Process Selection in RPA Projects – Towards a Quantifiable Method of Decision Making

Completed Research Paper

Jonas Wanner
University of Würzburg
Würzburg, Germany
jonas.wanner@uni-wuerzburg.de

Adrian Hofmann
University of Würzburg
Würzburg, Germany
adrian.hofmann@uni-wuerzburg.de

Marcus Fischer
University of Würzburg
Würzburg, Germany
marcus.fischer@uni-wuerzburg.de

Florian Imgrund
University of Würzburg
Würzburg, Germany
florian.imgrund@uni-wuerzburg.de

Christian Janiesch
Technische Universität Dresden
Dresden, Germany
christian.janiesch@tu-dresden.de

Jerome Geyer-Klingeberg
University of Augsburg
Augsburg, Germany
jerome.geyer-klingeberg@mrm.uni-augsburg.de

Abstract

The digital age requires companies to invest in value-creating rather than routine activities to drive innovation as a future source of competitiveness and business success. Thus, many companies are reluctant to invest in large-scale, costly backend integration projects and seek adaptable solutions to automate their front-office activities. Bridging artificial intelligence and business process management, robotic process automation (RPA) provides the promise of robots as a virtual workforce that performs these tasks in a self-determined manner. Many studies have highlighted potential benefits of RPA. However, little data is available on operationalizing and automating RPA to maximize its benefits. In this paper, we shed light on the automation potential of processes with RPA and operationalize it. Based on process mining techniques, we propose an automatable indicator system as well as present and evaluate decision support for companies that seek to better prioritize their RPA activities and to maximize their return on investment.

Keywords: Robotic Process Automation, Quantifiable Method, Organization of Work, Decision Support, Process Mining

Introduction

The digital age is changing the nature of business sustainably (Matt et al. 2015). Technological advances have paved the way for innovations that are tailored to increasingly sophisticated customer demands (Agarwal et al. 2010; Prahalad and Ramaswamy 2004; Zysman et al. 2011). Many of these new business opportunities arise from data, which is collected, stored, and analyzed for insights and patterns that facilitate more unique customer experiences (Davenport et al. 2012).

As a key driver of future competitiveness, companies must establish the strategic capabilities to retrieve data from different sources of which they derive valuable information to strengthen, expand, or adjust their actions (Imgrund et al. 2018; Sebastian et al. 2017). This demands a new organization of work, which promotes agility and as well as collaboration and puts a stronger focus on cognitive, knowledge-intensive, and value-creating activities. However, in many organizations, significant resources are tied up in work routines and repetitive tasks (van der Aalst et al. 2018). Despite a widespread use of process-aware information systems (PAIS), companies still rely on employees to trigger or terminate processes, adapt them to cope with different cases, or use their outcomes for a variety of purposes (van der Aalst et al. 2018).

Using the example of e-commerce, many retailers use multiple distribution channels (e.g., Amazon Marketplace and self-run web shops) to address different customer segments. These platforms compile and forward incoming orders to employees that are responsible for entering the data into internal management systems for further processing. Besides the significant continuous resource consumption, this procedure often involves media breaks, is prone to human error, and hampers business performance. Nevertheless, the project cost and duration of a comprehensive integration project is prohibitive. Another economically viable option must be found that allows improving (parts of) existing structures. One potential solution to address this is robotic process automation (RPA), an emerging concept in the field of business process automation technology.

RPA essentially builds on software robots that replicate human tasks without requiring adaptations to the underlying information system (Geyer-Klingeberg et al. 2018; van der Aalst et al. 2018) invalidating the need for said integration efforts. Enabling companies to free up resources and to reallocate them to activities with a focus on creating business value and customer satisfaction, RPA can foster the emergence of new work forms and drive organizational competitiveness in the digital age. Although research and practice show a growing interest in RPA, its full potential remains poorly understood. In fact, most research focuses on case studies or market research (Aguirre and Rodriguez 2017; Geyer-Klingeberg et al. 2018; Lacity et al. 2015a; Mendling et al. 2018; Schmitz et al. 2019; Dunie and Tornbohm 2017). This conforms to van der Aalst et al. (2018), who suggest that urgent questions pertaining to RPA have not been answered yet.

One major challenge of RPA projects is to select the most suitable processes for automation (van der Aalst et al. 2018). Complex organizations operate on differentiated architectures with processes that differ concerning various characteristics such as their execution frequency and length, their number of involved departments and stakeholders, and their variability in inputs and outcomes. As these characteristics influence a process's automation potential, RPA is not a one-size-fits-all solution but rather demands careful analyses and informed decision making (Geyer-Klingeberg et al. 2018). Hence, companies must obtain a detailed understanding of their processes to ensure a sufficient economic value and the overall success of their RPA initiatives.

To date, research has proposed a few high-level concepts to measure the suitability of processes for RPA (Aguirre and Rodriguez 2017; Geyer-Klingeberg et al. 2018). Typically, these approaches neglect objectivity and generalizability. We seek to close this gap by developing and operationalizing a multi-dimensional indicator system, which sets the groundwork for a decision-support method to guide companies in planning and designing RPA initiatives in practice. Our approach focuses on maximizing the (short-term) economic success of RPA projects. If RPA projects are selected and managed appropriately over time based on our framework, the (short-term) value realization can translate into long-term strategic value of RPA, which is freeing up resources and shifting focus to value-creating activities. The transparency created also counteracts the agency problem between short-term benefit promises of consulting companies and long-term strategic opportunity of their clients. We employ design science research (DSR) to answer the following research question (RQ):

(RQ) How can we guide companies in selecting processes that are beneficial for their RPA projects?

Answering this question comes with several contributions. First, we provide a first and software- as well as industry-independent conceptualization of indicators relevant for immediate success of RPA projects. Second, we present a method for process selection for RPA that provides decision support applicable in practice. Third, we validate the method by incorporating expert knowledge derived from an extensive interview study. Fourth, we open up several avenues for future research.

The remainder of this paper is organized as follows: in Section 2, we introduce fundamental knowledge on RPA and its related topics. We describe our research methodology in Section 3. In Section 4, we convey the current state-of-the-art in the field of RPA and integrate findings from several expert interviews. We introduce our decision framework in Section 5, before discussing and evaluating our approach in Section 6. Ultimately, in Section 7 we summarize findings, limitations, and future research opportunities.

Fundamental Knowledge

Business Process Management

Business process management (BPM) comprises methods, techniques, and technologies to constantly adapt and improve business processes by identifying, prioritizing, and analyzing, and monitoring them in a recursive manner. Thus, it helps companies to cope with competitive market conditions and is becoming increasingly important to cope with digital transformation (Dumas et al. 2018; Imgrund et al. 2018). Typically based on life cycle models and maturity models, companies use BPM to achieve organizational flexibility, agility, and responsiveness (Dumas et al. 2018). Along with the widespread adoption of BPM, several studies report of initiatives that do not deliver expected returns or that result in project failure. According to vom Brocke et al. (2016), reasons for this lie in the narrow focus of most BPM approaches, which are designed to fit a specific business context with structured processes as well as clear goals and requirements. Processes managed in these initiatives typically represent only the tip of the iceberg, as a large number of unstructured and widely implicit processes remain beneath the surface. In line with that, Imgrund et al. (2017) state that most companies only improve a handful of their processes at a time and, thus, neglect large parts of their process architecture. They conceptualize this observation in the theory of the long tail of business processes. Based on their importance, dysfunctionality, and feasibility, the traditional Pareto distribution model implied that approximately 20 % of a company's processes (short head) provide roughly 80 % of the available improvement potential. Due to their disadvantageous ratio of benefits to costs, lower-value processes are often not identified as suitable improvement candidates and neglected. A long-tailed distribution, however, suggests that the remainder of 80 % of the processes long tail constitutes more than 50 % of the available improvement potential rather than 20 %. As a consequence, companies must apply innovative approaches to access the, in sum, considerable improvement potential in the long tail of their process distribution (van der Aalst et al. 2018). One approach can be automation, which frees up resources in less valuable parts of an organization and reduces the need for continuous management and control.

Robotic Process Automation

RPA builds upon a set of tools that operate virtual robots on the user interface of PAIS in a human manner (van der Aalst et al. 2018). It is part of the broader discipline of business automation, which comprises related concepts such as business process automation (Scheer et al. 2004) or workflow automation (Georgakopoulos et al. 1995). While all approaches share similar goals and outcomes, RPA widens the scope of traditional automation solutions, as it is neither limited to a single software system nor require the use of APIs for data exchange and data processing. Instead, RPA works on the user layer, is compatible with a wide range of different software systems, and can thus automate processes that run through multiple systems and environments. RPA is suitable especially for 'swivel chair work', which entails that employees take inputs from one type of system, process it in a rule-based manner, and enter corresponding outputs into other types of systems (Willcocks et al. 2015a). More specifically, the robot can log into applications, copy and paste data, extract and process structured or semi-structured content from documents, scrape data from the Internet, or perform calculations. In contrast to traditional automation solutions, RPA does not require changes to applications but can be adapted to a wide range of application interfaces. Hence, it can be used to automate back office activities in different industries and application contexts without the need of application programming interfaces (API). From a technical perspective, RPA executes tasks based on simulating keyboard and mouse controls. Instead of accessing a system's backend, RPA integrates with its frontend to mimic human tasks by reacting to events on a computer screen (Asatiani and Penttinen 2016). This entails that RPA projects focus on frontend automation not backend automation. Software robots can either be programmed, configured using a graphical user interface, or setup based on macro-like recordings. More sophisticated approaches include self-learning systems or machine learning that access and analyze recorded employee activities to learn automatable tasks.

Several studies report that RPA offers benefits in the dimensions of quality, compliance, efficiency, and effectiveness (Aguirre and Rodriguez 2017; Lacity et al. 2015a; Schmitz et al. 2019). Other reports further project that the widespread use of RPA yields significant increases in productivity and a large-scale automation of back-office operations (Willcocks et al. 2015a) associated with a faster time to value. Further, economic research suggests that firms adopting (physical) robots as work force are more successful than non-adopters (Koch et al. 2019). While the research has not been transferred to virtual robots yet, the analogy is compelling.

Finally, it is important to note that robots typically do not automate complete end-to-end processes but only sub-processes or certain tasks thereof. Consequently, we distinguish processes, referring to end-to-end or sub-processes, and tasks, referring to individual activities within a process. An RPA project might yield the automation of one process, or one or multiple tasks respectively. While our unit of analysis is a process, the results of our analyses do not necessitate automating processes rather than only parts thereof.

Process Mining

Process mining focuses on collecting, analyzing, and interpreting execution data extracted from event logs of PAIS (van der Aalst et al. 2012). It assumes that events are recorded sequentially and refer to certain activities that together form cases (van der Aalst et al. 2004). Transaction logs provide various types of information including activity names, timestamps, and user identifications, which can be operationalized to shed light on a company's work routines, performance, and social structure (van der Aalst et al. 2007). Process mining provides companies with the means to achieve various objectives such as obtaining a better understanding and control of their processes or finding the root causes of performance deficiencies (Li et al. 2008). van der Aalst et al. (2012) distinguish three forms of process mining: *discovery*, *conformance checking*, and *enhancement*. Discovery techniques are used to derive the control flow of a process. They seek to produce process models without any a-priori information and often determine the starting point of process mining initiatives. With conformance checking, companies can compare the to-be design with the as-is design of a process to detect yet unknown structural deviations. Ultimately, enhancement builds upon process execution data to extend existing process models with additional information. This includes the use of timestamps to highlight for example bottlenecks, service levels, throughput times, or execution frequencies. Geyer-Klingeberg et al. (2018) describe process mining and RPA as complementary concepts. More specifically, they argue that process mining supports companies in capturing an end-to-end view on their processes, identifying high-value automation candidates, quantifying the economic value of corresponding initiatives, and eliminating costly and often subjective manual process evaluations.

Research Method

We employ DSR to develop our contribution (Hevner et al. 2004). DSR consists of multiple iterative steps, including *problem identification*, the *definition of objectives*, *design and development*, *demonstration*, *evaluation,* and *communication* (Peffers et al. 2007).

It is the goal of this research to design a quantifiable method of decision making for process selection in RPA initiatives. This conforms to van der Aalst et al. (2018)'s call for research, who emphasize the importance of understanding which characteristics make a process suitable for RPA. Based on our RQ, we derive more detailed objectives by conducting a literature analysis for requirements engineering. While the systematic literature review follows the recommendations proposed by Webster and Watson (2002), we base the process of empirical theory building on the principles of case study research (Yin 1989). This entails examining the current state-of-the-art and combining its implications with insights from practice to identify the most significant challenges and requirements of successful RPA projects.

For design and development, we use these insights to construct our method. Thereby, we analyze common event log configuration and define a set of indicators that can be used to assess the automation potential of process tasks. In addition, we specify procedures to analyze the benefits and costs of process task automation and to maximize the returns of RPA at project level. During the demonstration, we formalize our indicator system, specify the decision situation, and describe each step of our method. To do justice to the iterative character of DSR, we develop a first version of our method based on case studies before assessing the preliminary result model based on expert interviews. This enables us to include a practical real-world perspective that is less prone to factual theorization. We repeated the design and development phase until the benefit of including further insights by the means of literature or interviews became neglectable. In total, this resulted in three iterations of the design and development phase (cf. Section 'Best Practices'). Ultimately, we evaluate the usefulness and quality of our method by applying it to several process logs collected from companies and institutions of different industries.

Preparatory Work for Quantifying RPA Process Selection

Requirements Analysis

We build a catalog of general requirements for RPA from literature. As exposed by van der Aalst et al. (2018), we seek to provide decision support for the prioritization and selection of processes or process tasks that are most suitable to be automated in RPA projects, thereby answering a research question that has not adequately been addressed yet. In a first step, we rely on findings from academic literature to conceptualize three essential challenges of RPA. Summarized in Table 1, we link them to their causes and implications, before we give actionable recommendations on how to address them.

Table 1: Problem Definitions and Implications from Current RPA Projects

Challenge	Cause	Implication	Approach	Source
Ad hoc and demand-driven implementation of RPA	Complexity of RPA projects and lack of methods to systematically identify suitable processes	RPA requires a structured approach to process selection to avoid rework and project failure	A comprehensive indicator system to support companies in selecting processes for RPA	(Lindström et al. 2018; Marrella 2018)
RPA fails to deliver expected cost savings	Inaccurate predictions on the economic value of RPA at process level	RPA demands a system to accurately pin down the economic value of automating a process	A decision-support system for process selection as a means for detailed cost-benefits-analysis	(Willcocks et al. 2015a)
Processes vary in their automation potential over time	Changing requirements may alter the feasibility of processes to be automated with RPA	RPA requires methods that detect and account for changes in environmental requirements	Recurring and periodic evaluations of the automation potential of processes	(Lacity and Willcocks 2016; Schmitz et al. 2019)

The summary of implications suggests that RPA projects demand a comprehensive indicator system for process task selection, a way to evaluate the profitability and advantageousness of process task automation, and an agile approach that accounts for dynamics in process task requirements.

Best Practices

To approach the challenges related to process task selection in RPA projects, we built our research on case studies (Iteration I) and streamlined its results by interviewing practitioners that we regard as experts in the researched domain based on their profession and experience (Iteration II). We further evaluated the developed indicator system with these experts to incorporate additional and yet neglected implications (Iteration III).

Case Studies

We examined six case studies with a focus on challenges during RPA projects (Aguirre and Rodriguez 2017; Asatiani and Penttinen 2016; Lacity et al. 2015a; Lacity et al. 2015b; Schmitz et al. 2019; Willcocks et al. 2015b). Three case companies are highly automated businesses that provide outsourcing services in the area of insurance (*Xchanging*), finance (*OpusCapita*), and CRM-systems (anonymous). Furthermore, we include insights of *Deutsche Telekom* and *Telefonica O2* from the sector of telecommunication, as well as one of Europe's biggest utilizing company (anonymous - *bAN*). Despite several preparatory steps to initiate RPA projects, such as defining objectives, deciding for a software provider, or establishing a Center of Excellence (Lacity et al. 2015a; Schmitz et al. 2019), all companies started with collecting processes before entering into the phase of process task selection for RPA automation itself.

First, we encountered a lack of indicators ensuring measurability and comparability. We found that none of the case companies used standardized tools to select the most appropriate process tasks for RPA. Instead, they attempted to quantify the automation potential of their process tasks with a wide set of highly individualized measures. In particular, the cases of *bAN* and *Xchanging* show that a lack of adequate measures can yield low decision quality, eventually jeopardizing the success of the entire venture (Lacity et al. 2015b; Willcocks et al. 2015b). Better results were achieved by companies that used measurable indicators such as complexity and execution frequency (Asatiani and Penttinen 2016; Lacity et al. 2015a; Schmitz et al. 2019). All companies provide a set of lessons learned from their first experiences with RPA. By consolidating these insights, we identified four measures for quantifying the automation potential of process tasks in RPA projects: (1) *execution frequency*, as volume of transactions; (2) *degree of stability*, as little or almost no exceptions occur while completing a task, (3) *degree of rule-based procedure*, as process tasks and decisions can be easily disassembled into steps without misinterpretation; and (4) *degree of standardization*, as mature process tasks with extensive knowledge about known deviations and outcomes.

Second, RPA typically fails to deliver the expected cost savings if the company fails to conduct a detailed cost-benefits analysis. All companies shared the objective of reducing full-time equivalents (FTE) with RPA to save costs or to free up resources for more complex and creative tasks. However, prior to the project, most companies assessed potential savings from RPA in a widely subjective manner based on opinions formed by individual experiences. Only after project completion, they assessed the true savings by measuring the reduction in throughput time in automated process tasks (e.g. Lacity et al. 2015a). Two of the six case companies conducted a more precise assessment procedure. At *Deutsche Telekom*, the RPA team estimated potential savings and costs of each process candidate and conducted two-week workshops with stakeholders from IT, business, and consultancy to strengthen their understanding of each task's characteristics (Schmitz et al. 2019). Similarly, *bAN* initially identified processes or tasks whose automation offered potential cost reductions of up to 200 % compared to the cost induced by their manual performance within a 12 months period. *bAN* included both fixed and variable cost necessary to instantiate an RPA project (e.g., costs for software licenses, hardware, IT service costs, and RPA staff). The company measured the benefits from RPA based on several metrics such as FTE avoidance, FTE redeployment or FTE savings. Furthermore, it incorporated manual costs by considering the throughput time of process tasks or the costs for salaries and organizational overhead (Lacity et al. 2015b).

For the third problem, we found that the automation potential of processes varies over time. Due to changes., especially Lacity et al. (2015b) and Willcocks et al. (2015b) highlight that companies must continually improve their RPA initiatives projects efforts by performing the corresponding steps iteratively. In conclusion, this suggests that the success of RPA projects depends significantly on the availability of a standardized method to quantify and compare the automation potential of process tasks. This does not only support companies in decision making during process selection, but also helps them in identifying the processes that yield the highest returns on investment.

Expert Survey

Based on the case study analysis carried out, we have conducted a two-part expert survey on success factors for RPA projects in practice, in two phases (15; 8). Thereby, we found tactics and tools to address the challenges faced by companies during process selection (cf. Section 'Requirement Analysis') and to further improved our research artifact. As such, we encompassed used open-ended questions in a semi-structured interview guide questionnaire (Myers and Newman 2007) and conducted the interviews via face-to-face, via phone, or via web-conference. Therefore, we have been able to discuss different viewpoints of relevant subtopics. The survey comprises fifteen participants, all with several years of experience in the digitalization of business processes, who have been involved in at least three RPA projects. To avoid biases, we have mainly focused on RPA service providers as they deal with a wide range of customers. The participants themselves were from the big major RPA software vendors *AutomationAnywhere, UiPath,* and *Blue Prism*, *Celonis* as a leader in process mining, *Ernst & Young*, *Brightcape,* and *Capgemini* as consulting firms, and four companies from different industrial sectors that have already implemented RPA solutions in their day-to-day business operations. Foremost, there was high interest in the topic of quantifying process selection as this posed a key problem related to a high rate of unsuccessful RPA projects. None of the experts had an objective solution to quantify and compare processes and tasks but sought to have one.

RPA software vendors with their own tools provide the most advanced indicator systems. These are based on a step-by-step subjective interviewing procedure of the client to identify the complexity of each process and to use a simple classification algorithm based on existing experiences to determine whether or not it is suitable to be automated using RPA. Other approaches employ similar methods as the ones reported in the case studies. They try to define the complexity of a process by substitution of measurable indicators. The result is than compared with the number of process executions as the degree of repetition, and thus of potential savings. Nevertheless, all approaches have in common that they are mainly based on experience and subjective estimation rather than on a quantifiable approach. The most common quantifiable measures mentioned were *execution frequency* (15), *duration of process tasks* (10), *degree of standardization* (9), *degree of stability* (8), *failure rate* (7), *automation ratio* (7), *number of intersections* (5), *structure of data* (5), and *application stability* (4). This confirms and extends the indicators found when analyzing the case studies. An exception was the *degree of rule-based procedures*.

Currently, the problem of the missing cost-benefit analysis is addressed for the most part by estimation of the FTE that will be saved if the process (or task) is automated. To calculate the FTE equivalent, the frequency of each process task is multiplied with its execution time. If there is no log data available, the value is estimated, based on statements of internal employees directly involved in daily operations. Often, the comparison is simplified using a fix cost rate for the RPA robot's implementation. To 'keep things simple', the savings are measured by the time reduction. It is determined ex post (cf. Section 'Case Studies'). Afterwards, the processes are prioritized for the automation by RPA robots. The costs of each process task's automation are compared to the companies' defined RPA project aims such as an improvement of process quality, cost savings, or a higher customer satisfaction. Half of the experts considered idea of the return of investment as the core of the successive processing. The other half of the experts were in favor of a budget limit, as it is common in any project. Further, all experts agreed on the need for an economic efficiency and, thus, a way to pre-indicate the potential savings in a reliable manner as there will be a time when the 'RPA hype' is over but the need is still there.

Most expert confirmed a variation of the automation potential of processes over time due to internal and external changes and indicated that it should be considered for the final solution. In sum, the recommendations suggest a standardized step-by-step method to quantify the potential of process tasks and their cost-benefit-ratio in relation to budget restrictions. This is in line with the conclusions from the case studies in the previous section.

Method Development for Quantifying RPA Process Selection

Method Construction

Based on the above, we have found that RPA projects demand a comprehensive indicator system for process selection, a way to evaluate the profitability and advantageousness of process automation, and a standardized method that accounts for dynamics in process requirements. As illustrated in Figure 1, our method to address this consists of three steps, taking the input from the phase of *Process Pre-Selection* and transferring its result to the phase of *Final Process Decision*.

Depending on the scope of an RPA project, our method requires a phase of *Process Pre-Selection* to identify processes of interest first. This ensures that special challenges, for example legal compliance, or defined aims such as improving customer satisfaction, are addressed before comparing them in a quantifiable manner. The experts mentioned further measures, which we did not include due to complexity and restrictions of log data availability (e.g. number of intersections, application stability, or data structure).

In the first step of our method, the automation potential of every process task is calculated on the basis of its properties based on measurable indicators. In the second step, the method enables companies to analyze the profitability of process task automation. Therefore, companies can compare the benefits of RPA, such as cost reduction or quality increase, with fixed and variable costs imposed by configuring, implementing, and maintaining corresponding robots. Having identified the process tasks that are both automatable and profitable, companies must account for their project budget (e.g. 1,000,000 €) or amortization restrictions (e.g., 12 months) to determine the scope of an RPA project and to maximize its economic return in the third step. The result is a list of recommended process tasks taking into account the above limitations.

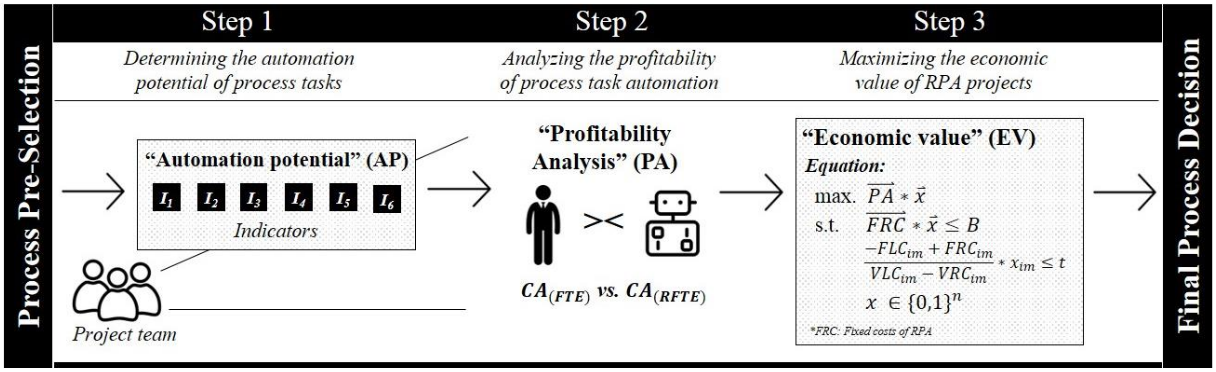

Figure 1: Methodology for Process Selection in RPA Projects

This improves the decision-making in the subsequent phase of *Final RPA Decision* and consequently, increases the likelihood of the RPA project's success.

Determining the Automation Potential of Processes

To acquire a detailed understanding of their process organization, companies must define and operationalize indicators that facilitate the separation of promising processes and process tasks from those with a disadvantageous ratio of benefits to costs. Despite the growing importance of RPA in research and practice (van der Aalst et al. 2018), only little is known about process characteristics that can be used for process task selection. In fact, most studies treat process and process task selection as a case-by-case challenge that requires situational analysis and adaptation (Asatiani and Penttinen 2016; Lacity et al. 2015a; Schmitz et al. 2019). Many established approaches define high-level indicators but build upon the collection of qualitative data. Because corresponding implications are prone to measurement and assessment biases, they lack comparability and reproducibility and hardly can be adopted by companies encountering similar challenges in a different context.

To overcome these limitations, we suggest to analyzing process execution data collected from PAIS. In event logs, PAIS store a wide range of process features, including *execution* and *waiting times*, *handovers of work*, or *executing systems,* or *users*. Further data such as *execution frequencies* and *variances* can be derived. We analyze the current body of literature and combine implications from research and practice to build a system of indicators and to analyze which of the features in event log data can be used for process task selection in RPA. To ensure generalizability, we abstract from case-specific observations and translate them into high-level requirements that can be adapted to different application scenarios. Table 2 presents six process characteristics relevant to process selection in RPA projects.

We use the features of process execution data to define indicators that enable companies to assess the characteristics specified in Table 2 in a reliable and reproducible manner. Therefore, we introduce a formal notation, whereby an event log E consists of a finite set of processes P_i. Processes are further composed of cases C_{ij}, which represent process executions. Ultimately, cases describe a stream of activities A_{ijk} that stand for the tasks of a process. Each activity is associated with an activity type $\bar{A}_m$. We use indices to refer to processes *(i)*, cases *(j)*, or tasks *(k)*.

Most cases and all experts emphasize the importance of a task's execution frequency for process selection in RPA (Asatiani and Penttinen 2016; Lacity et al. 2015a). Frequently performed tasks offer economies of scale and, thus, allow companies to realize significant cost reductions and to leverage the returns of automation. Equation 1 operationalizes the feature of *execution frequency* as an indicator for the automation potential of processes. We measure EF_{im} as the count of each activity m belonging to the same process i.

Table 2: Indicators for the Automation Potential of Processes		
Process characteristics	**Low-level properties**	**Example references**
Execution Frequency (EF)	Repetitive process tasks with a high volume of transactions, sub-tasks, and frequent interactions between different systems or interfaces	(Asatiani and Penttinen 2016; Fung 2014)
Execution Time (ET)	Average execution time of a process task	Cf. Section 'Best Practices'
Standardization (SD)	Streamlined process tasks with a-priori knowledge of possible events and outcomes of process task executions	(Asatiani and Penttinen 2016; Willcocks et al. 2017)
Stability (ST)	Process tasks with a low probability of exception and a high predictability of outcomes	(Anagnoste 2018; Lindström et al. 2018)
Failure Rate (FR)	Throwbacks ratio of process tasks, i.e. unusual and repetitive (partial) tasks until completion	Cf. Section 'Best Practices'
Automation Rate (AR)	Process tasks with a small number of steps that are already automated and offer less significant economic benefits	(Geyer-Klingeberg et al. 2018; Cf. Section 'Best Practices')

$$EF_{im} = \sum_{j=1}^{|P_i|} |\{A_{ijk} \in C_{ij} | \mathrm{A_{ijk}} \approx \bar{A}_m\}| \tag{1}$$

Similarly, not only activities that are executed frequently provide a high saving potential, but also activities that are time consuming. Hence, we consider the average duration of an activity during a process execution, formalized in equation 2.

$$ET_{im} = \frac{\sum_{j=1}^{|P_i|} \sum_{k=1}^{|C_{ij}|} T(A_{ijk}, \bar{A}_\mathrm{m})}{EF_{im}} \tag{2}$$

$$\text{with } T\left(A_{ijk}, \overline{A}_m\right) = \begin{cases} execution\ time\ of\ activity\ A_{ijk}\ if\ A_{ijk} \approx \overline{A}_m \\ 0\ else \end{cases}$$

Our results further suggest that the automation potential of an activity depends on their degree of standardization (Lacity et al. 2015a). In general, standardization describes the act of establishing a best practice of how to execute a process and yields several benefits, including higher productivity and output, less complex process improvement, and easier onboarding (Asatiani and Penttinen 2016; Fung 2014; Willcocks et al. 2015a). Standardization builds upon on a profound understanding of an activity's environment and its goals and requirements. It facilitates the reduction of process variants and structural deficiencies (Geyer-Klingeberg et al. 2018; Moffitt et al. 2018). Hence, it enables companies to reduce the cost of RPA, as robots must cover fewer variants to support a business case completely. As is more natural to measure a lack of standardization in the form of different outcomes, we measure an inverse standardization. Equation 3 presents a formalism to measure the inverse of an activity's degree of *standardization*. SD_{im}^{-1} analyzes each activity in a given process instance and returns the number of different prior and following activities. This gives an indication about the number of different contexts an activity is executed in.

$$SD_{im}^{-1} = |\bigcup_{j=1}^{|P_i|} \bigcup_{k=1}^{|C_{ij}|} \{ (\bar{A}_m^{-1}, \bar{A}_m^{+1}) | A_{ijk-1}, A_{ijk}, A_{ijk+1} \in C_{ij}, A_{ijk-1} \approx \bar{A}_m^{-1}, A_{ijk} \approx \bar{A}_m, A_{ijk+1} \approx \bar{A}_m^{+1} \}| \tag{3}$$

We further found that the stability of an activity is an important determinant for process selection in RPA. Here, stability refers to a condition in which all causes of variations in an activity's performance is known, so that its outcomes are robust and predictable within certain boundaries. An activity's stability correlates with its automation potential as it prevents companies from encountering unforeseen events that hamper and delay the success of an RPA project. As with standardization, we measure the inverse stability of an activity based on the variance of its execution times. We present the underlying formalism in Equation 4. Thus, we initially compute the average duration of a process. ST_{im}^{-1} then returns the sum of the squared differences of the execution times of an activity in a process instance and the average duration of that activity and normalizes the results based on the number of activity executions. We noticed a strong correlation between execution time and execution time variance. Therefore, we normalized the execution time variance by the execution time to obtain a variance factor.

$$ST_{im}^{-1} = \frac{\sum_{j=1}^{|P_i|} \sum_{k=1}^{|C_{ij}|} \left(T(A_{ijk}, \bar{A}_m) - ET_{im}\right)^2}{EF_{im} * ET_{im}} \tag{4}$$

A common assumption that the experts made was that automated process steps are less prone to errors than activities done by humans. Therefore, they suggested that activities with a high failure rate are more suitable for RPA since the reduced errors can lead to overall savings. Activities that were not performed correctly must be repeated to correct this error. Therefore, we estimate the error rate by counting redundant activities during the process execution per case. This is equivalent to the execution frequency of the activity per case in which the activity is executed. Activities repeated often should be considered for RPA. We formalize this in equation 5.

$$FR_{im} = \frac{EF_{im}}{|\{C_{ij} \in P_i \mid there\ exists\ A_{ijk} \in C_{ij}\ such\ that\ A_{ijk} \approx \bar{A}_m\}|} \tag{5}$$

Ultimately, we found that processes suitable for RPA are characterized by a low automation rate. For example, Geyer-Klingeberg et al. (2018) state that the automation of widely manual processes is promising especially in the early phases of RPA projects, as they offer opportunities to simultaneously realize quick wins and a high impact in various dimensions, including costs, effectiveness, and productivity. Naturally we do not consider activities that are already fully automated as stated in equation 6.

$$AR_{im} = \begin{cases} 1 & if\ \bar{A}_m\ is\ automated \\ 0 & else \end{cases} \tag{6}$$

Analyzing the Profitability of Process Automation

Having specified a set of indicators to assess the automation potential of processes, we present an integrated profitability analysis PA_i in the subsequent section. In general, we reduce decision making to the two components of benefits and costs. Thereby, benefits represent cost reductions as the result of replacing human labor by robots. In contrast, costs result from the investment necessary to set up, configure, and maintain RPA robots.

To create a basis for decision support to analyze the profitability of each process activity for its economical suitability for automation, it is necessary to calculate the underlying costing rates, which are passed on as input variables for the method developed. Both direct and indirect costs must be considered. As individual processes activities are examined for their overall cost to compare on, it is of decisive importance to carry out a cause-related cost analysis. This follows results from the individual cost rates, such as those between companies and external providers, between countries, and tariffs, as well as or between industries and internal wages workplaces.

Therefore, we recommend companies to use the well-established method of activity-based costing to calculate these individualities (Hansen et al. 2007). To determine the total cost function to analyze the profitability of RPA process automation, the four cost rates summarized in Table 3 must be presumed.

For human labor, we assume that executing a process yields both fixed and variable costs. Fixed costs are execution-independent costs. They are often called 'overhead cost' and categorized with indirect costs, such as rent or equipment (Hansen et al. 2007). We represented this by the constant FLC_i (equation 7) as the sum of the overhead cost represented by the variable *fix labor cost rate* $FLCR$.

$$FLC_{im} = FLCR \tag{7}$$

By contrast, variable costs of human labor are direct costs of a process. They correlate with its execution frequency and the execution time. We summarize the variable cost of human labor with the function

$VLC_{im}(EF_{im}^{+}, ET_{im}^{+}, AR_{im}^{-})$, whereby the total cost is determined by the product of the execution frequency EF_{im} and the average duration $VLCR$. We only consider already non-automated activities in the economical selection of the best cost saving (equation 8).

$$VLC_{im} = VLCR * EF_{im} * ET_{im} * (1 - AR_{im}) \tag{8}$$

In addition to the costs of human labor, we further account for those imposed by implementing a fully automated RPA solution. Analogously, we assume that costs for RPA have a fixed and a variable cost component. Fixed cost includes expenses for configuring and maintaining robots and depend on a process's stability and degree of standardization. Less standardized processes are characterized by a larger number of process variants, which must be implemented using RPA software. This also applies to unstable processes that impose the risk of frequent adaptations subsequent to the deployment of RPA. We capture fixed costs of RPA with the function $FRC_{im}(SD_{im}^{-}, ST_{im}^{-}, AR_{im}^{-})$, which captures the inverse relationship between variable costs and a process's current automation rate, as well as the correlation between costs, process variance, and process stability (equation 9). $FRCR$ can be considered the equivalent of writing a certain amount of lines of code.

$$FRC_{im} = FRCR * SD_{im}^{-1} * ST_{im}^{-1} * (1 - AR_{im}) \tag{9}$$

We further assume that RPA comes with variable costs that result from executing a process with RPA software. Cost rates typically depend on agreements between the implementing company and RPA vendors and can vary significantly (Grand View Research 2018). We account for variable costs of RPA with the function $VRC_{im}(EF_{im}{}^{+}, FR_{im}^{-})$, whereby a larger number of process executions yields increases in the variable costs of RPA (equation 10).

$$VRC_{im} = VRCR * \frac{EF_{im}}{FR_{im}} \tag{10}$$

Finally, the individual cost variables must be brought into the overall cost analysis context. By operationalizing the equation system, we can define the following linear cost functions for human labor HL and robotic labor RL to enable economic comparison (equation 11 and 12):

$$CA_{HL} = FLC_{im} + VLC_{im} \tag{11}$$

$$CA_{RL} = FRC_{im} + VRC_{im} \tag{12}$$

Table 3: Indicators for the Profitability of Activities

Profitability indicator	**Low-level properties**	**Example references**
Fixed Cost of Human Labor (FLC)	Fixed cost rate for each process with execution-independent costs such as rent or equipment	(Callahan et al. 2011; Hansen et al. 2007)
Variable Cost of Human Labor (VLC)	Variable cost rate for each process execution with a certain degree of human interaction based on the relative amount of salary payments	(Callahan et al. 2011; Hansen et al. 2007)
Fixed Cost of RPA (FRC)	Fixed cost rate to configure and maintain robots, depending on process's stability and degree of standardization as well as the cost of writing code	(Grand View Research 2018), Practitioner talks
Variable Cost of RPA (VRC)	Variable cost rate for each process step execution based on cost agreement(s) with RPA service vendor(s)	(Grand View Research 2018), Practitioner talks

To clarify the understanding of the economic comparison, Figure 2 illustrates the two cost functions of HL and RL in relation to the number of process executions and process variants.

Controlling for a process's automation rate, the left part of Figure 2 illustrates the impact of a process degree of standardization on the costs of RPA. More variants increase the costs for implementing and configuring a robot and thus diminish the returns of an RPA project. This also applies to the stability of a process, which serves as an indicator of the probability of exceptions and predictable outcomes. The negative impact of the number of process variants and instability diminishes with higher automation rates. The right part of Figure 2 captures the profitability of automating a process as a function of costs for human labor and RPA. The cost function of human labor is characterized by lower fixed costs but increases strongly with each process execution. By contrast, the costs function of RPA starts out at a higher value that accounts for the fixed costs imposed by implementing an RPA robot but has a less significant increase in cost with more process executions. Because of $VLC_{im} > VRC_{im}$, process automation becomes economical at the functions' point of intersection, shaping a *line of automatability* where the variable costs of both alternatives break even. The size of the dotted area indicates the overall economic value of automating a process, as a function PA_{im} of the predefined indicators.

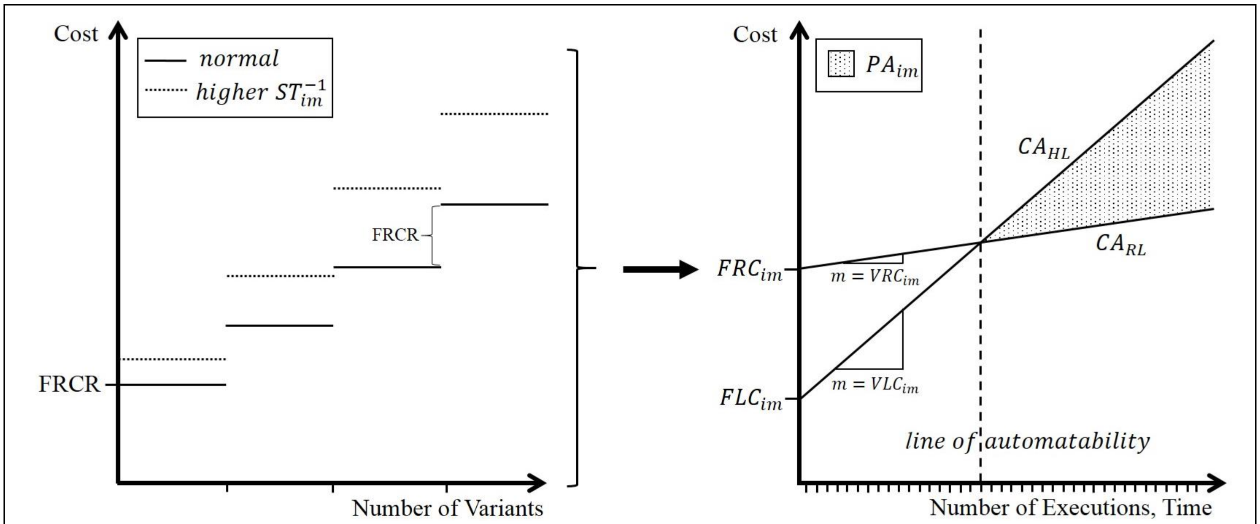

Figure 2: Profitability Analysis

Maximizing the Economic Value of RPA Projects

From an economic perspective, we assume that companies act rational and seek to maximize the economic return of an RPA project (equation 13), by deciding whether an activity $\bar{A}_{im}$ should be automated ($x_{im} = 1$ or $x_{im} = 0$). Thereby, companies typically face three constraints: *a budget constraint* (equation 14), an *amortization constraint* (equation 15) and *a completeness constraint* (equation 16). Hereby $\overrightarrow{PA}$ and $\overrightarrow{FRC}$ denote row vectors with entries PA_{im} and FRC_{im}respectively, whereas $\vec{x}$ is a column vector with entries x_{im}.

$$\max \quad \overrightarrow{PA} * \vec{x} \tag{13}$$

$$\text{s. t.} \quad \overrightarrow{FRC} * \vec{x} \leq B \tag{14}$$

$$\frac{-FLC_{im} + FRC_{im}}{VLC_{im} - VRC_{im}} * x_{im} \leq t \text{ for all activities } (i, m) \tag{15}$$

$$\vec{x} \in \{0,1\}^n \tag{16}$$

Equation 14 implies that RPA is naturally limited by a project budget B, which captures the available financial resources that can be used for the implementation, configuration, and maintenance of robots. Equation 15 implies that the RPA project should amortize at latest after the time t. In practice, the project should typically pay off after 6 to 36 months. We frame the resulting optimization problem for maximizing the economic value of RPA as a *zero-one knapsack problem* (Toth 1980). This entails weighing all relevant activities with their potential returns and costs of automation and building a subset of processes that maximizes the returns of RPA while remaining within the boundaries of the given restrictions. Although other solutions exist, we recommend employing dynamic programming to solve the optimization problem in pseudo-polynomial time. The resulting x_{opt} is a vector that indicates, which processes should be considered for automation.

Application and Evaluation

This section demonstrates and evaluates our method for process selection in RPA projects. In an ideal scenario, the underlying information system is a PAIS and creates a data log for all activities. The experts stated, that process logs are a crucial foundation to identify suitable candidates for RPA projects, as it gives a 'much better background and reliability for people to discuss on'.

In the first step we conducted a technical evaluation to show that the identified key figures can be extracted from a variety of process logs. Therefore, we use process execution data from a Dutch financial services institute, building permit application logs from different Dutch municipalities and a Dutch hospital, which are provided by the *4TU Centre for Research Data*[1]. We demonstrate our approach using the logs from the financial institute. More specifically, the dataset contains approximately 45,000 cases and 450,000 events stemming from the company's IT service management processes. In addition to activity names, timestamps, and users, the event log provides various features, including the urgency of a service request, its alert status, or a case's open and closing time. Several authors further note that the composition of event logs can vary from case to case and depends strongly on the goals of analysis and the capabilities of the underlying information system (van der Aalst 2016). For our quantifiable method we only require the common features of *activity name* and *timestamp*, which mandatory attributes in all process logs we encountered.

To operationalize the event log for our evaluation example, we must break down the event log into process tasks (process discovery). This entails deriving the control flow of a process task based on its time-related and structural features without any a-priori knowledge about its composition. Research has proposed several methods for process discovery, including the *α-algorithm*, the *fuzzy miner*, or the *heuristic miner (van Dongen et al. 2009)*. Each method has strength and weaknesses, whereby its suitability depends strongly on the underlying research problem (Song et al. 2009). In this study, we draw upon the *trace clustering* approach as proposed by (Song et al. 2009). Trace clustering builds upon concepts of text mining and interprets cases as strings. It uses similarity metrics and hierarchical agglomerative clustering to group event logs into homogenous clusters. Each cluster represents a process, whereby the cases within that cluster are defined as executions of that process. Furthermore, we use bi-grams and tri-grams to incorporate contextual information of activities and use the *Calinski-Harabaz index* to determine a suitable cluster solution (Caliński and Harabasz 1974). Based on this configuration, trace clustering yields a total of 29 IT service processes with overall 723 process steps.

First, we determine the automation potential of process activities, which entails computing the indicators introduced in Section 5.2 (cf. 'Determining the Automation Potential of Processes'). Assessing the indicators of *execution frequency*, *standardization*, and *stability* is straightforward. While many BPM systems indicate whether a process task was performed manually or automated, this event log did not explicitly flag automated tasks. Hence, we assume, that activities that have identical start- and end-times are automated, and the rest is performed manually. This leaves us with 530 unautomated activities, that could be considered for automation.

Summarizing the results in Figure 3, we suggest that the automation potential of the company's processes varies significantly. While a considerable amount of activities is characterized by a high execution frequency, most other processes are performed rarely shaping a long-tailed distribution. The scores in Figure 3 are independently sorted to provide an overview. Processes with high indicator values in one category do not imply high value in other categories.

Having assessed the tasks automation potential, we analyze the profitability of task automation. As described in Section 5.3 (cf. 'Analyzing the Profitability of Process Automation'), the profitability analysis entails comparing fixed and variable costs of human labor with the costs imposed by implementing and maintaining robots as well as determining the execution frequency for which automation becomes beneficial. For reasons of clarity, we assume a cost rate for human labor of 25 € ($VLCR$) per hour and fixed

[1] For more information, see https://data.4tu.nl/.

costs of 1,000 € ($FLCR$). Thereby, we consider analyzing the profitability of RPA as a case-by-case challenge and recommend companies to perform a thorough analysis of their labor costs to update these placeholder values with their cost rates. Process mining be used to conduct a more differentiated cost analysis, for example to examine data on the handover of work to identify involved employees as well as the time spent on executing activities in a process. We further assume the variable costs of RPA ($VRCR$) at 0.01 € per execution and a cost rate for line of code of 20 € ($FRCR$) for the implementation, configuration, and maintenance of robots. Again, we do so for reasons of simplification but note that companies must conduct a detailed analysis of the RPA market and request information from RPA vendors to acquire more accurate values. If robots are implemented with fixed capacity, an additional robot must be purchased when the capacity is exceeded. This would result in a step cost function making the optimization problem more complex. Therefore, our model considers robots to be elastic with near infinite capacity based on cloud computing. Fixed costs of RPA further vary with the stability, automation rate, and degree of standardization of a process. Based on the assumed cost parameters, we perform our profitability analysis and present the results in Table 4. The last column specifies whether it is profitable to automate tasks.

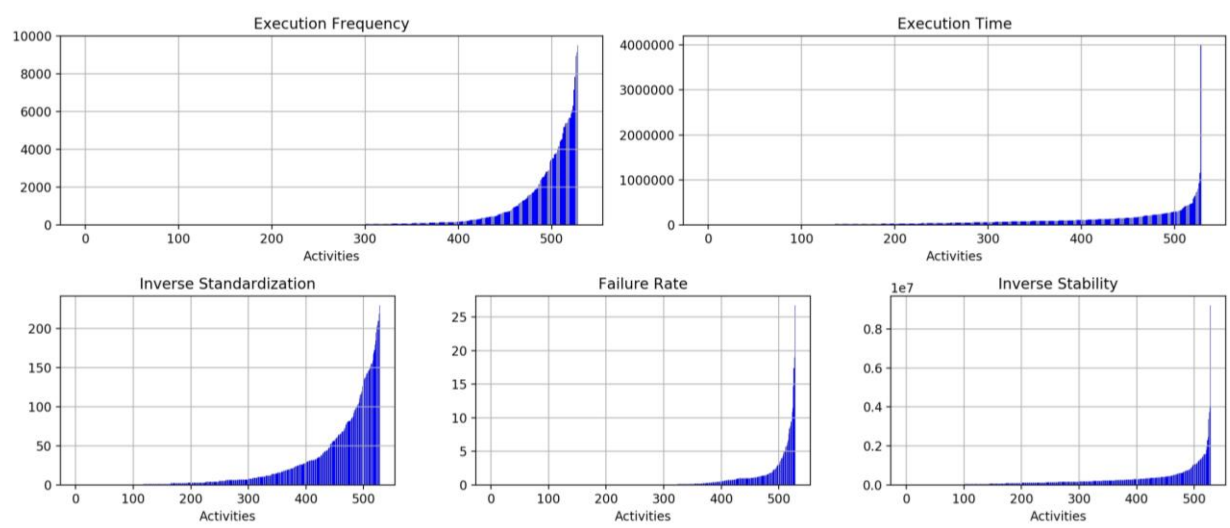

Figure 3: Assessing the Automation Potential of Processes' Activities

Because the processes are characterized by different costs for human labor and automation, they also differ regarding their suitability for RPA. As shown in Table 4, our results suggest that the tasks 529 and 530 are promising candidates for automation while automating processes 1 and 2 takes too long to amortize. The third step of our method supports companies in maximizing the economic value of RPA by determining the optimal number of tasks to be automated under the constraint of a given budget. Therefore, we solve the optimization problem introduced in the previous section, where we assume the RPA budget to be 1,000,000 €.

In this case, only tasks with an amortization time of less than three years are considered for optimization. The resulting RPA project includes a total of 14 processes and yields a total profit of 1,699,496 € with costs at 935,290 € within the next three years. Consequently, the company faces a significant potential for cost reduction when engaging with RPA.

The technical evaluation showed that the process mining approach produces meaningful key figures and shows reasonable profitability values. We additionally implemented a demonstrator that we evaluated with five experts. The experts stated that the key figure system and profitability optimization yields a good addition to the manual selection process for RPA projects. The key figure system gives an objective overview on how well activities are suitable for RPA projects and gives a suitable discussion basis for decisions. The main critique was the low availability of process logs for most clients, especially for small and medium enterprises (SME). Additionally, they mentioned that the key figure system is not exhaustive for every application, but could require additional key figures, if the automatization goal is not saving cost. RPA projects can have other goals such as customer satisfaction.

Table 4: Indicators for Process Profitability Analysis							
Task ID	FLC_i	VLC_i	FRC_i	VRC_i	CA_{HL}	CA_{RL}	**Return**
1	1,000	2,169,501	2,490,811	34.4	2,170,501	2,490,845	6.21 years
2	1,000	4,555,255	6,899,536	13.6	4,556,255	6,899,550	4.80 years
…	…	…	…	…	…	…	…
529	1,000	360,947	101,353	9.8	145,449	100,362	2,91 years
530	1,000	511,691	33,821	52.4	361,947	33,873	0.53 years

Conclusion, Limitations, and Future Research Opportunities

Our goal was to develop a quantifiable method to support companies for process selection in RPA projects. We employed DSR, which guided us in analyzing a problem domain and designing as well as evaluating a corresponding IT artifact (Hevner et al. 2004; Peffers et al. 2007).

Based on a literature analysis and practical requirements of RPA case studies, we identified indicators for process selection in RPA projects. In two iterations, we improved the method by involving expert knowledge. The resulting method is based on a pre-selection of relevant processes and consists of three steps: (1) determining the automation potential of processes, (2) analyzing the profitability of process task automation, and (3) maximizing the economic value on RPA projects. The method returns a list with quantified indicators and recommendations to support the subsequent decision-making. Our theoretical contribution is – to the best of our knowledge – the first comprehensive method that uses a software- and industry-independent collection of indicators.

Thereby, our method has immediate practical value and operationalizes techniques from process mining. It enables companies to analyze the execution frequency of process tasks, their degree of stability, their standardization, their failure rate, their average activity time, as well as their automation rate and thus, to assess their automation potential for RPA. A profitability analysis that incorporates both fixed and variable costs of human labor and robots helps companies to identify suitable process tasks and determining the optimal number of process tasks that can be automated within a given project budget or within a given amortization time. We have evaluated our approach based on a real-life event log and demonstrated its applicability and usefulness. We have not found any indication that the indicators cannot be transferred to other industries than those surveyed as no one of them is industry-specific. Nevertheless, adding industry-specific indicators for certain industries may enhance the results.

This research is not without limitations. First, we did not analyze the literature with the goal of building an exhaustive collection of indicators on RPA but rather to acquire an understanding of the main challenges associated with it. Hence, we cannot eliminate the possibility that we have neglected insights that would have offered additional perspectives and requirements. Ultimately, we will have to test the indicators for deficit, redundancy, overload, and excess. Furthermore, while RPA is gaining traction in IT consulting, mature research on RPA is scarce and long-term profitability can only be estimated. Second, our analysis builds common event log configurations. As emphasized by van der Aalst (2016), the composition and structure of event logs can vary, which implies that additional features may be available in practice that could increase the quality of our results. Third, we analyzed process tasks independent of each other and neglected dependencies and reuse of tasks where automation in one process might influence other related process tasks. Fourth and last, we limited our analysis to a few determinants and exclusively focused on the economic value of RPA. Thereby, we neglected social and organizational implications such as a changing nature of work and the need for new skills and competencies.

Summarizing, we would like to emphasize that more multi-dimensional research is necessary to account for the consequences of RPA. Apart from accounting for social, cultural, and organizational perspectives, research on RPA should also focus on augmenting selection methods with an additional perspective that captures expected costs for change management when analyzing the profitability of RPA. Further, research should also consider the cost of fragmentation when parts of a process are being performed by robots while data integration is neglected in favor of front-end automation. Moreover, the results from the expert interviews provide a first indication for differences in the perceived importance of the indicators. Future research could use this as basis for further exploring weighted indicators to account for this. Ultimately, RPA also touches the issue of machine autonomy, which should be considered in the specification and implementation of the decision-making abilities of the robots.

References

Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "Research Commentary—the Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Aguirre, S., and Rodriguez, A. 2017. "Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study," *Applied Computer Sciences in Engineering*, Bogotá: Springer, pp. 65-71.

Anagnoste, S. 2018. "Setting up a Robotic Process Automation Center of Excellence," *Management Dynamics in the Knowledge Economy* (6:2), pp. 307-322.

Asatiani, A., and Penttinen, E. 2016. "Turning Robotic Process Automation into Commercial Success: Case Opuscapita," *Journal of Information Technology Teaching Cases* (6:2), pp. 67-74.

Caliński, T., and Harabasz, J. 1974. "A Dendrite Method for Cluster Analysis," *Communications in Statistics* (3:1), pp. 1-27.

Callahan, K. R., Stetz, G. S., and Brooks, L. M. 2011. *Project Management Accounting, with Website: Budgeting, Tracking, and Reporting Costs and Profitability*. John Wiley & Sons.

Davenport, T. H., Barth, P., and Bean, R. 2012. "How 'Big Data' Is Different," *Sloan Management Review* (54:1), pp. 43-46.

Dumas, M., Rosa, M. L., Mendling, J., and Reijers, H. A. 2018. *Fundamentals of Business Process Management*, (2nd ed.). Berlin: Springer.

Dunie, R., and Tornbohm, C. 2017. "Market Guide for Robotic Process Automation Software," Gartner.

Fung, H. P. 2014. "Criteria, Use Cases and Effects of Information Technology Process Automation (Itpa)," *Advances in Robotics & Automation* (3), pp. 1-10.

Georgakopoulos, D., Hornick, M., and Sheth, A. 1995. "An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure," *Distributed and parallel Databases* (3:2), pp. 119-153.

Geyer-Klingeberg, J., Nakladal, J., Baldauf, F., and Veit, F. 2018. "Process Mining and Robotic Process Automation: A Perfect Match," *16th International Conference on Business Process Management, CEUR Workshop Proceedings 2196*, Sydney, pp. 124-131.

Grand View Research. 2018. "Robotic Process Automation (RPA) Market Size, Share & Trends Analysis Report by Type (Software, Services), by Application (BFSI, Retail), by Organization, by Services, by Region, and Segment Forecasts, 2018 - 2025," San Francisco, CA, pp. 1-115.

Hansen, D., Mowen, M., and Guan, L. 2007. *Cost Management: Accounting and Control*. Cengage Learning.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Imgrund, F., Fischer, M., Janiesch, C., and Winkelmann, A. 2017. "Managing the Long Tail of Business Processes," in: *25th European Conference on Information Systems*. Guimarães: pp. 595-610.

Imgrund, F., Fischer, M., Janiesch, C., and Winkelmann, A. 2018. "Approaching Digitalization with Business Process Management," in: *Multikonferenz Wirtschaftsinformatik 2018*. Lüneburg, pp 1725-1736.

Koch, M., Manuylov, I., and Smolka, M. 2019. "Robots and Firms," *CESifo Working Paper* (No. 7608).

Lacity, M., Willcocks, L. P., and Craig, A. 2015a. "Robotic Process Automation at Telefonica O2," *The Outsourcing Unit Working Research Paper Series* (No. 15/02).

Lacity, M., Willcocks, L. P., and Craig, A. 2015b. "Robotic Process Automation: Mature Capabilities in the Energy Sector," *The Outsourcing Unit Working Research Paper Series* (No. 15/06).

Lacity, M. C., and Willcocks, L. P. 2016. "A New Approach to Automating Services," *MIT Sloan Management Review* (Fall 2016).

Li, C., Reichert, M., and Wombacher, A. 2008. "Mining Process Variants: Goals and Issues," *IEEE International Conference on Services Computing (SCC)*, Honolulu, HI, pp. 573-576.

Lindström, J., Kyösti, P., and Delsing, J. (eds.). 2018. "European Roadmap for Industrial Process Automation," Norrbrotten: ProcessIT.EU.

Marrella, A. 2018. "What Automated Planning Can Do For business Process Management," *15th International Conference on Business Process Management*, Barcelona: Springer, pp. 7-19.

Matt, C., Hess, T., and Benlian, A. 2015. "Digital Transformation Strategies," *Business & Information Systems Engineering* (57:5), pp. 339-343.

Mendling, J., Decker, G., Richard, H., Hajo, A., and Ingo, W. 2018. "How Do Machine Learning, Robotic Process Automation, and Blockchains Affect the Human Factor in Business Process Management?," *Communications of the Association for Information Systems* (43:19), pp. 297-320.

Moffitt, K. C., Rozario, A. M., and Vasarhelyi, M. A. 2018. "Robotic Process Automation for Auditing," *Journal of Emerging Technologies in Accounting* (15:1), pp. 1-10.

Myers, M. D., and Newman, M. 2007. "The Qualitative Interview in Is Research: Examining the Craft," *Information and organization* (17:1), pp. 2-26.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Prahalad, C. K., and Ramaswamy, V. 2004. "Co-Creation Experiences: The Next Practice in Value Creation," *Journal of Interactive Marketing* (18:3), pp. 5-14.

Scheer, A.-W., Abolhassan, F., Jost, W., and Kirchmer, M. E. 2004. *Business Process Automation*. Heidelberg: Springer.

Schmitz, M., Dietze, C., and Czarnecki, C. 2019. "Enabling Digital Transformation through Robotic Process Automation at Deutsche Telekom," in *Digitalization Cases: How Organizations Rethink Their Business for the Digital Age,* N. Urbach and M. Röglinger (eds.). Cham: Springer International Publishing, pp. 15-33.

Sebastian, I. M., Ross, J. W., Beath, C., Mocker, M., Moloney, K. G., and Fonstad, N. O. 2017. "How Big Old Companies Navigate Digital Transformation," *MIS Quarterly Executive* (16:3), pp. 197-213.

Song, M., Günther, C. W., and van der Aalst, W. M. P. 2009. "Trace Clustering in Process Mining," *6th International Conference on Business Process Management*, Milan: Springer, pp. 109-120.

Toth, P. 1980. "Dynamic Programming Algorithms for the Zero-One Knapsack Problem," *Computing* (25:1), pp. 29-45.

van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., and Baier, T. 2012. "Process Mining Manifesto," *Business Process Management Workshops,* Clermont-Ferrand: Springer, pp. 169-194.

van der Aalst, W., Weijters, T., and Maruster, L. 2004. "Workflow Mining: Discovering Process Models from Event Logs," *IEEE Transactions on Knowledge and Data Engineering* (16:9), pp. 1128-1142.

van der Aalst, W. M. P. 2016. *Process Mining: Data Science in Action*, (2nd ed.). Berlin: Springer.

van der Aalst, W. M. P., Bichler, M., and Heinzl, A. 2018. "Robotic Process Automation," *Business Information Systems Engineering* (60:4), pp. 269-272.

van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M., and Verbeek, H. M. W. 2007. "Business Process Mining: An Industrial Application," *Information Systems* (32:5), pp. 713-732.

van Dongen, B. F., Alves de Medeiros, A. K., and Wen, L. 2009. "Process Mining: Overview and Outlook of Petri net Discovery Algorithms," in *Transactions on Petri Nets and Other Models of Concurrency II,* K. Jensen and W.M.P.e. van der Aalst (eds.). Berlin: Springer, pp. 225-242.

vom Brocke, J., Zelt, S., and Schmiedel, T. 2016. "On the Role of Context in Business Process Management," *International Journal of Information Management* (36:3), pp. 486-495.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. xiii-xxiii.

Willcocks, L., Lacity, M., and Craig, A. 2017. "Robotic Process Automation: Strategic Transformation Lever for Global Business Services?," *Journal of Information Technology Teaching Cases* (7:1), pp. 17-28.

Willcocks, L. P., Lacity, M., and Craig, A. 2015a. "The IT Function and Robotic Process Automation," *The Outsourcing Unit Working Research Paper Series* (No. 15/05).

Willcocks, L. P., Lacity, M., and Craig, A. 2015b. "Robotic Process Automation at Xchanging," *The Outsourcing Unit Working Research Paper Series* (No. 15/03).

Yin, R. K. 1989. *Case Study Research: Design and Methods*. Newbury Park, CA: Sage Publications.

Zysman, J., Murray, J., Feldman, S., Nielsen, N. C., and Kushida, K. E. 2011. "Services with Everything: The ICT-Enabled Digital Transformation of Services," in: *Berkely Roundtable on the International Economy*. Berkeley, CA.