processes for the implementation of RPA is cited as one of the most fundamental reasons for the failure of the implementation of RPA in the company (Osmundsen et al. 2019; Lamberton 2016). Thus, it is useful to apply the principles of business process

management (BPM) and process mining, as a result of the raise of interest in the area of business process performance and process optimisation (Isik et al. 2012). Similarly to BPM, many implementation failures could have been avoided if, prior to starting the implementation, the nature of the processes were examined (Szelągowski 2021). Therefore, it is necessary to address a non-technological perspective on the implementation of RPA using a holistic approach.

The purpose of the research in this paper is to present a data-driven framework assessing the impact of implementation of RPA using process mining. RPA studies are scarce in the mainstream IS journals (Ivančić et al. 2019). Furthermore, they very often focus on the implementation of RPA from the technological perspective, while mostly ignoring the non-technological perspective. Thus, the objective of this research is to design and demonstrate a process mining based performance assessment framework of RPA implementation (PPAFR). This framework should help organisations to identify potentially suitable activities for RPA implementation and not overestimating potential gains. This research is addressed to decision makers and people who implement RPA. Its focus is on the performance of the overall process. With this objective in mind, we postulated following research questions:

- RQ1: What RPA factors can be monitored and analysed using process mining techniques?
- RQ2: What process performance metrics apply to RPA characteristics based on RQ1?
- RQ3: How can process mining and business process simulations be used to assess activities suitable for RPA implementation?

In the first section, we present a literature review. The second section introduces the research objective and adopted research methodology. The third section presents the design of the assessment framework based on systematic literature research in the areas of RPA, BPM and process mining. The fourth section describes the demonstration of the PPAFR with the use of real data from loan application process and discusses the results. To conclude, we summarise our results and the prospects of its further development.

2 Literature review

In this section, we firstly introduce the fundamental concepts of process mining. Secondly, we introduce the RPA technology and its relation to BPM research. As is shown based on the literature review, both process mining and BPM complement RPA and its implementation in several areas.

2.1 Process mining

Despite all the benefits, in order to identify and implement RPA, extensive knowledge about the process is required and benefits of RPA are far less significant

as much time and effort has to be put into gaining that knowledge (König et al. 2020). Here, process mining techniques come to use. Process mining has been receiving considerable attention in recent years. The wave of digitisation is one of the main reasons for the popularisation of process mining techniques in connection with business processes. Besides that, information systems are increasingly appearing in small and medium-sized companies. Process mining techniques seek to recognise patterns and other information within data produced by business information systems (van der Aalst 2016; Dumas et al. 2018). Process mining thus goes hand in hand with the current trend, where companies base their decisions on information obtained from available data. The essence of process mining is to analyse business processes that are objectively represented by data, so-called event logs. Event logs capture recorded actions in the form of events carried by agents operating over systems infrastructure (Szimanski et al. 2013).

There are five types of process mining approaches: process discovery, conformance checking, enhancement, deviance mining and online support (Dumas et al. 2018; van der Aalst 2016). The important areas of process mining for this research are presented in more detail. The main goal of process discovery is to find patterns in the logs, based on which a process model of the analysed process is constructed. Currently, one of most successful process discovery techniques is split miner, which is based on criteria such as accuracy, generalisation, complexity and soundness of the discovered model (Augusto et al. 2019).

None of the discovery techniques guarantees that the discovered model really corresponds to the original process. It is therefore necessary to verify that the discovered process model is of good quality. The quality of the model is assessed using various criteria. These cases are addressed using conformance checking. The fundamental metrics are fitness, precision, generalisation and simplicity. When building a simulation model, fitness and precision are of greater importance. Fitness is the ability of a model to reproduce the behaviour seen in a log and can be measured according to Adriansyah et al. (2011). Precision is the ability of a model to generate only the behaviour found in a log. If we use a state machine enriched with log behaviour and penalty-optimal alignment to calculate precision, we can measure it according to Adriansyah et al. (2015).

The essence of process enhancement is the extension or improvement of existing process models using information from the log of the monitored process. Techniques such as adding attributes to events (such as bottlenecks, service levels, frequency of occurrence, etc.), sorting traces, or correcting (for example, redesigning the model to better reflect reality, etc.) are used to improve processes. Processes can also be improved by adding different perspectives such as organisational perspective, data perspective, etc. (van der Aalst 2016). Based on the above, process enhancement can be defined as an extension or improvement of an existing process model using information from the current process record (Yasmin et al. 2018). Moreover, the performance of the process can vary considerably over time, geographically, or across business units, product types, or customer types. And even within these cases we can observe significant differences between the cases with the best and with the worst performance. Thus, deviance mining techniques allow us to analyse differences across subsets of cases in a process (Dumas et al. 2018).

In this research, we use process mining in two ways: firstly, to acquire as-is process model of a loan application process and its analysis (van der Aalst 2016). Secondly, for simulation model creation based on which to-be model should be proposed. Business processes are usually described at a high level of abstraction using modelling languages such as BPMN, EPCs or Petri nets, etc. While business processes are defined and modelled at a high level, the analysis of runtime behaviour through process mining techniques is based on the low-level events recorded in the event log, as each agent carries out its operations (Szimanski et al. 2013). There is ongoing research involving use of process mining techniques for creation of simulation models of business processes (Rozinat et al. 2009a, b; Camargo et al. 2018). In this regard, process mining techniques are faster and more reliable than other sources, e.g., process documentation, interviews or direct observations.

2.2 Robotic process automation

Recently, there has been a strong interest in industry in a specific area of automation called RPA. RPA technology is used to automate virtually running business processes previously performed by employees and, therefore, enable employees to be involved in more complicated tasks, which can bring organisation more value. However, it is true that implementation of RPA may lead to dismissal from employment or the need to retrain the employee.

Baranauskas (2018) defines RPA as an imitation of everyday human activities based on IT, in which only a limited number of autonomous decisions are required and in most cases this activity is performed in large numbers and in a short period of time. Slaby (2012) defines RPA as the technological imitation of a human worker with the goal of automating structured tasks in a fast and cost-efficient manner. According to Syed et al. (2020), for authors who have tried to define RPA, the most recurring topics are the replacement of human activity by software agents and the integration of these agents with front-end systems similar to human agents. Example of such interaction is the transfer of data from ERP or CRM systems to a web application and vice versa (Aguirre and Rodriguez 2017), inserting customer data and placing an order on a website, data updates, cross-data validation, data migration and entry, mass email generation, claims processing, and other activities related to information processing. RPA is a software that delivers business processes thus, it is necessary that the process inputs and outputs be structured and in a machine-readable format. To summarise, RPA is a software technology that makes it easy to build, deploy and manage software robots that emulate human actions interacting with digital systems and software. Ultimately, therefore, it copies actions performed by employees and does not change the flow or logic of the process (Fung 2014). Thus, the result being that the deployment of RPA on inefficient business process does not improve such process as is generally the case for any type of automation (van der Aalst et al. 2018).

Advances in artificial intelligence technologies and machine recognition technologies greatly increase RPA capabilities (König et al. 2020). There are two different approaches towards RPA software agents (Lasso-Rodriguez and Winkler

2020): firstly, software agents are based on relatively simple and clear rules performing repetitive tasks at a high frequency. Secondly, software agents are trained on data and executing complex tasks while being flexible and adaptive mostly due to combination of RPA with machine learning and/or further artificial intelligence functionalities. In case of traditional RPA, process to be automated should be well defined and have a low change rate and decision complexity. Otherwise, success of RPA implementation and its performance are severely limited (van der Aalst et al. 2018; Aguirre and Rodriguez 2017; Lacity and Willcocks 2016; Baranauskas 2018). However, integration of RPA with artificial intelligence, cognitive computing and other advanced digital technologies allows RPA to be reallocated from performing repetitive and error-prone routines in business processes towards more complex knowledge-intensive and value-adding tasks.

The main advantages of implementing RPA in business processes include (Lacity and Willcocks 2016; Aguirre and Rodriguez 2017; Slaby 2012; Baranauskas 2018; van der Aalst et al. 2018): increase in productivity and effectivity, flexibility, reliability, accuracy, consistency, customer and employee satisfaction, cost reduction and non-invasiveness of technology. Syed et al. (2020) summarise these benefits into four main groups: operational efficiency (e.g., reduction in durations, cost and human resources, reduction of manual tasks and workload, and increased productivity), quality of service (e.g., amount of errors, incorrect data inputs, mistakes, availability), implementation and integration (e.g., relatively faster, cheaper and easier to implement, configure and maintain than other forms of automation), and risk management and compliance (e.g., reduction of risks, increase in compliance regulatory requirements).

Even though RPA is often used in connotation with the term “robot”, it is not a physical robot but software-based solution that is operating on an interface of different systems. It is unlike traditional software, which communicates with other IT systems via back-end. Thus, RPA does not disturb the underlying IT systems and only replaces the existing manual task with the automated one through a presentation layer. This allows the integration of RPA with no required programming skills and with practically any software system, regardless of its openness to third party integration (Aguirre and Rodriguez 2017). The underlying concept of RPA is that software robots replace human resources, which results in decreasing costs and increasing efficiency and consistency. Thus, the implementation of RPA technology has similar goals to that of BPM. Many researchers suggest integrating RPA with BPM as it solves its limitations (König et al. 2020). RPA itself does not provide techniques required to gather information necessary for the deployment of RPA, dealing with exceptions during the execution of automated processes and managing process automation on an organisational level. There is research for integration of RPA with Business Process Management Systems focused on a technological point of view like, e.g., König et al. (2020). According to Aguirre and Rodriguez (2017), there are following characteristics that distinguish RPA from technologies like Business Process Management Systems:

- RPA sits on the top of existing systems and access these platforms through the presentation layer, so no underlying systems programming logic is touched.

- In contrast to most BPMN modelling packages, RPA solutions do not require programming skills for software interface configuration. RPA is set to work by just dragging, dropping and linking icons.
- RPA does not create a new application and does not store any transactional data, so there is no need of a data model or a database like BPMS systems.

According to Ivančić et al. (2019), RPA is more often implemented in practice than it is investigated by researches, even though there is a number of authors reporting various benefits of RPA implementation (Lacity and Willcocks 2016; Aguirre and Rodriguez 2017). Moreover, the research on how to successfully utilise RPA technology is lacking (Syed et al. 2020). RPA complements BPM in many areas and vice versa, but does not replace it. Therefore, it is important to discuss differences, similarities and complementarities between RPA and similar technologies and approaches. According to van der Aalst (2021), old-school BPM tends to be unable to capture human behaviour, deal with complexity of real-life systems, and realize actual improvements. Process mining helps by looking at real processes in an objective manner before and after interventions. Moreover, just like RPA, process mining does not try to replace existing systems and face the complexity of real-life systems.

2.3 Robotic process automation and business process management

Digitalisation can have significant impact on organisations concerning profitability and competitiveness. According to resource-based view (RBV), adequate management of organisations' resources (i.e., processes, assets, capabilities, knowledge, etc.) can bring higher performance and competitive advantage; however, only resources characterised by VRIN (valuable, rare, inimitable, non-substitutable) attributes allow to build sustained competitive advantage (Barney 1991). Nevertheless, RBV is too static for dynamic environments experiencing fast market and technological changes. Thus, RBV was extended by dynamic capabilities framework, which defines dynamic capabilities as ability of organisation to integrate, build and reconfigure internal competences to address rapidly changing environments (Teece 2018; Teece et al. 1997). Besides dynamic capabilities, there are ordinary capabilities targeting efficiency in operations. Thus, managerial and organisational processes are one of the key factors for competitive advantage and profitability of the organisation. Similarly to BPM, RPA gives organisations opportunity to improve their organisational processes. Based on the Teece's characterisation of capabilities, RPA technology is an ordinary capability. However, it is an emerging technology for automating business processes and its implementation process is yet to be standardised.

It is obvious, that the successful implementation of RPA is dependent on the organisation's business processes regarding both technological and non-technological perspective (Plattfaut et al. 2022). Business process is a collection of inter-related events, activities and decision points that involve a number of actors and objects, and that collectively lead to an outcome that is of a value to a customer (Dumas et al. 2018). Thus, business processes are about getting business results for customers or about getting internal intermediate results that contribute to the end results,

in an effective and efficient way. Consequently, BPM aims at providing techniques and software to design, enact, control, and analyse business processes involving humans, organisations, documents and other sources of information (Di Ciccio et al. 2015). Moreover, due to the nature of BPM, it can be considered dynamic capability itself (Niehaves et al. 2014). It is therefore appropriate to use established concepts of BPM in combination with process mining together with RPA.

The adoption of new technologies assumes that the organisation is ready to implement such technology. For the assessment of BPM capabilities of an organisation, the so-called maturity models (Poepelbuss and Roeglinger 2011) are used. Maturity models evaluate maturity of organisations' business processes from low maturity, which is characterised by ad hoc processes, to high maturity, which is characterised by broadly embedded BPM in organisation's operations and strategy. BPM maturity models are subject to a great deal of criticism, i.e., oversimplification, lacking empirical foundation, limited extent of prescriptive properties, etc., (Poepelbuss and Roeglinger 2011; Szelągowski and Berniak-Woźny 2019). According to Szelągowski (2021), there is a lack of holistic approach in implementations of BPM, especially concerning emerging new technologies like RPA. It is due to the lack of tools enabling the quick diagnosis of the nature of business processes of organisations, which BPM maturity models no longer fulfil. Moreover, BPM maturity models are generally adequate for descriptive purposes, but are lacking in their prescriptive qualities (Röglinger et al. 2012).

The methods needed for conducting a successful BPM project can be structured into the BPM lifecycle and thus providing the iterative methodology for the enactment of BPM on the level of business processes (König et al. 2020). There are several versions of BPM lifecycle (Szelągowski 2021; Dumas et al. 2018; van der Aalst et al. 2007) with differently structured phases; however, the including activities and their order stay the same. Weske (2019) presents BPM lifecycle consisting of 4 phases: design, configuration, enactment and evaluation. He puts emphasis on the use of process mining in evaluation phase processing data retrieved from enactment phase. In this case, process mining is used for evaluation of an on-going as-is process model. However, van der Aalst et al. (2007) already discussed the potential of process mining techniques supporting the entire BPM lifecycle. To make any impact on business processes, it is fundamental to capture and characterise it in some way (Reijers 2021). Thus, process modelling is one of the core activities of BPM (van der Aalst 2016; Dumas et al. 2018). If the implementation of RPA is approached holistically, then process modelling has to play a crucial role to help capture and characterise business processes in some way, e.g., control flow, data perspective and resource perspective. Process modelling is then closely related to process design. According to Reijers (2021) process models are often useful to capture the design of processes, while also allowing analysis and enactment of the process. Process design involves decisions about organisation of processes, used technologies and assignment of responsibilities apart from mere modelling. Thus, process mining can be used as well in design phase of BPM lifecycle regarding the implementation of RPA with focus on performance of the process. In this case, the information acquired from analysis of as-is process model is used to identify activities suitable for RPA implementation with regard to its performance.

There is ongoing research on the nature of business processes, which considers contextual factors that at least partly determine success and expected results of process-management initiatives (Zelt et al. 2018), i.e., scope of RPA is limited and cannot be used at the organisational level (König et al. 2020). Thus, based on the functionality of the processes (Szelagowski and Berniak-Woźn 2019), RPA should focus on operational and support processes. According to Ivančić et al. (2019), the general concept of RPA is industry agnostic and thus, applicable within business processes regardless of economic sectors. Furthermore, depending on the dynamism of execution, business processes can be divided into (Szelagowski and Lupeikiene 2020): (1) structured (static) processes, (2) structured processes with ad hoc exceptions, (3) unstructured processes with pre-defined fragments and (4) unstructured processes. Thus, based on the structure of the process, RPA is best used within structured processes with ad hoc exceptions and unstructured processes with pre-defined fragments. Structured processes are automated using heavy automation, and unstructured processes are too complex for traditional RPA. However, many of the process typologies are broadly defined and include a wide range of different processes. This limits their use for the assessment of the suitability of the process for RPA implementation and similarly the impact on the overall performance. Zelt et al. (2018) derived 36 process dimensions, which are referred to as factors within the RPA literature. Furthermore, the interactions of different dimensions of the process are not clear and business process simulations may provide a guidance in this regard.

Vendors expect from RPA cost reduction, accuracy, timeliness and improved compliance (Ivančić et al. 2019). Thus, the performance of the process may be depicted using the so-called Devil's Quadrangle (Dumas et al. 2018) consisting of four dimensions allowing for the meaningful assessment of the process. The dimensions of the Devil's Quadrangle are: time, cost, quality and flexibility. These four dimensions are very well traceable through simulations of business processes. Moreover, they are also well suited for the measurement of benefits of RPA. According to Dumas et al. (2018), task automation as a process-redesign pattern should generally result in improvements with regard to time, cost and quality dimensions, while there should be opposite effect when it comes to flexibility dimension. In our research, we focus on time and quality perspective of Devil's Quadrangle from an efficiency perspective. Process performance can be improved in two ways: (1) change of process flow; and (2), quality of the process execution can be improved by minimising variations in the process performance for a given process design (Frei and Harker 1999). Since RPA technology does not change the flow of the process, we focus on process execution through development of formal process model as a basis for process analysis. From the IS development perspective, it is necessary to understand and possibly recognise business processes so that the introduction of information technology has the highest possible impact on them.

Wastell et al. (1994) introduced a PADM framework which deals with process innovation. However, the authors do not specify which principles should guide the selection of the process, but rather identify several criteria upon which selection can be based. Becker et al. (1999) introduced a framework for the identification of workflow relevant business processes based on technical, organisational and economic criteria. Technical criteria for workflow automation relate to the process structure, the

resources involved, the throughput and the overall quality of the process (e.g., number of process instances, number of different application systems, number of exceptions, etc.). Organizational criteria relate to the organizational environment of the business process (e.g., mentality for innovations, procedural orders, quality of documentation, etc.). Economic criteria relate to the benefits in relation to the specific business goals that can be expected from an automation of the business process (e.g., digitalization of routine work, enhancement of process transparency, etc.). Škrinjar and Trkman (2013) identified BPM critical success factors and their critical practices as follows: strategic alignment, performance measure, organisational changes, IT support, and employee training and empowerment. Mutschler et al. (2008) identified three groups of critical success factors for IS implementation: organisation-specific, project-specific and technology-specific critical success factors. Even though BPM emphasises organisations processes, it has a firm link to the business performance. It depends on organisation, culture and maturity. The organisation is being recognised as a system of interacting processes whose performance must be balanced, and the BPM methods have an indirect contribution to process performance by establishing a process management culture (Lasso-Rodríguez and Winkler 2020). According to Ongena and Ravesteyn (2019), organisational performance is posited as outcome of BPM maturity. Initiatives aiming at increasing an organisation's BPM maturity thus subsequently lead to better organisational performance. For our study, we comply with the notion that the relation between BPM maturity and actual organisational performance is mediated by the performance of an organisation's processes (Rosemann and de Bruin 2005). Hence, rather than examining performance at an organisational level as used by several studies, we use process performance or success as proxy for organisational performance, similarly as this has been used in prior studies. Evaluation of the performance of business processes and its involved elements is important as it is used as a tool to control and improve the processes. More recent research also found that business process outcomes typically relate to efficiency, effectiveness, and agility/flexibility (Ongena and Ravesteyn 2019). The performance is usually evaluated in the form of quantitative measurements which help to indicate about quality. Business processes and its elements are evaluated in different dimensions like time, cost, and quality (Lodhi et al. 2014). For each of these performance dimensions different key performance indicators can be defined such as (van der Aalst 2013a): lead time, service time, waiting time, resource utilisation, number of complains or number of product defects, etc.

3 Research objective and methods

Based on the previous sections, it is necessary to create a diagnostic tool enabling a holistic approach to RPA implementation. The appropriate tools in this respect appear to be BPM and process mining methods and techniques. The research gap and practical need for the assessment of the impact of RPA implementation on the performance of the overall process and identification of potentially suitable activities were a direct inspiration to formulate the research objective. Thus, the objective of this research is to design and demonstrate assessment framework

of RPA implementation using process mining. This framework should help organisations to identify potentially suitable activities for RPA implementation and not overestimating potential gains based on the performance of the process. To achieve this objective, the Design Science Research Methodology (Peffer et al. 2007; vom Brocke et al. 2021; Szelągowski 2021) was applied as follows.

3.1 Problem specification and motivation phase

In this phase, a literature review and research focused on positioning of RPA within BPM and process mining concepts (see Sect. 1). In the second stage, systematic literature review focused on identification of factors important for the implementation of RPA with focus on the performance of the overall process (see Sect. 4.1). Moreover, we focused on factors that can be evaluated using process mining, e.g., even though culture might have impact on the implementation of RPA (Willcocks et al. 2015a), we exclude such factors. After the factors were identified, we grouped them into characteristics relevant for assessment of the process performance. The grouping was based on an overlap of the factors or close relation to other factors. Meaning that overlapping or closely related factors were grouped into the same characteristic.

3.2 Objective of the solution phase

The aim of this phase was to develop a flow of the process mining based performance assessment process of RPA implementation (see Sects. 4.1 and 4.2), which will enable organisations, managers and practitioners to select methods and techniques supporting the analysis of the performance of the implementation of RPA based on its characteristics, which were derived from systematic literature review of RPA literature in the previous phase. Concretely, we focus on the use of process mining techniques with relation to the specified characteristics and the overall performance of the process.

3.3 Design and development phase

This phase, based on the previous literature research and analysis of the presented data, involved the design and development of the PPAFR assessment framework. In this phase, we specified parameters (Table 2) that are measurable with the use of process mining techniques to help organisations in the assessment of activities suitable for RPA implementation with the focus on the performance of the overall process. This was a necessary step, because the factors specified in the literature as crucial for RPA implementation are not specified in a way that allows organisations, managers and practitioners to use them for the assessment of the implementation of the RPA.

3.4 Demonstration phase

This phase demonstrates the process of the process mining based the performance assessment process of RPA implementation, methods and techniques used in PPAFR assessment framework and the framework itself (see Sect. 5). PPAFR assessment framework was demonstrated using simulation of business processes, which are part of the assessment process. It is not an uncommon approach to use simulations for evaluation or redesign of business processes as they are widely used in the area of BPM, i.e., Märušter and van Beest (2009); Rozinat et al. (2009b).

3.5 Evaluation phase

Evaluation of the presented PPAFR assessment framework was partially covered in the demonstration phase through simulation of business processes. Simulations in business process area is an important issue, frequently used for predicting the systems performance (Pérez-Álvarez et al. 2018). Furthermore, based on the previous research, several practitioners have shown interest in the application of process mining techniques within the implementation of RPA. In the future research, we would like to investigate applicability of PPAFR framework by practitioners and investigate the weights of monitored parameters introduced in Table 2, so that the assessment framework is more self-contained and less dependent on business process simulation.

4 Performance assessment framework of RPA implementation

4.1 Identification of RPA characteristics

To use the RPA to improve business processes, it is necessary to identify the activities contained in the process that should be automated. For this reason, it is necessary to know the characteristics of these processes, based on which the activities will be determined. The systematic literature review on the implementation of RPA was used for the general enumeration of all characteristics. The Table 1 shows the most important factors that are mentioned in the literature as central to the implementation of RPA. The table is based on a systematic literature search in the Web of Science and Scopus databases, and the web search engine Google Scholar was also used. In the first stage, the searches included the terms “RPA” AND “Implementation”, “RPA” AND “Factors”, “RPA” AND “Success factors” and “RPA” AND “Process assessment”. The search terms were allowed to appear in all accessible fields. The entries were then filtered based on areas of interest, which are business, management, economics and computer science. In the second stage, the search included terms “RPA” or “robotic process automation” and the entries were again filtered based on the areas of interest. The coding was done in both stages. In the first stage, we read the articles

Table 1 Factors critical for RPA implementation

Factors	Factor description	Frequency
Rule based	Unambiguous description of business logic through scenario specific business rules	27
Standardisation	Higher standardisation of processes is better for RPA implementation	24
Highly frequent tasks	Sufficient volume that helps to maximise benefits of RPA	24
Application involvement	Processes that interact with several systems are better candidates for RPA implementation	19
Repetitive	Automated processes should be repetitive enough to yield better return on investment	18
Error prone	Processes that are prone to error are better candidates for RPA implementation	16
Low level of exceptions	Exceptional behaviours in the process limits the implementation of RPA	12
Stable	Stable processes that are in place for a while and are well understood	10
Limited human intervention	Processes that are capable of running without human intervention are better RPA candidates	10

and identified factors that were important for RPA implementation according to the authors of the articles. In the second stage, we read the articles and again identified factors that were important for RPA implementation; however, this time with the information acquired from the first stage. To be able to answer RQ1, the search focused on RPA factors that could be monitored and analysed using process mining techniques. Articles without full access, extended abstracts and articles citing the term RPA with a different meaning were excluded. The search was not constrained to any time frame to ensure full coverage of published literature. Factors with a frequency of occurrence of less than 5 were excluded. In addition, factors not related to process structure such as company culture, employee involvement, management approach to implementation, etc., were excluded. After the initial search in the first stage was conducted, we used the snowball method to identify literature not included in Web of Science and Scopus databases, including industry white papers. Due to the limited amount of research on RPA, we also included factors found in the professional literature published by, e.g., Deloitte, UiPath, Minit, etc. Table 1 was created based on 56 different sources of professional and academic literature. Methodologically, one researcher extracted the data, and the other checked the extraction. When there was disagreement, we discussed the issues until we reached an agreement.

Based on the literature review and Table 1, we identified following five characteristics: frequency-significant activities, error-prone and rework, involvement of manual work, process flow and logic, and productivity and efficiency. We grouped the factors from Table 1 into characteristics relevant for assessment of the process performance. The grouping was based on overlap of the factors or close relation to other factors. Meaning that overlapping or closely related factors were grouped into the same characteristic.

4.1.1 Frequency-significant activities

Frequency-significant activities characteristic is based on the following factors from Table 1: rule based, highly frequent tasks and repetitive. Activities that are carried out in large numbers are typically routine (and thus, based on clear rules), rule based and are constantly repetitive, with automation often being an appropriate choice (Syed et al. 2020; Radke et al. 2020; Madakam et al. 2019; Cooper et al. 2019; Huang and Vasarhelyi 2019; Kokina and Blanchette 2019; Baranauskas 2018; Asatiani and Penttinen 2016; Fung 2014). This means that such processes occur very often in the observed period and fluctuations in the number of occurrences have a major impact on the productivity and efficiency of the process. In addition, targeting these types of activities and automating them is also crucial on the cost side, as these activities help maximise the benefits of RPA implementation. For these reasons, it is generally better to target the implementation of RPA to frequency-significant activities. However, the exception is highly valued activities. For such activities, the need for accuracy and reliability outweighs the costs associated with automation and the transaction itself.

4.1.2 Error-prone and rework

Error-prone and rework characteristic is based on the following factors from Table 1: error prone. One of the important benefits of RPA implementation is accuracy, consistency and reliability. These play an important role when the process is prone to errors and unforced rework due to the involvement of people in the process and the activities performed (Radke et al. 2020; Huang and Vasarhelyi 2019; Madakam et al. 2019; Cooper et al. 2019; Sutherland 2013). The error-prone and rework is closely related to the characteristics of frequency-significant activities, as this is the type of transaction. Especially if these transactions require the attention and involvement of employees for a long time. Similarly, error-prone and rework is closely linked to productivity and efficiency, as fewer errors and reprocessing contribute to fewer working and waiting times, or process efficiency. In this respect, the RPA can be used to enforce compliance with the established rules that define the RPA itself (Radke et al. 2020; Cooper et al. 2019). Together with other characteristics, the reduction of errors and reworking causes an increase in value to the customer through higher quality together with higher satisfaction.

4.1.3 Involvement of manual work

Involvement of manual work characteristic is based on the following factors from Table 1: limited human intervention, low level of exceptions and application involvement. The need for manual work occurs in many definitions of RPA, because from its principle, its implementation replaces human workers (Baranauskas 2018; Fung 2014). In addition, this characteristic is closely linked to the susceptibility to error and reworking that human workers are prone to. Manual work is often characterised by intensity and constantly recurring activity (Radke et al. 2020; Cooper et al. 2019). Although, in the end, such activities do not necessarily require human intervention or require them only minimally when dealing with exceptions. In addition, the workload of such activities must be handled during working hours, while the RPA is able to handle it at any time of the day, throughout the year (van der Aalst et al. 2018; Osmundsen et al. 2019; Lacity and Willcocks 2016). RPA is therefore best used for repetitive, standardised tasks based on clear rules (Radke et al. 2020; Cooper et al. 2019; Willcocks et al. 2015a, b). In the case of deployment, it reduces employees' workload and burden associated with constantly recurring routine tasks, and thus saves time (Aguirre and Rodriguez 2017).

4.1.4 Process flow and logic

Process flow and logic characteristic is based on the following factors from Table 1: standardisation and stable. Business process management typically focuses on the logic and flow of business processes. However, RPA is used to mimic the workload of human workers and does not change the logic or flow of the process. Process flow and logic is therefore an important characteristic of the process when considering the implementation of RPA in it (Halaška and Šperka 2020). It is neglected in the literature, although it can have a significant impact on the selection of activities and

processes suitable for implementation. It is necessary to consider the process patterns within which the RPA is implemented, as well as the complexity or number of steps required to carry out the activity. In addition, the process flows and process logic must be stable and unchanged in the long run. Automating a process that changes frequently is a waste of time as developers spend a lot of time maintaining it. Therefore, rather stable processes are suitable for the deployment of RPA.

4.1.5 Productivity and efficiency

Productivity and efficiency characteristic is related to the outcomes and benefits of RPA. Achieving higher productivity and process efficiency is one of the key expectations when implementing RPA. The reduction of the workload of human workers, which is transferred to software robots (van der Aalst et al. 2018), contributes to this. These software robots are able to process a high volume of workload, which takes significantly more time for human workers. The result should be a reduction in process and waiting times and an increase in the efficiency and flexibility of the redistribution of resources needed to carry out the activities. This needs to be considered in particular in the case of processes and activities that show significant fluctuations in workload due to changes in transactional demand, which creates a demand for temporary labour that is often not easy to obtain. The throughput of the process also needs to be considered in this regard. From the point of view of efficiency, therefore, the involvement of RPA leads to a better use of resources resulting from the minimisation of errors, reduction of costs associated with defects and increased productivity. Increased productivity can be combined with scalability, but this is not the only reason to increase it. Another reason is to increase the efficiency of processes.

In the following subsection, we present data (see Sect. 4.2) and process stages of the assessment framework (see Fig. 1 and Sect. 4.3). In our research, we deal with a complex real-life process of a Dutch financial institution. Namely, it is the loan application process, which involves employees of the financial institution as well as customers requesting the loan. The fact that it is a real-life event log is one of the two reasons why the loan application process was chosen. The second reason is that financial processes are generally appropriate for the implementation of RPA.

4.2 Data

The research uses publicly available event log (Dongen 2012). The event log describes the execution of the loan application process. The log contains 13,087 process instances or in other words cases, which are formed by 262,200 events. One of 24 activity names contained in the log is ascribed to all events. The reference period of the log is 1 October 2011–14 March 2012. The average case duration in the log is 8.6 days. The log contains three types of events. Each event name starts either with A, O or W. The A events are related to applications, the O events are related to offers sent to customers, and the W events are related to processing of work items of applications. Each event in the log has a total of 9 recorded attributes, where: “Case. ID” identifies each case, “Activity” identifies activities within each case, “Resource”

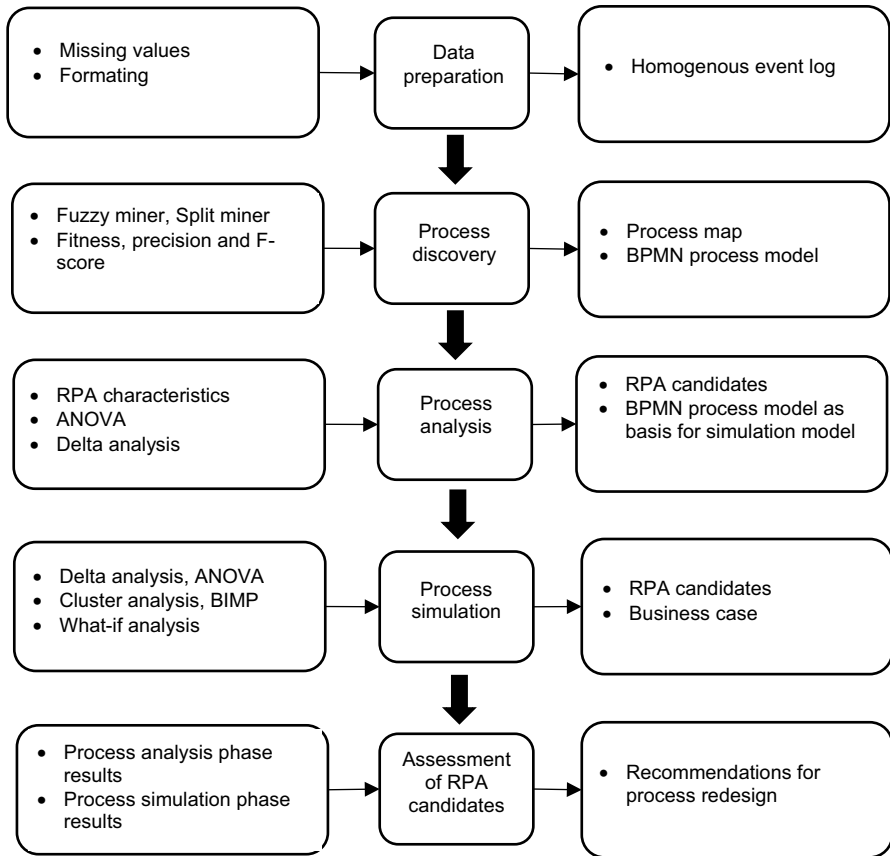


Fig. 1 Process phases of the assessment framework, its methods, inputs and outputs

identifies resources performing particular events, “Complete.timestamp” identifies an occurrence of events and “Lifecycle:transition” identifies the state of activity. The overall workflow of the process is as follows: after submitting an application, small part of the applications is controlled for fraudulent behaviour, the rest of them are controlled for completeness, after that the application is pre-accepted and the application is processed. Some applications are cancelled and the offer is sent to the rest of the customers and the contact with customer follows. In case that the customer accepts the offer, application is assessed and the loan is approved. In some cases, after assessment of the application, further contact with the customer might be required to complete the application.

The publicly available dataset is appropriate to use for process mining techniques. The data set was used with many purposes. Pourbafrani et al. (2020) used the data set in combination with time-granularity detection framework, which is a technique used for detection of time step-size for time-series analysis. Van der Aalst (2013b) uses data set to demonstrate the use of process mining techniques with regard to

services. Pasquadibisceglie et al. (2020) demonstrate the relation of computer vision to process mining. Verbeek and van der Aalst (2015) demonstrate a framework for decomposed process discovery and decomposed conformance checking using integer linear programming. Rafiei et al. (2020) use the data set to present approach that allows to hide confidential information in a controlled manner. Verbeek et al. (2017) used the data set to present the use of several ProM plug-ins like, e.g., resource work analysis, fuzzy map miner, pattern abstraction, heuristic miner, conformance checker etc. Adriansyah and Buijs (2013) used alignments between process model and even log to manually improve the process models obtained by algorithms and projected performance information on it. Augusto et al. (2019) used the data set for the evaluation of presented process discovery technique called split miner. However, to the best of our knowledge, there is no research focusing on use of process mining for assessment of activities suitable for RPA implementation.

4.3 Process stages of the assessment framework

In this subsection, the individual stages of the assessment framework process and applied methods will be presented. The process consists of five stages: (1) data preparation, (2) process discovery, (3) process analysis, (4) process simulation and (5) assessment of RPA candidates.

4.3.1 Data preparation

At this stage, it is necessary to prepare the logs for the application of individual process mining techniques. This means an extraction of logs from various database sources. Process mining techniques are performed in software tools that work with certain data file formats (CSV, XML, MXML). The log of the loan application process was already available in XES format. Thus, it was checked whether all events in the log contain the basic required attributes in the appropriate formats, i.e., unique case ID's, all events have assigned activity names, all time stamps are in the format "dd.mm.yyyy hh:mm:ss ", lifecycle transition and finally, all events have assigned resources. Events and related cases that did not possess required attributes or did not respect necessary formats were modified to respect them. Otherwise, they were excluded. Missing values were handled similarly. As the focus was solely on the aforementioned attributes (case ID, activity, timestamp, resources, lifecycle transition), the rest of the attributes were ignored and no cases nor events were removed from the log based on these attributes (e.g., amount requested, variant, index variant, etc.). The output of data preparation stage is the event log, which is used in the process discovery stage to discover the process model of the loan application process.

4.3.2 Process discovery

When deciding on the implementation of RPA, the key issue is first and foremost the identification of suitable business processes. In this regard, process mining as a data-oriented approach provides a significant advantage over other practices commonly

used in BPM to identify processes, such as workshops, interviews, etc., where process owners may not necessarily be familiar with the entire process. To find suitable candidates for RPA implementation and their assessment, it is necessary to know how the process works. For this purpose, we used process discovery techniques implemented within process mining tools Disco¹ and Apromore.² Process discovery in Disco is based on fuzzy mining and the discovered model is represented in the form of a process map (Günther and van der Aalst 2007), which is basically directly-follow graph. Then, the discovered process map is adjusted using an integrated path and activity filters to acquire a readable process map that is appropriate for further analysis. In our approach, process maps are used for preliminary examination of the process and relations between activities within the process. However, process maps do not show the process flow and logic, nor have executable semantics, which are necessary for the simulation purposes, thus to acquire an executable process model, we use Apromore. Process discovery in Apromore is based on split miner (Augusto et al. 2019) and the discovered model is represented in the form of a BPMN diagram (Business Process Model and Notation 2.0). Furthermore, split miner performs very well among other process discovery techniques regarding fitness and precision of discovered process model, and performs among the best in terms of F-score (Augusto et al. 2019). Similarly to Disco, noise in the log is filtered out using integrated nodes, arcs and parallelism filters. The nodes filter is similar to activity filter in Disco and it filters out activities based on the frequency of their occurrence. The arcs filter is similar to paths filter in Disco and it filters out arcs based on the frequency of their occurrence. The parallelism filter offers a possibility to adjust the amount of parallelism (e.g., AND and OR gateways) discovered. Finally, the quality of the process model discovered using Apromore is assessed based on fitness, precision and F-score metrics. Only BPMN process model is assessed using quality metrics, because it cannot be done on process maps. The output of the process discovery stage is the process model of the loan application process, which is used in the following stages for process analysis and process simulation to identify suitable candidates for RPA implementation.

4.3.3 Process analysis

At this stage, the acquired process model is analysed in relation to introduced RPA characteristics (Sect. 4.1), which are crucial for RPA implementation. First, the characteristic of the involvement of manual work (1) was evaluated in relation to the processing times of activities in the log, as non-automated activities show processing times of activities other than instantaneous. The characteristic of frequency-significant activities (2) is based on the discovered process model enriched with a frequency perspective. The log was searched for activities and pathways showing significant values in relation to the total number of cases in the log. The characteristic of productivity and efficiency (3) is based on a process model enriched with a

¹ Fluxicon Disco software can be accessed from <https://fluxicon.com/disco/>.

² Apromore software can be accessed from <http://apromore.org/platform/tools/>.

performance perspective. Productivity refers to the processing times of the activities performed. Efficiency is related to the waiting times of individual activities, also resulting from a process model enriched with a performance perspective. In addition, the overall impact on the whole process will also be monitored in terms of the total times that individual activities and traces take over in the records. The error-prone and rework characteristic (4) was monitored through unit-length loops in process models. Furthermore, this characteristic will be evaluated on the basis of the number of repetitions of individual activities in the records. The use and involvement of resources will also be monitored through so-called dotted charts and graphs of the dependence of time stamps on the number of active cases. The characteristics of process flow and logic (5) are assessed using BPMN process model, where we were looking for specific patterns like, e.g., AND gateways. The outputs of this stage are identified tasks suitable for RPA implementation. These tasks are then assessed in the following stage using simulation of the loan application process.

4.3.4 Process simulation

At this stage, the simulation model is created for the assessment of the impact of RPA implementation within the observed process. For the simulation purposes we adopted approach presented by Mărușter and van Beest (2009); Rozinat et al. (2009a, b) Firstly, the discovered process model was further adjusted for the simulation purpose using filters nodes, arcs and parallelism. The purpose was to obtain simpler version with a focus on the preservation of the maximum number of activities in the log and bottlenecks discovered in the previous stage, while preserving sufficient level of quality of the model. The decision points (e.g., XOR gateways) within the model were simulated as percentages based depending on the frequency of outgoing arcs (see Fig. 2). In other words, probabilities come out as a mathematical quotient of cases with respect to those affected by the selected path. The number of business instances is derived directly from the original log. As mentioned previously, the number of cases within the original log is 13,087. The maximum number of process instances allowed in BIMP is 10,000. Within the simulation model, we work with 10,000 cases as the number of cases is high enough to properly represent the workflow of the process.

The time perspective of the simulation model is derived also from the original log. There are three time-related parameters: arrival of new cases, processing times and waiting times. Parameter arrival of new case represents the point in time when the first event of the case occurs. The arrival of new cases is typically considered a Poisson process and thus, arrival times are modelled with Poisson distribution, which was estimated from the original log. Parameter processing time represents the time required to execute each task. The probability distributions used for estimation of processing times of each activity within the log were gamma distribution, lognormal distribution, normal distribution, exponential distribution and Weibull distribution. The probability distribution was chosen based on how well the distribution can fit the data using Kolmogorov–Smirnov statistics, Cramer-von Mises statistics and standard error. If two distributions had value of Kolmogorov–Smirnov statistics same to the hundredths, the distribution with

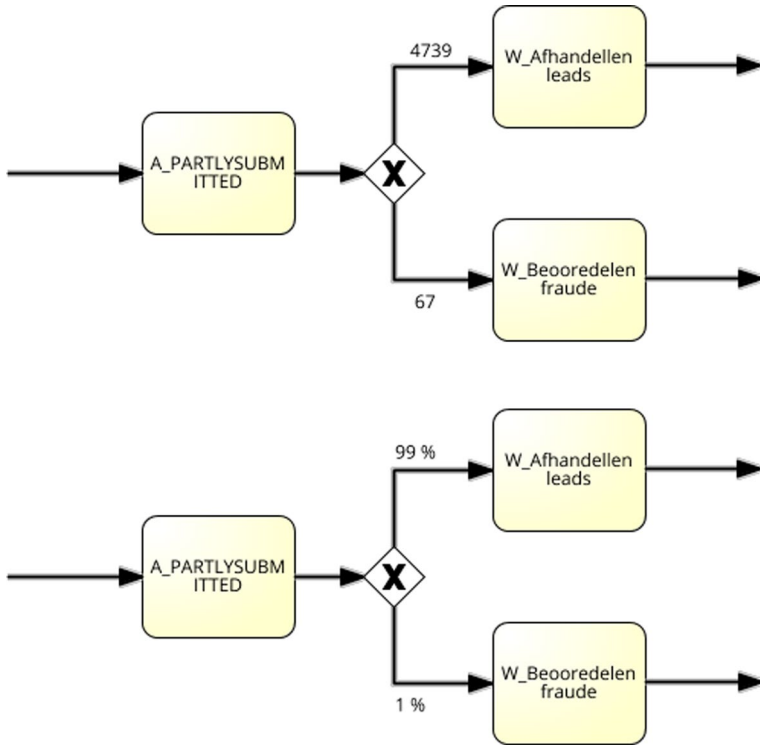


Fig. 2 Simulation of decision points in the loan application process

Table 2 PPAFR assessment framework

RPA characteristics	Monitored parameters
Frequency-significant activities	<ul style="list-style-type: none"> • Number of activity executions • Number of activity executions in successfully completed cases
Error-prone and reworks	<ul style="list-style-type: none"> • Number of reworks • Processing and waiting times of reworks
Involvement of manual work	<ul style="list-style-type: none"> • Involvement of at least one software application* • Number of users executing each activity • Reduction of workload (FTE)
Productivity and efficiency	<ul style="list-style-type: none"> • Average processing times of activities • Average waiting times of activities • Total processing times of activities • Total waiting times of activities • Potential time savings • Average cycle time
Process flow and logic	<ul style="list-style-type: none"> • Number of involved software applications* • Number of predecessors and successors* • Percentage of paths of predecessors and successors in the monitored activity*

lower Cramer-von Mises statistics was chosen. In some cases, the best fitted distribution had very high standard error resulting in mean value being significantly off. In such cases, standard error was considered when selecting the distribution. The same probability distributions were also used for estimation of waiting times and the same procedure was applied as well. The waiting time represents the time required for activity to start after the precedent activity ended. Each activity in the log has lifecycle transition consisting of two states “Start” and “Complete”. Thus, the processing time is a difference between the complete timestamp and start timestamp of each activity, while the waiting time is a difference between the start timestamp of an activity and the complete timestamp of a precedent activity. Activities with instant processing time in the original log are activities that have equal complete and start timestamps, and the same applies to the simulation model. Activities that have instant waiting times in the original log are activities that have average waiting times in the order of milliseconds. We approximate average waiting times that are in the order of seconds to be instant with a value of waiting time equal to 0.001 s.

The organisational perspective of the simulation model is derived from the original log. The organisational perspective deals with the assignment of resources to the performed activities. Based on the data, most of the resources perform most of the individual activities. The difference is the degree to which each resource performs each individual activity, meaning how many times throughout the log the resource performed the activity. Activities with instant processing times are assigned resource “System”. To assign each non-instant activity one type of resource, we use k-means clustering. The clustering is based on the profile of each resource. The profile of a resource is a one-dimensional vector, where each row represents how many times the resource executed each non-instant activity. The optimal number of clusters required for k-means clustering is derived based on Elbow method, Silhouette method and Gap statistics. The optimal number of clusters was determined to be 6. The cluster was assigned to each activity based on the maximum of total number of executions of activity by each resource within the cluster (see Eq. 1).

$$\max_n \sum_i e_i^n \quad (1)$$

n is the number of clusters, i is the number of resources within the cluster, e_i^n is the number of times activity a was executed by resources i belonging to cluster n .

The closer the approximation, the more reliable is the prediction of the to-be model. Since the intention is to obtain a simulation model similar to the real data, the following indicators of similarity are used:

- Process flow and semantics: Both models should show the same process flow expressed as ordering of activities and BPMN constructs. In addition, with respect to semantics, the activity labels should be identical.
- Throughput times: Throughput times per activity for the discovered process model should be comparable with the simulated process model.

- Bottlenecks: The location and severity of bottlenecks of the process model should be comparable with the location and severity of bottlenecks of the simulation model.

4.3.5 Assessment of RPA candidates

The last stage consists of suggestions of suitable candidates for RPA implementation with focus on productivity and efficiency of the process. Based on the results obtained in the previous stages, the last stage will present the processes that are based on the results of process mining analysis suitable for the deployment of RPA in connection with the evaluation of process characteristics. Individual solutions will be achieved by expanding or improving the existing as-is process model by deploying RPA. In connection with the redesign of the as-is process model, the case study will focus on the following strategies:

- improving the customer experience;
- cost reduction;
- improving process efficiency;
- more efficient use of resources;
- shortening the response time.

4.4 PPAFR assessment framework design

Factors critical for RPA implementation are not specific enough to be used for assessment of activities suitable for the implementation of RPA. Thus, RPA characteristics derived from factors critical for RPA implementation are also not specific enough. To the best of our knowledge, there does not exist any proposal of process performance indicators (PPIs) that allow to assess suitability of process activities for RPA implementation with focus on the overall performance of the process. In order to make the derived RPA characteristics measurable using process mining techniques, we propose PPAFR assessment framework presented in Table 2. The PPAFR assessment framework was designed with the following principles in mind (del-Río-Ortega et al. 2013):

- Four requirements for the definition of PPIs can be established: (1) expressiveness, (2) understandability, (3) traceability with business process and (4) possibility to be automatically analysed.
- It is recommended that the PPIs satisfy the SMART criteria (Specific, Measurable, Achievable, Reliable and Time-bounded).
- Must be possible to use with different business process modelling languages.

The characteristics and consequently also monitored parameters in Table 2 may be of dual nature. Some parameters have a mandatory nature like, e.g., involvement of at least one software application which does not require simulation to determine suitability of activity for RPA implementation. Some parameters have a nature of

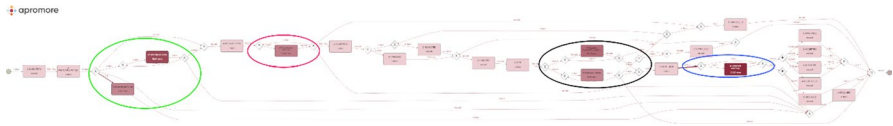


Fig. 3 BPMN process model of the loan application process (values of filters nodes, arcs and parallelism set up to 100, 60, 100)

outcome measure and thus the simulation is required. Mandatory parameters not requiring the simulation are marked with “*” in Table 2.

5 Demonstration of PPAFR assessment framework

5.1 Process discovery

The discovered process map captures the overall behaviour of the loan application process described in Sect. 4.2. The loan application process is complex and unstructured; thus, it is necessary to apply discovery techniques to acquire appropriate process model, which can be analysed. Figure 3 is an illustration of BPMN process model of loan application process discovered by Apromore with parameters nodes, arcs and parallelism set up to 100, 60 and 100. Further analysis of activities suitable for RPA implementation is based on the discovered model. We gradually analyse the discovered process model regarding the RPA characteristics presented in Sect. 4.1.

5.2 Process analysis

Firstly, we checked which activities meet the characteristics of involvement of manual work within one or more software applications. This results in 7 activities that are further analysed: “W_Handle leads”, “W_Assess potential fraud”, “W_Call incomplete files”, “W_Complete application”, “W_Call after offers”, “W_Validate application” and “W_Change contract details”. The involvement of manual work is considered a mandatory characteristic of activities that are being considered for the implementation of RPA. For the illustration purpose, highlighted parts of the Fig. 3 represent the main bottlenecks of the loan application process. In the green oval in Fig. 3, activities “W_Assess potential fraud” and “W_Handle leads” are highlighted. Activity “W_Assess potential fraud” is an activity in which an application for a fraud loan is evaluated. Activity “W_Handle leads” is related to the submission of an incomplete initial application. In the red oval in Fig. 3, activity “W_Call incomplete files” is highlighted. Activity “W_Call incomplete files” is an activity of completing pre-accepted applications. In the black oval in Fig. 3, activities “W_Call after offers” and “W_Complete application” are highlighted. Activity “W_Call after offers” is concerning the submission of offer to a qualified applicant. As part of activity “W_Complete application”, additional information is obtained during the

Table 3 Characteristics of frequency-significant activities, productivity and effectivity, and error-prone and reworks of activities highlighted as bottlenecks

Activity	Frequency	Processing times (min)	Frequency of waiting times	Waiting times	Frequency of reworks	Waiting times
W_Handle leads	4755	16.94	5502	5.15 h	763	56.97 min
W_Assess potential fraud	108	9.27	67	1.4 days		
W_Call incomplete files	7367	10.02	7367	5.85 h	4396	1.17 days
W_Call after offers	5011	9.08	5009	2.83 days	4207	3.27 days
W_Complete application	1647	12.71	357	36.59 min	1468	17.13 h
W_Validate application	3210	20.91	3208	2.44 days	1438	10.53 h

assessment stage of the application. Activity “W_Validate application” is an application review activity. After that, we focused on the analysis of processing and waiting times, frequencies of occurrence of individual activities and reworks represented in the model as loops with length 1. We will therefore involve a combination of three characteristics, namely frequency-significant activities, productivity and efficiency, and error-prone and rework. Based on the process map (values of filter activities and path equal to 100 and 5.2), activity “W_Change contract details” acts as a bottleneck of the process with mean waiting time of 30.7 days. However, on closer inspection the activity is an exceptional behaviour as it occurs within the log only 12 times and the waiting time of 30.7 days is based only on 4 occurrences; thus, the activity is no longer considered in the analysis, because it is not suitable for RPA implementation.

Table 3 presents the summary of frequencies, processing times, waiting times and waiting times of reworks from previous paragraph of activities highlighted as bottlenecks of the loan application process. Based on Table 3, the activity “W_Assess potential fraud” is excluded from consideration of RPA implementation. The activity is not a frequency-significant activity and its impact on productivity and efficiency is negligible. Moreover, it is not reworked. If we consider the overall process and built in Apromore filter, then with the exception of activity “W_Handle leads”, the rework is done in more than 10% of occurrences of each activity. Namely, 5.83% reworks of activity “W_Handle leads”, 10.99% of reworks of activity “W_Validate application”, 11.22% of reworks of activity “W_Complete application”, 32.15% of reworks of activity “W_Call after offers” and 33.59% of reworks of activity “W_Call incomplete files”. In the context of Table 3, the rework rate of individual activities is even worse. Namely, 16.05% of reworks of activity “W_Handle leads”, 44.80% of reworks of activity “W_Validate application”, 59.67% of reworks of activity “W_Call incomplete files”, 83.96% of reworks of activity “W_Call after offers” and 89.13% of reworks of activity “W_Complete application”. Similar rates of rework occur if we consider only successfully completed cases.

Table 4 Total processing times, total waiting times and rework waiting times of activities highlighted as bottlenecks

Activity	Processing times	Waiting times	Rework waiting times
W_Handle leads	2.31 months	2.79 years	1.49 months
W_Call incomplete files	5.56 months	4.92 years	52.63 years
W_Call after offers	4.83 months	51.25 years	145.78 years
W_Complete application	3.36 months	1.44 weeks	17.65 years
W_Validate application	3.82 months	22.82 years	3.51 years

The activity "W_Handle leads" has an average processing time of 17.80 min in the case of the first occurrence, while the average processing time in the case of its repeated occurrence in the case is 13.34 min. The ANOVA test and the Kolmogorov–Smirnov test were used to compare the two samples, where $p_{ANOVA} = 0,6720$ and $p_{KS.test} = 1.649 \cdot 10^{-13}$. The activity "W_Call incomplete files" has an average processing time of 13.04 min in the case of the first occurrence, while in the case of its repeated occurrence in the case the average processing time is 7.70 min, where $p_{ANOVA} = 3.358 \cdot 10^{-6}$ and $p_{KS.test} = 2.2 \cdot 10^{-16}$. The activity "W_Complete application" has an average processing time of 17.83 min in the case of the first occurrence, while in the case of its repeated occurrence in the case the average processing time is 17.31 min, where $p_{ANOVA} = 0.9248$ and $p_{KS.test} = 2.2 \cdot 10^{-16}$. The activity "W_Call after offers" has an average processing time of 15.00 min in the case of the first occurrence, while in the case of its repeated occurrence in the case the average processing time is 313.56 min, where $p_{ANOVA} = 2.2 \cdot 10^{-16}$ and $p_{KS.test} = 2.2 \cdot 10^{-16}$. The activity "W_Validate application" has an average processing time of 21.80 min in the case of the first occurrence, while in the case of its repeated occurrence in the case the average processing time is 19.56 min, where $p_{ANOVA} = 0.3228$ and $p_{KS.test} = 2.2 \cdot 10^{-16}$. This has an impact on the interpretation of the results of the simulation models, as the simulation model is based on a further simplified discovered model, in which loops of length 1 do not occur for individual activity types.

Table 4 presents total processing times, waiting times and waiting times of reworks of activities highlighted as bottlenecks in the loan application process. Total processing time of an activity is a sum of all processing times of each activity; similarly, total waiting times and waiting times of reworks are derived. Table 4 shows that "W_Handle leads" is the least demanding activity, while "W_Call after offers" is the most demanding activity. Table 5 presents frequencies, processing time and waiting times of activities highlighted as bottlenecks in the loan application process. Table 5 shows that average processing times of activities are similar to those presented in Table 3. Correspondingly, average waiting times of activities of successfully completed cases are also similar to the average waiting times presented in Table 3. If we consider waiting times of reworks, it is apparent that with the exception of activities "W_Handle leads" and "W_Call incomplete files", waiting times are again similar to those in Table 3. Thus, the problematic areas of the process are not limited to unsuccessfully completed cases.

Table 5 Characteristics of frequency-significant activities, productivity and effectivity, and error-prone and reworks of successfully completed cases of activities highlighted as bottlenecks

Activity	Frequency	Processing times	Frequency of waiting times	Waiting times	Frequency of reworks	Waiting times
W_Handle leads	573	18 min	572	5.48 h		
W_Call incomplete files	2246	9.36 min	2246	5.39 h	1101	7.24 h
W_Call after offers	2246	8.5 min	2245	2.81 days	1768	3.13 days
W_Complete application	1231	11.51 min	306	37.94 min	1090	16.25 h
W_Validate application	2246	20.91 min	2245	2.41 days	1049	10.08 h

Figure 4 shows the number of events in the log distributed over time. At the highest point, there were 895 events, while at the lowest point, there were 583 events. It is apparent that the number of events in time across the reference period is relatively steady. Table 6 presents the number of events of highlighted activities that occurred in individual months of the observed period. Next, we divided the log into 5 periods, where first four periods last 33 days each and the last period lasts 35 days, and we discovered the process model for each period. Then we compared the process model discovered based on the entire log with process model of each period. The bottlenecks of each period correspond to the bottlenecks discovered over the entire log with no significant divergences in processing and waiting times (Fig. 5). Thus, on a monthly basis, there are no significant deviations in the performance of the process, even though there are some small deviations in the number of occurring events throughout the observed period.



Fig. 4 Number of events in the log depending on time

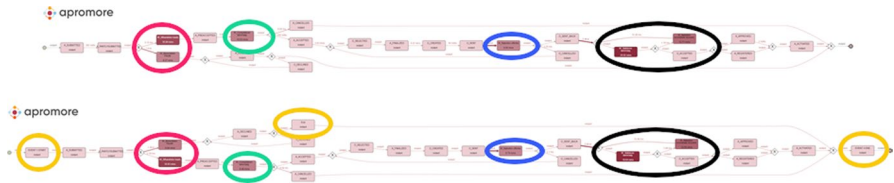


Fig. 5 Process flow of discovered (upper part) and simulation (bottom part) model

Furthermore, the occurrence of events related to the monitored activities is very similar over time in terms of the trend of growth and decrease in the number of events associated with individual activities. In these cases, RPA is a very suitable tool for capturing fluctuations in demand, as it can be scaled very easily. However, none of the pre-selected activities differs fundamentally in this respect in terms of declining and rising trends. For this reason, the waiting and processing times of individual activities are also decisive here, since even in terms of resources, the respective activities are performed by essentially all resources from the group of resources assigned to the process. This is also indicated in Table 6, which presents the number of events associated with selected activities in individual months of the observed period. Likewise, over time there are no changes in the distribution of bottlenecks, or their severity in connection with the procedural and waiting times of the monitored activities. A total of 56 resources ensure the whole process and each activity highlighted as bottleneck is ensured by 50 and more resources. In turn, it can be deduced that the activities do not have a high degree of specialisation. Lastly, process flow and process logic are important to avoid limiting the effect of RPA deployment in the process. It is therefore necessary to check whether the selected activities do not occur in patterns that reduce the effectiveness of RPA due to the existing dependencies between the activities in the process. In this respect, it is therefore necessary to monitor whether the selected activities are not in certain patterns, especially if they are activities located inside the AND or OR gates. However, the activities monitored by us occur in the patterns of XOR gates, which, on the contrary, logically exclude the dependence of the respective activities in this direction. Standardisation is also connected with the characteristics of process flows and the logic of the process, which evaluates the level of structure of the process. If we look at the individual activities considered for the deployment of RPA, we get the following:

- “W_Handle leads”—the activity has 3 predecessors and 5 successors. Two predecessors occur in 99.71% of all routes, while three successors occur in 99.60% of all routes.
- “W_Call incomplete files”—the activity has 7 predecessors and 9 successors. Two predecessors occur in 99.03% of all routes, while four successors occur in 98.65% of all routes, of which two activities account for 80.06% of all routes.

Table 6 Frequency of occurrence of highlighted activities in individual months of the observed period

Activity	October 2011	November 2011	December 2011	January 2012	February 2012	March 2012
W_Handle leads	797	1237	909	1264	1172	230
W_Call incomplete files	1367	2080	1856	2159	2375	816
W_Call after offers	966	1505	1380	1556	1729	689
W_Complete application	327	510	448	531	594	204
W_Validate application	661	976	835	968	1044	322

- “W_Call after offers”—the activity has 7 predecessors and 10 successors. Two predecessors occur in 99.20% of all routes, while four successors occur in 89.44% of all routes, of which two activities account for 70.41% of all routes.
- “W_Complete application”—the activity has 5 predecessors and 12 successors. Three predecessors occur in 99.94% of all routes, of which two activities account for 89.31% of all routes. Five successors account for 90.32% of all trips, of which three activities account for 83.19% of all trips.
- “W_Validate application”—the activity has 8 predecessors and 13 successors. Three predecessors occur in 99.37% of all routes, while seven successors occur in 91.57% of all routes, of which four activities account for 75.60% of all routes.

As can be seen, all activities have a high level of standardisation. The most standardised activities are “W_Handle leads” and “W_Call incomplete files”, while another standardised activity is “W_Validate application”. However, for all activities, three or fewer predecessors occur in more than 99% of all these routes, and four or fewer predecessors occur in more than 75% of all subsequent routes. The number of different route variants in the record should also be considered, which is 4366, which is a very high number and would suggest high non-standardisation of the whole process. However, this number is very misleading if several loops appear in the record, including the loops of length 1, which are behind such a large number of different variants. However, the 31 most common variants account for 54.6% of all cases, which indicates further good standardisation of the process.

5.3 Process simulation

First of all, it is necessary to verify the fulfilment of indicators for assessing the suitability of the simulation model, which were presented in Sect. 4.3.4—Research stages. The process flow and semantics of the discovered and simulation model are the same except for the initial and final activities. Initial and final activities are artificially inserted into the simulation model; however, these activities do not affect the process logic, nor do they affect process performance or process bottlenecks, as they are simulated as instant activities that have instant processing and waiting times (see Fig. 5—orange ovals). The discovered BPMN process model was used to build a simulation model. The discovered process model is composed of 23 activities, while the simulation model is composed of 26 activities. Figure 5 represents the process flows and bottlenecks of the discovered and simulation model. The red oval highlights the activities “W_Handle leads” and “W_Assess potential fraud”, the green oval highlights the activity “W_Call incomplete files”, the blue oval highlights the activity “W_Call after offers” and the black oval highlights the activities “W_Application assessment” and “W_Validate application”. The BPMN model, on which the simulation model is based, has the following quality values: 0.75 for fitness, 0.76 for precision, and 0.76 for F-score. The model shows a lower quality; however, this is mainly due to the absence of length 1 loops, which are not considered in this model. Loops of length 1 also negatively affect other discovery techniques. The quality of

Table 7 Throughput times of the original log and simulation model log

	W_Handle leads	W_Assess potential fraud	W_Call incomplete files	
Original log	Processing times	16.94 min	9.27 min	10.02 min
	Waiting times	5.15 h	1.40 days	5.85 h
Average of simulation logs	Processing times	16.94 min	9.87 min	9.47 min
	Waiting times	5.20 h	1.38 days	5.74 h
Std Dev of simulation logs	Processing times	0.36 min	0.79 min	0.12 min
	Waiting times	0.06 h	0.13 days	0.08 h
	W_Call after offers	W_Complete application	W_Validate application	
Original log	Processing times	9.08 min	12.71 min	20.92 min
	Waiting times	2.83 days	1.78 h	2.44 days
Average of simulation logs	Processing times	8.76 min	12.83 min	19.63 min
	Waiting times	3.06 days	1.70 h	2.82 days
Std Dev of simulation logs	Processing times	0.10 min	0.16 min	0.20 min
	Waiting times	0.03 days	0.03 h	0.01 days

Table 8 Estimated probability distributions of chosen activities

Activities	Probability distributions	
	Processing times	Waiting times
W_Handle leads	Gamma	Exponential
W_Assess potential fraud	Exponential	Exponential
W_Call incomplete files	Exponential	Weibull
W_Call after offers	Exponential	Exponential
W_Complete application	Exponential	Weibull
W_Validate application	Weibull	Normal

the model is therefore sufficient. In addition, loops of length 1 are considered separately in the analysis of the results.

A comparison of the bottlenecks of the discovered model and the simulation model is also presented in Table 7. These are the average values of the process and waiting times of 15 simulation runs. The severity of bottlenecks in the case of the simulation model is similar to the situation in the discovered process model. According to Fig. 5 and Table 7, it can be stated that the indicators for the assessment of the simulation model are satisfactory, namely the process flow and semantics, process throughput and bottlenecks are adequate in both models.

Table 8 describes estimated probability distributions of processing and waiting times of chosen individual activities that provide the best fit based on data from the log. Exponential distribution and Weibull distribution are the most used distributions. Exponential distribution prevails among both processing and waiting

Table 9 Simulations of to-be process model after implementation of RPA

Activity		Average case duration	Reduction of average case duration	Relative frequency	Reduction of workload	Potential time savings
Partial	W_Call incomplete files	6.9067	0.2119	20.75	1,562.23	–
	W_Validate application	6.9053	0.2312	9.35	2,664.08	–
	W_Call after offers	6.9133	0.1156	20.75	1,188.39	–
	W_Assess potential fraud	6.9000	0.3082	0.05	16.78	–
	W_Complete application	6.8714*	0.7214	20.75	997.22	–
	W_Handle leads	6.8688*	0.7590	20.75	2,794.32	–
Full	W_Call incomplete files	6.6733*	3.5831	15.90	1,562.23	58,097.92
	W_Validate application	4.2753*	38.2296	0	2,664.08	524,348.32
	W_Call after offers	3.3587*	51.4737	15.07	1,188.39	570,132.49
	W_Assess potential fraud	6.9033	0.2601	0	16.78	3,393.38
	W_Complete application	6.7949*	1.8267	15.39	997.22	8,878.23
	W_Handle leads	6.6540*	3.8625	15.90	2,794.32	54,136.68

*We used ANOVA, thus, significant differences among group means were calculated using F statistic, with level of significance equal to 5%. F statistic is the ratio of variation between samples to variation within the samples

times. Weibull distribution is the second most used distribution. Processing time of activity “W_Handle leads” was estimated as gamma distribution and waiting time of activity “W_Validate application” was estimated as normal (Gaussian) distribution.

Table 9 presents the impact of partial and full automation on the performance of the discovered process through the deployment of RPA. Within the simulation model, partial implementation of RPA is characterised by the elimination of process times of selected activities, while full implementation of RPA is characterised by the elimination of process and waiting times of selected activities. This option allows us to accommodate for external factors, e.g., customers interacting with the process. Thus, Table 9 represents a total of 12 scenarios (6 scenarios for partially automated selected activities and 6 scenarios for fully automated selected activities), where the results for each scenario are the average of 15 simulation runs. The average duration of a case indicates the time needed to carry out all activities in the case and is shown in days. The reduction of average case duration indicates the reduction in the average case duration after the deployment of RPA in the process compared to the

simulation model in percent. The relative frequency of resource involvement indicates the percentage of resource utilisation in relation to individual process activities. Workload reduction refers to time savings due to partial automation through the deployment of RPA in hours and relates to processing times of activities. Potential time savings indicates the potential time savings associated with full automation in hours and relates to processing and waiting times of activities.

Table 9 shows the impact of RPA implementation within the simulation model with respect to individual scenarios. For partial implementation of RPA, the impact on the overall case duration is very low. Typically, in selected activities, there was a change in the order of tenths of a percent, of which four times this change was not statistically significant. Statistically significant changes at the significance level $\alpha = 5\%$ are marked with an "*" for the average case duration. With the partial implementation of RPA, there were statistically significant changes in the activities "W_Complete application" and "W_Handle leads" by 0.7214% and 0.7590%, respectively. For the full implementation of RPA, all changes are statistically significant, except for the activity "W_Assess potential fraud", which is not frequency significant. The largest changes were achieved in the activities "W_Call after offers", "W_Validate application" and "W_Handle leads", namely by 51.4737, 38.2296 and 3.8625%, respectively. For the relative frequency of resource, there was the largest change in the activity "W_Validate application" by 9.35% and the smallest change in the activity "W_Assess potential fraud", namely by 0.05%. Changes in the remaining activities were around 5%. In terms of reduction of workload, the implementation of RPA achieves the best results for the activities "W_Handle leads", "W_Validate application" and "W_Call incomplete files". For potential time savings, the best results are achieved for the activities "W_Handle leads", "W_Validate application" and "W_Call incomplete files".

5.4 Assessment of RPA candidates

As already mentioned, the log contained a total of 24 different activities. The selection of suitable candidates was made on the basis of the following characteristics: involvement of manual work and software application, frequency of significant activities, productivity and efficiency, susceptibility to errors and rework, and process flows and logic. After applying the characteristics of the involvement of manual work and software application, the number of activities considered was reduced to 7 activities: "W_Handle leads", "W_Assess potential fraud", "W_Call incomplete files", "W_Complete application", "W_Call after offers", "W_Validate application" and "W_Change contract details".

Activities that do not meet the mandatory characteristics were excluded, as there is no point in considering them for the implementation of RPA. Subsequently, the frequency-significant characteristic is applied to the considered activities, on the basis of which the activity "W_Assess potential fraud" occurring only 108 times and the activity "W_Change contract details" occurring only twice are excluded. In the next step, we focused on the efficiency characteristic represented by waiting and processing times. If we focus on the processing times of individual activities, we can see in

Table 3 that the processing times of individual activities ranges between 9.08 and 20.91 min, while waiting times range between 36.59 min and 2.83 days. It is therefore more sensible to consider waiting times for activities here in the first place. However, it is also necessary to consider to which extend waiting times are affected by internal or external factors. For the activity "W_Complete application", the frequency of occurrence of waiting times on the main flow is only 357, and in addition, considering the nature of the activity, it is clear that the waiting times are largely due to the applicant. We cannot therefore expect major improvements in this area here. As a result, we will exclude the activity "W_Complete application" from further analysis. If we look at the activity "W_Handle leads", we see that in terms of efficiency, the activities "W_Call incomplete files", "W_Call after offers" and "W_Validate application" are several times more serious. Furthermore, if we consider the characteristics of susceptibility to errors and reworking, the activity "W_Handle leads" is by far the best. Therefore, we will keep these 3 activities for the following analysis. Considering the characteristics of productivity, it should be mentioned that the remaining activities are provided between 51 and 53 employees (in addition, with a few exceptions, they are the same employees), and activities also respond similarly to fluctuations in demand. In terms of productivity and potential savings, the remaining activities are very similar. In terms of process flows and logic, the mentioned activities do not occur in patterns that make one activity dependent on another, and thus do not limit them. Typically, it is desirable to make changes within the flows that lead to the successful completion of the cases. If we perform this analysis only on successfully filtered cases (Table 5), we arrive at the same conclusions.

We will reach similar conclusions on the basis of process simulations (see Table 8). As can be seen, the highest impact on the duration of the process is recorded for the activities "W_Call after offers", "W_Validate application", "W_Handle leads" and "W_Call incomplete files". Of these activities, "W_Call after offers" and "W_Validate application" have the highest impact on the performance of the process, in all categories of Table 8. Implementation of RPA for the activity "W_Handle leads" has a higher impact on reducing the duration case and there is a higher reduction of workload; however, the activity "W_Call incomplete files" provides higher potential time savings. Moreover, the activity "W_Handle leads" has a frequency of reworks 763 according to Table 7 with an average waiting time of 56.97 min and with an average processing time of 13.34 min, and the activity "W_Call incomplete files" has a frequency of reworks of 4,396 with an average waiting time of 1.17 days and an average processing time of 7.70 min. Ultimately, the implementation of RPA for the "W_Call incomplete files" activity will provide higher process performance. Emphasis is placed on the efficiency of the process, so the resulting recommendations are based on the resulting effects of the full RPA implementation. In the case of partial implementation of RPA, the impact on the duration of case is statistically significant only for the activities "W_Complete application" and "W_Handle leads". However, in the case of partial implementation of RPA, the impact is only in tenths of a percent for all monitored activities. It is therefore better to consider here primarily the reduction of workload. In this case, according to Table 8, the highest values are again reached by the activities "W_Handle leads", "W_Validate application", "W_Call incomplete files" and "W_Call after offers".

However, it should be mentioned here again that while for the activity "W_Handle leads" the frequency of rework is only 763 with an average waiting time of 56.7 min and an average processing time of 13.34 min, for the activities "W_Validate application", "W_Call incomplete files" and "W_Call after offers" the frequency of reworks are 1,438, 4,396 and 4,207, respectively, with significantly longer average waiting times of 10.53 h, 1.17 days and 3.27 days, respectively, and average processing times of 19.56 min, 7.70 min and 313.56³ min, respectively. Loops of length one in this case will affect both the average duration of the cases and the reduction of the workload. This again corresponds to the selection of suitable candidates for RPA below.

Based on the performed analysis and the results of simulation experiments, the activities "W_Call incomplete files", "W_Validate application" and "W_Call after offers" were selected as candidates. It is not recommended to start with the activities "W_Handle leads" and "W_Complete application", as they do not have as serious an impact on the process as the selected candidates. Once the selected candidates have been automated, it is also possible to automate these activities in the order listed. The activities "W_Assess potential fraud" and "W_Change contract details" are not recommended for automation.

6 Discussion of the PPAFR assessment framework

6.1 Discussion of the PPAFR with regard to selected activities

As part of the activity "W_Call incomplete files", applications for loan approval are completed before their acceptance. These are activities related to manual document editing, which are generally suitable for RPA deployment. Activities of this type occur in the main flow of approval processes, regardless of the financial or banking institution. The implementation of RPA can be expected to reduce the human error rate caused by actions such as data entry, copying and pasting, miscalculations etc., and thus the need for reworking. As this is an internally performed manual activity associated with documents, its throughput would be significantly improved, which would also affect the throughput of the entire process due to the reduction of processing and waiting times. In the case of procedural times, they would be completely eliminated, while in the case of waiting times, they would be shortened. In addition, the activity is carried out by 53 different employees working in two shifts. The involvement of RPA would reduce the FTE costs of this activity and would also allow employees to shift their focus to other activities. Although the activity based on simulations is only in the fourth place, the simulation model does not consider loops of length 1, and therefore this activity was preferred to the activity "W_Handle leads". Discovered model contains traces with following patterns <a,b,c,d,d,e,f>, where sub-trace <d,d> represents loop of length 1. To limit the number of different variants of the process, we omitted loops of length 1 from simulation model,

³ This is due to several remote values, the average process time of the activity "W_Call after offers" across all occurrences is 9.08 min.

especially since it is straightforward to account for them without the need to simulate them.

The “W_Validate application” activity is related to the assessment of loan applications and whether the applicant obtains them. This activity takes place in SUP processes, regardless of which financial or banking institution it is. In the case of automation, the validation of each application would be forced, as the approval of the loan should cover all applications. Due to the fact that this is again an internally performed activity associated with document editing, there would be a significant improvement in the throughput of the activity, and therefore the throughput of the whole process in connection with shortening the process and waiting times of the activity itself. In the case of procedural times, they would be completely eliminated, while in the case of waiting times, they would be shortened. In addition, the activity is carried out by 50 different employees working in two shifts. Involving the RPA would not only reduce the FTE’s cost of performing this activity, but it would release these employees as well.

The activity “W_Call after offers” represents the contact of the applicant after the offer has been made. Following the monitored characteristics, this is an activity that has considerable potential for the deployment of RPA. However, the nature of the process and the activity itself do not clearly define the actions needed to carry out the activity. Deploying RPA for cognitively demanding tasks is not appropriate. These include tasks that require judgement, interpretation, assessment of results, unstructured tasks, creative tasks, and tasks demanding for empathy and social interactions. In addition, it is not a purely internally performed activity, the procedural and waiting times depend on the applicant. However, the activity “W_Call after offers” has such a significant impact on the productivity and efficiency of the process that partial automation could be considered to increase the throughput of the activity, and thus the whole process. Here, too, there would be savings in the FTE costs associated with this activity, as well as the redundancy of staff, considering the fact that the activity is carried out by 52 different employees.

6.2 Discussion of the PPAFR with regard to business process simulations

In the simulations, emphasis was placed on the fact that the individual activities show all the characteristics. The more problems the RPA implementation solves, the easier it will be to implement it at the strategic and tactical level of management. Table 2 presents the parameters associated with the individual characteristics in terms of PPIs, which were monitored and used in the evaluation of activities suitable for the implementation of RPA. There are two major advantages to use of business process simulations. Firstly, interactions between different characteristics of RPA implementation are unclear and it is assumed that future developments of RPA will bring further characteristics. And the combination of different characteristics makes business processes relatively complex systems. Furthermore, it is much easier to follow the complexity of the system using business process simulations. Moreover, use of simulations makes the PPAFR assessment process and framework more generalisable, as business process simulations are applicable in many different areas.

Moreover, the research in the area of business process simulations is still continuing, e.g., Camargo et al. (2020), Pourbafrani et al. (2020).

In addition to the parameters listed in Table 2 in the analysis stage, attention was also paid to the stability of the process, i.e. there have been no major changes in the process model over time. Although the stability of the process does not have a direct impact on its performance, if there were changes in the model, automation would not make sense, as there would be changes in the predecessors and successors of the monitored activities. Furthermore, fluctuations in the number of occurrences of events represented by individual events were monitored, and even these did not have a significant impact on process performance, as all activities reacted similarly to fluctuations in the number of active cases, i.e. if there was an increase in the number of occurrences of one activity, there was an increase in the number of occurrences of remaining activities and vice versa. In the same way, there were no major fluctuations associated with changes in waiting times. In the future research, the stability of the process offers the use of technique called concept drift (Adams et al. 2021; Ceravolo et al. 2020) and its applicability towards RPA, which needs to be examined further.

As follows from simulation experiments, the division of data for processing and waiting times also plays a crucial role, where significant distortion can occur only with the use of average times. Related to this is the fact that the BPM literature states the assumption that the processing times of manually performed activities are typically represented by a normal distribution, which is not true in the case of the records used. The probability distribution of the monitored activities showed a positive slope, so the distributions have the so-called right tail. Furthermore, it is evident from simulation experiments that if waiting times are significantly dominant in the record, it is appropriate for partial automation to focus mainly on the efficiency of the process through processing times and the associated resource allocation and workload reduction. This is because the implementation of RPA has the greatest effect, while the reductions in the average duration of cases are often negligible. On the other hand, in the overall automation, the implementation of RPA will have a significant effect not only on efficiency, but also on process productivity.

6.3 Discussion of the PPAFR with regard to process mining

There are several synergies with regard to the use of process mining capabilities to assess activities suitable for RPA implementation. In this research, PPAFR assessment framework is used for assessment of suitability of activities at the process level. However, process mining can also be used for automated generation of RPA scripts based on analysis of UI (user interaction) logs. The core idea behind the so-called robotic process mining is that repetitive routines amenable for automation can be discovered from logs of interactions between workers and Web and desktop applications in the form of UI logs (Leno et al. 2020). UI log consists of recorded interactions of users with different applications such as web, desktop, system, application, etc., which are based on mouse and keyboard inputs. These UI logs are basically event logs, which can be analysed using automated process discovery

techniques such as split miner. Thus, process mining techniques can be applicable both at the process level and task level of automated activities.

In this research, we focused on the analysis of time and quality dimensions of the Devil's Quadrangle. However, there is a gap for the use of process mining for the analysis of the cost side of the Devil's Quadrangle. The traditional costing systems use volume-based cost drivers which may have difficulty to reallocate overhead costs and costs related to technologies like, e.g., RPA. In this case, the costs should not be controlled by the volume of production, but over time in relation to the occurrence of activities in the process (Halaška and Šperka 2021). This approach to cost analysis called time-driven activity-based costing is suitable not only in the field of RPA, but in non-manufacturing companies in general due to the changing structure of the costs.

6.4 Discussion of PPAFR limitations

In this research, we focused on the analysis of aspects related to the operational efficiency of the process. However, operational efficiency is only one aspect that needs to be assessed by the management of the company. In this study, we omit the organisational aspect of RPA implementation. Thus, the PPAFR framework has to be considered in the context of organisational aspect. RPA implementation has to be aligned with goals, challenges and capabilities of the organisation. RPA is suitable in cases where organisations seeks to reduce costs, improve quality of the process, process efficiency, better compliance, integration of many different systems. Furthermore, RPA implementation shows higher satisfaction levels for more mature adopters.⁴ The PPAFR framework however does not consider BPM maturity of the organisation and RPA readiness. RPA readiness does not involve just existing technology, but also required financial and human resources, data availability, organisation's culture and even customer's readiness. Thus, there is a risk that efforts made by the organisation does not yield the benefits, especially for organisations with lower BPM maturity. However, we argue that this is a strong message for organisations for development of maturity.

Another limitation of our study is the lack of economic lens. The assessment of cost dimension of RPA implementation requires three perspectives. Firstly, it is necessary to assess the costs of to-be automated tasks. However, as discussed previously, many organisations use traditional costing systems and thus, cost assessment of certain activities might not be a trivial task. Secondly, in this research we did not focus on relevant cost drivers for cost estimation of RPA projects and how can we incorporate them in our framework. Furthermore, there are no dedicated guidelines for defining cost components in these types of projects beyond comparison with person-hours and salary cost. It is also not clear if the cost estimation used for heavy-weight automation projects are appropriate for RPA implementation projects. The

⁴ Everest group research. [online][29. 7. 2022]. Accessed from <https://www.fusionsol.com/wp-content/uploads/sites/22/2019/07/Everest-Group-UiPath-Defining-Enterprise-RPA.pdf>

third perspective which was not considered was the value of the to-be automated process. Similarly to cost drivers of RPA implementation, it is not clear what are the value drivers of RPA implementation and how they should be assessed with regard to various stakeholders, e.g., there is evidence of symbiosis for RPA and BPM, RPA and ERP, RPA and AI.

Characteristics for RPA implementation in literature are often vague. Thus, one of the objectives for the use of process mining for assessment of RPA implementation was to quantify the parameters of RPA characteristics (see Table 2). It is also not clear from the evidence to date to what extent individual characteristics of RPA affect the performance of a given process. Both can be affected by different factors. This lack of data resulted in the need of business process simulations and assessment of the RPA implementation at the process level, even though assessment at the activity level might be in many cases sufficient enough. Moreover, it might limit the usability of PPAFR framework as simulation models might not be easily attainable. We believe, this will be addressed in the future as more research is done in this area.

7 Conclusions

The aim of the research was to propose data-driven approach towards selection of business processes using process mining techniques with the intention of RPA implementation. Emphasis was placed on improving the productivity and efficiency of business processes. The following procedure was applied in assessment process: (1) data preparation, (2) process discovery, (3) process analysis, (4) process simulation and (5) assessment of RPA candidates. Emphasis was placed on improving the productivity and efficiency of business processes. In order to implement RPA, the activities considered must first meet the mandatory characteristics of manual work and the related involvement of at least one software application. Furthermore, there are important characteristics of frequency-significant activities, productivity and efficiency of monitored activities, susceptibility to errors and rework, and process flows and logic, as these have a major impact on the productivity and efficiency of the entire process. The parameters monitored within the individual characteristics are listed in Table 2, which also provides answer to the RQ1 and RQ2. All these characteristics are individually mentioned in the literature, but so far no one has dealt with their parallel assessment on one sample in a comprehensive analysis and their impact on the performance of the whole process. According to the simulation experiments, frequency-insignificant activities had a minimal impact on both the average case duration and the reduction of the workload, although the average waiting and processing times could be several times higher than for frequency significant activities. The productivity and efficiency of the monitored activities is also reflected in the overall performance of the process. If waiting times are significantly dominant in the record, it is appropriate for partial automation to focus mainly on the efficiency of the process through the process times and the related allocation of resources and reduction of the workload. This is because the implementation of RPA has the greatest effect, while the reductions in the average time of cases are often negligible. On the other hand, in the overall automation, the implementation of

RPA will have a significant effect not only on the efficiency but also on the productivity of the process, i.e., both in the reduction of the workload and the shortening of the average process durations. As follows from simulation experiment, the distribution of data for processing and waiting times also plays a crucial role, where significant distortion can occur only with the use of average times. Related to this is the fact that the BPM literature states the assumption that the processing times of manually performed activities are typically represented by a normal distribution, which is not true in the case of the records used. The susceptibility to errors and reworking is closely related to the previous characteristics and again fundamentally affects both the duration of the case and the reduction of the workload. While the average case duration in the log is 8.61 days, the average case duration for an unmodified simulation model is 6.92 days, which is primarily due to the absence of length 1 loops in the simulation model. The impact of process flows and logic is not entirely clear, as AND gates and similar constructs do not occur in simulated processes, as well as no significant branching of the process; however, their impact is implicitly monitored using simulation experiments.

To answer the RQ3, we showed the application of our framework on a case study. PM techniques can be used to find activities suitable for the implementation of RPA with an emphasis on process performance. These methods and techniques are able to detect the above characteristics in the data, provided that suitable data are available to allow the analysis of the processes. The data contained in the record should sufficiently represent the monitored process, as both the analysis stage and the simulation stage are built on them. A limitation of the case studies is that the process times were simulated independently of the company's resources, in a sense that there were always enough resources to perform an activity. As a result, it is not possible to monitor the effect of reallocation of resources in the company. Resource modeling is currently one of the most problematic areas of business process simulation. In the current form, the PPAFR assessment framework can provide the guidance to which parameters should be monitored; however, the decision about the RPA implementation is still dependent on the business process simulation. To the best of our knowledge, there does not exist assessment framework focusing on impact of RPA implementation on the performance of the overall process. Furthermore, we show what capabilities are provided by process mining techniques for such a purpose with a combination of business process simulations and we present the process of the assessment framework.

Table 10 Translation of the log activities

W_Afhandelen leads	W_Handle leads
W_Beoordelen fraude	W_Assess potential fraud
W_Completeren aanvraag	W_Call incomplete files
W_Nabellen incomplete dossiers	W_Complete application
W_Nabellen offertes	W_Call after offers
W_Valideren aanvraag	W_Validate application
W_Wijzigen contractgegevens	W_Change contract details

Acknowledgements This paper was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the Institutional Support for Long-term Development of a Research Organisation in 2022.

Authors' contributions RŠ was responsible for the design and development of the research. MH was responsible for data collection and analysis. Both were responsible for data interpretation.

Funding This paper was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the Institutional Support for Long-term Development of a Research Organisation in 2022.

Availability of data and material All the data and used software are available. They are referenced within the manuscript.

Code availability Not applicable.

Declarations

Conflict of interest We have no relevant financial or non-financial interest to disclose.

Ethical approval Not applicable.

Consent approval Not applicable.

Consent for publication Not applicable.

Appendix 1 See Table 10

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acceleration Economy (2021). State of robotic process automation. <https://accelerationeconomy.com/digital-business/state-of-rpa-survey-report/>. Accessed 11 April 2022
- Adams JN, van Zelst SJ, Quack L, Hausmann K, van der Aalst WMP, Rose T (2021) A framework for explainable concept drift detection in process mining. <http://arxiv.org/abs/2105.13155> [Cs], vol 12875, pp 400–416. https://doi.org/10.1007/978-3-030-85469-0_25
- Adriansyah A, Buijs JCAM, La Rosa M, Soffer P (2013) Mining process performance from event logs. Business process management workshops. Springer, Berlin, pp 217–218. https://doi.org/10.1007/978-3-642-36285-9_23
- Adriansyah A, Munoz-Gama J, Carmona J, van Dongen BF, van der Aalst WMP (2015) Measuring precision of modeled behavior. *Inf Syst E-Bus Manage* 13(1):37–67. <https://doi.org/10.1007/s10257-014-0234-7>
- Adriansyah A, van Dongen BF, van der Aalst WMP (2011) Conformance checking using cost-based fitness analysis. In: 2011 IEEE 15th international enterprise distributed object computing conference, IEEE, pp 28–42. <https://doi.org/10.1109/EDOC.2011.12>
- Aguirre S, Rodriguez A (2017) Automation of a business process using robotic process automation (RPA): a case study. In: Figueroa-García JC, López-Santana ER, Villa-Ramírez JL, Ferro-Escobar

- R (eds) Applied computer sciences in engineering. Springer, Berlin, pp 65–71. https://doi.org/10.1007/978-3-319-66963-2_7
- Asatiani A, Penttinen E (2016) Turning robotic process automation into commercial success—case Opus-Capita. *J Inf Technol Teach Cases* 6(2):67–74. <https://doi.org/10.1057/jittc.2016.5>
- Augusto A, Conforti R, Dumas M, La Rosa M, Maggi FM, Marrella A, Mecella M, Soo A (2019) Automated discovery of process models from event logs: review and benchmark. *IEEE Trans Knowl Data Eng* 31(4):686–705. <https://doi.org/10.1109/TKDE.2018.2841877>
- Baranauskas G (2018) Changing patterns in process management and improvement: using RPA and RDA in non-manufacturing organizations. *Eur Sci J* 14(26):251–264
- Barney J (1991) Firm resources and sustained competitive advantage. *J Manag* 17(1):99–120. <https://doi.org/10.1177/014920639101700108>
- Becker J, Uthmann CV, zur Muhlen M, Rosemann M (1999) Identifying the workflow potential of business processes. In: Proceedings of the 32nd annual hawaii international conference on systems sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers. IEEE. <https://doi.org/10.1109/HICSS.1999.772962>
- Camargo M, Dumas M, González-Rojas O (2020) Automated discovery of business process simulation models from event logs. *Decis Support Syst* 134:113284. <https://doi.org/10.1016/j.dss.2020.113284>
- Ceravolo P, Marques Tavares G, Junior SB, Damiani E (2020) Evaluation goals for online process mining: a concept drift perspective. *IEEE Trans Serv Comput*. <https://doi.org/10.1109/TSC.2020.3004532>
- Computer Economics Avasant Research (2021). Robots on the Rise: RPA investment grows. <https://www.computereconomics.com/article.cfm?id=2992>. Accessed 11 April 2022.
- Cooper LA, Holderness DK, Sorensen TL, Wood DA (2019) Robotic process automation in public accounting. *Account Horizons* 33(4):15–35. <https://doi.org/10.2308/acch-52466>
- del-Río-Ortega A, Resinas M, Cabanillas C, Ruiz-Cortés A (2013) On the definition and design-time analysis of process performance indicators. *Inf Syst* 38(4):470–490. <https://doi.org/10.1016/j.is.2012.11.004>
- Di Ciccio C, Marrella A, Russo A (2015) Knowledge-intensive processes: characteristics, requirements and analysis of contemporary approaches. *J Data Semant* 4(1):29–57. <https://doi.org/10.1007/s13740-014-0038-4>
- Dumas M, La Rosa M, Mendling J, Reijers H (2018) Fundamentals of business process management. Springer, Berlin
- Frei FX, Harker PT (1999) Measuring the efficiency of service delivery processes: an application to retail banking. *J Serv Res* 1(4):300–312
- Fung HP (2014) Criteria, use cases and effects of information technology process automation (ITPA). *Adv Robot Autom* 3:1–10
- Günther CW, van der Aalst WMP (2007) Fuzzy mining—adaptive process simplification based on multi-perspective metrics. In: Alonso G, Dadam P, Rosemann M (eds) Business process management. Springer, Berlin, pp 328–343
- Halaška M, Šperka R (2020) Importance of process flow and logic criteria for RPA implementation. In: Jezic G, Chen-Burger J, Kusek M, Sperka R, Howlett R, Jain L (eds) Agents and Multi-agent systems: technologies and applications 2020. Smart innovation, systems and technologies, vol 186. Springer, Berlin, pp 221–231. https://doi.org/10.1007/978-981-15-5764-4_20
- Halaška M, Šperka R (2021) TDABC and estimation of time drivers using process mining. In: Jezic G, Chen-Burger J, Kusek M, Sperka R, Howlett RJ, Jain LC (eds) Agents and multi-agent systems: technologies and applications 2021. Springer, Berlin, pp 489–499. https://doi.org/10.1007/978-981-16-2994-5_41
- Huang F, Vasarhelyi MA (2019) Applying robotic process automation (RPA) in auditing: a framework. *Int J Account Inf Syst* 35:100433. <https://doi.org/10.1016/j.accinf.2019.100433>
- Isik Ö, Bergh JVD, Mertens W (2012) Knowledge intensive business processes: an exploratory study. In: 2012 45th Hawaii international conference on system sciences. IEEE, pp 3817–3826. <https://doi.org/10.1109/HICSS.2012.401>
- Ivančić L, Suša Vugec D, Bosilj Vukšić V (2019) Robotic process automation: systematic literature review. In: Di Ciccio C, Gabryelczyk R, García-Bañuelos L, Hernaus T, Hull R, Štemberger MI, Kő A, Staples M (eds) Business process management: Blockchain and central and Eastern Europe forum. Springer, Berlin, pp 280–295. https://doi.org/10.1007/978-3-030-30429-4_19
- Kokina J, Blanchette S (2019) Early evidence of digital labour in accounting: innovation with robotic process automation. *Int J Account Inf Syst* 35:100431. <https://doi.org/10.1016/j.accinf.2019.100431>

- König M, Bein L, Nikaj A, Weske M (2020) Integrating robotic process automation into business process management. In: Asatiani A, García JM, Helander N, Jiménez-Ramírez A, Koschmider A, Mendling J, Meroni G, Reijers HA (eds) Business process management: Blockchain and robotic process automation forum. Springer, Berlin, pp 132–146. https://doi.org/10.1007/978-3-030-58779-6_9
- Lacity M, Willcocks LP (2016) A New approach to automating services. *MIT Sloan Manage Rev* 58(1):1–16
- Lamberton C (2016) Get ready for robots: why planning makes the difference between success and disappointment. [https://www.ey.com/Publication/vwLUAssets/Get_ready_for_robots/\\$FILE/ey-get-ready-for-robots.pdf](https://www.ey.com/Publication/vwLUAssets/Get_ready_for_robots/$FILE/ey-get-ready-for-robots.pdf). Accessed 11 Jan 2021
- Lasso-Rodríguez G, Winkler K (2020) Hyperautomation to fulfil jobs rather than executing tasks: The BPM manager robot vs human case. *Rom J Inf Technol Autom Control-Revista Romana De Informatica Si Automatica* 30(3):7–22. <https://doi.org/10.33436/v30i3y202001>
- Leno V, Dumas M, La Rosa M, Maggi FM, Polyvyanyy A (2020) Automated discovery of data transformations for robotic process automation. <http://arxiv.org/abs/2001.01007> [Cs]
- Lodhi A, Köppen V, Wind S, Saake G, Turowski K (2014) Business process modeling language for performance evaluation. In: 2014 47th Hawaii international conference on system sciences. pp 3768–3777. <https://doi.org/10.1109/HICSS.2014.469>
- Madakam S, Holmukhe RM, Jaiswal DK (2019) The future digital work force: robotic process automation (RPA). *J Inf Syst Technol Manage* 16:1–17. <https://doi.org/10.4301/s1807-1775201916001>
- Mărușter L, van Beest NRTP (2009) Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowl Inf Syst* 21(3):267. <https://doi.org/10.1007/s10115-009-0224-0>
- Mutschler B, Reichert M, Bumiller J (2008) Unleashing the effectiveness of process-oriented information systems: problem analysis, critical success factors, and implications. *IEEE Trans Syst Man Cybern Part C (appl Rev)* 38:280–291. <https://doi.org/10.1109/TSMCC.2008.919197>
- Niehaves B, Poeppelbuss J, Plattfaut R, Becker J (2014) BPM capability development—a matter of contingencies. *Bus Process Manag J* 20(1):90–106. <https://doi.org/10.1108/BPMJ-07-2012-0068>
- Ongena G, Ravesteyn P (2019) Business process management maturity and performance: a multi group analysis of sectors and organization sizes. *Bus Process Manag J* 26:132–149. <https://doi.org/10.1108/BPMJ-08-2018-0224>
- Osmundsen K, Iden J, Bygstad B (2019) Organizing robotic process automation: balancing loose and tight coupling. <https://scholarspace.manoa.hawaii.edu/bitstream/10125/60128/0688.pdf>. Accessed 11 Jan 2021
- Pasquadibisceglie V, Appice A, Castellano G, Malerba D (2020) Predictive process mining meets computer vision. In: Fahland D, Ghidini C, Becker J, Dumas M (eds) Business process management forum. Springer, Berlin, pp 176–192. https://doi.org/10.1007/978-3-030-58638-6_11
- Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. *J Manag Inf Syst* 24(3):45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pérez-Álvarez JM, Maté A, Gómez-López MT, Trujillo J (2018) Tactical business-process-decision support based on KPIs monitoring and validation. *Comput Ind* 102:23–39. <https://doi.org/10.1016/j.compind.2018.08.001>
- Plattfaut R, Borghoff V, Godefroid M, Koch J, Trampler M, Coners A (2022) The critical success factors for robotic process automation. *Comput Ind* 138:103646. <https://doi.org/10.1016/j.compind.2022.103646>
- Poeppelbuss J, Roeglinger M (2011) What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. In: 19th European Conference on Information Systems, ECIS 2011
- Pourbafrani M, van Zelst SJ, van der Aalst WMP (2020) Semi-automated time-granularity detection for data-driven simulation using process mining and system dynamics. In: Dobbie G, Frank U, Kappel G, Liddle SW, Mayr HC (eds) Conceptual modeling. Springer, Berlin, pp 77–91. https://doi.org/10.1007/978-3-030-62522-1_6
- Radke AM, Dang MT, Tan A (2020) Using robotic process automation (RPA) to enhance item master data maintenance process. *LogForum* 16(1):129–140. <https://doi.org/10.17270/J.LOG.2020.380>
- Rafiei M, von Waldthausen L, van der Aalst WMP (2020) Supporting confidentiality in process mining using abstraction and encryption. In: Ceravolo P, van Keulen M, Gómez-López MT (eds) Data-driven process discovery and analysis. Springer, Berlin, pp 101–123. https://doi.org/10.1007/978-3-030-46633-6_6

- Reijers HA (2021) Business process management: the evolution of a discipline. *Comput Ind* 126:103404. <https://doi.org/10.1016/j.compind.2021.103404>
- Röglinger M, Pöppelbuß J, Becker J (2012) Maturity models in business process management. *Bus Process Manag J* 18(2):328–346. <https://doi.org/10.1108/14637151211225225>
- Rosemann M, De Bruin T (2005) Towards a business process management maturity model. In: Rajola F, Avison D, Winter R, Becker J, Ein-Dor P, Bartmann D, Bodendorf F, Weinhardt C, Kallinikos J. (eds.) ECIS 2005 Proceedings of the 13th European conference on information systems. Verlag and the London School of Economics, pp 1–12
- Rozinat A, Mans RS, Song M, van der Aalst WMP (2009a) Discovering simulation models. *Inf Syst* 34(3):305–327. <https://doi.org/10.1016/j.is.2008.09.002>
- Rozinat A, Wynn MT, van der Aalst WMP, ter Hofstede AHM, Fidge CJ (2009b) Workflow simulation for operational decision support. *Data Knowl Eng* 68(9):834–850. <https://doi.org/10.1016/j.datak.2009.02.014>
- Škrinjar R, Trkman P (2013) Increasing process orientation with business process management: critical practices. *Int J Inf Manage* 33:48–60. <https://doi.org/10.1016/j.ijinfomgt.2012.05.011>
- Slaby J (2012) Robotic automation emerges as a threat to traditional low-cost outsourcing. <https://www.blueprism.com/resources/white-papers/robotic-automation-emerges-as-a-threat-to-traditional-low-cost-outsourcing/>. Accessed 11 Jan 2021
- Stople A, Steinsund H, Iden J, Bygstad B (2017) Lightweight IT and the IT function: experiences from robotic process automation in a Norwegian bank. *Nokobit* 25(1):1–11
- Sutherland C (2013) Framing a constitution for robotistan: racing with the machine of robotic automation. https://neoops.com/wp-content/uploads/2014/03/RS_1310-Framing-a-constitution-for-Robotistan.pdf. Accessed 11 Jan 2021
- Syed R, Suriadi S, Adams M, Bandara W, Leemans SJJ, Ouyang C, ter Hofstede AHM, van de Weerd I, Wynn MT, Reijers HA (2020) Robotic process automation: contemporary themes and challenges. *Comput Ind* 115:103162. <https://doi.org/10.1016/j.compind.2019.103162>
- Szelągowski M (2021) Practical assessment of the nature of business processes. *Inf Syst E-Bus Manag* 19(2):541–566. <https://doi.org/10.1007/s10257-021-00501-y>
- Szelągowski M, Lupeikiene A (2020) Business process management systems: evolution and development trends. *Informatica* 31(3):579–595. <https://doi.org/10.15388/20-INFOR429>
- Szelągowski M, Berniak-Woźn J (2019) The adaptation of business process management maturity models to the context of the knowledge economy. *Bus Process Manag J* 26(1):212–238. <https://doi.org/10.1108/BPMJ-11-2018-0328>
- Szimanski F, Ralha CG, Wagner G, Ferreira DR (2013) Improving business process models with agent-based simulation and process mining. In: Nurcan S, Proper HA, Soffer P, Krogstie J, Schmidt R, Halpin T, Bider I (eds) Enterprise, business-process and information systems modeling. Springer, Berlin, pp 124–138. https://doi.org/10.1007/978-3-642-38484-4_10
- Tallon P, Queiroz M, Coltman T, Sharma R (2016) Business process and information technology alignment: construct conceptualization, empirical illustration, and directions for future research. *J Assoc Inf Syst* 17(9):563–589. <https://doi.org/10.17705/1/jais.00438>
- Teece DJ (2018) Business models and dynamic capabilities. *Long Range Plan* 51(1):40–49. <https://doi.org/10.1016/j.lrp.2017.06.007>
- Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. *Strateg Manag J* 18(7):509–533
- van der Aalst WMP (2013a) Business process management: a comprehensive survey. *ISRN Softw Eng* 2013:e507984. <https://doi.org/10.1155/2013/507984>
- van der Aalst WMP (2013b) Service mining: using process mining to discover, check, and improve service behavior. *IEEE Trans Serv Comput* 6(4):525–535. <https://doi.org/10.1109/TSC.2012.25>
- van der Aalst WMP (2016) Process mining: data science in action. Springer, Berlin
- van der Aalst WMP (2021) Hybrid Intelligence: to automate or not to automate, that is the question. *Int J Inf Syst Proj Manag* 9(2):5–20
- van der Aalst, WM, Netjes M, Reijers HA (2007) Supporting the full BPM life-cycle using process mining and intelligent redesign. In: Siau K (ed) Contemporary issues in database design and information systems development. IGI Global, pp 100–132. <https://doi.org/10.4018/978-1-59904-289-3.ch004>
- van der Aalst WMP, Bichler M, Heinzl A (2018) Robotic process automation. *Bus Inf Syst Eng* 60(4):269–272. <https://doi.org/10.1007/s12599-018-0542-4>

- van Dongen BF (2012) BPI challenge: event log of a process loan application process. <https://data.4tu.nl/repository/uuid:3926db30-f712-4394-aebc-75976070e91f>. Accessed 11 Jan 2020
- Verbeek HMW, van der Aalst WMP (2015) Decomposed process mining: the ILP case. In: Fournier F, Mendling J (eds) Business process management workshops. Springer, Berlin, pp 264–276. https://doi.org/10.1007/978-3-319-15895-2_23
- Verbeek HMW, van der Aalst WMP, Munoz-Gama J (2017) Divide and conquer: a tool framework for supporting decomposed discovery in process mining. *Comput J* 60(11):1649–1674. <https://doi.org/10.1093/comjnl/bxx040>
- vom Brocke J, Baier M-S, Schmiedel T, Stelzl K, Röglinger M, Wehking C (2021) Context-aware business process management. *Bus Inf Syst Eng* 63:533–550. <https://doi.org/10.1007/s12599-021-00685-0>
- Wastell DG, White P, Kawalek P (1994) A methodology for business process redesign: experiences and issues. *J Strateg Inf Syst* 3:23–40. [https://doi.org/10.1016/0963-8687\(94\)90004-3](https://doi.org/10.1016/0963-8687(94)90004-3)
- Weske M (2019) Business process management. Springer, Berlin
- Willcocks L, Lacity M, Craig A (2015a) The IT function and robotic process automation. <http://eprints.lse.ac.uk/64519/>. Accessed 11 Jan 2021
- Willcocks L, Lacity M, Craig A (2015b) Robotic process automation at Xchanging. http://eprints.lse.ac.uk/64518/1/OUWRPS_15_03_published.pdf. Accessed 11 Jan 2021
- Yasmin FA, Bukhsh FA, Silva PDA (2018) Process enhancement in process mining: a literature review. In: CEUR workshop proceedings, vol 2270, pp 65–72
- Zelt S, Schmiedel T, vom Brocke J (2018) Understanding the nature of processes: an information-processing perspective. *Bus Process Manag J* 24(1):67–88. <https://doi.org/10.1108/BPMJ-05-2016-0102>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.