

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

Use of Process Mining in ERPs:

A model for the use of Process Mining techniques
to improve ERPs usability.

João Pedro Rolo Lucas

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**USE OF PROCESS MINING IN ERPS:
A MODEL FOR THE USE OF PROCESS MINING TECHNIQUES TO
IMPROVE ERPS USABILITY**

by

João Pedro Rolo Lucas

Dissertation report presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Business Analytics

Supervisor: Professor Doutor Vítor Manuel Pereira Duarte dos Santos

November 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 27-11-2023

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Dr. Vitor Santos for his guidance and patience, it truly made a difference on helping me conclude this thesis. I would also like to thank the two Nova professor that took their personal time to answer my interviews that helped me finish my research.

ABSTRACT

Enterprise Resource Planning (ERP) Systems collect data and information across the enterprises, helping users and managers perform operations and processes more efficiently while also allowing for more informed decision making. Process Mining on the other hand is a specific technique of data mining that is process-centric and uses event data to provide a holistic view of the different processes. ERPs produce event log data suitable to be used in Process Mining to better understand the different processes of a particular business. This study presents a model that consists of an abstract representation of an architecture that efficiently integrates process mining techniques within the workflow on an ERP system improving its usability and increasing its value to the organizations.

This research follows a design science research methodology to create an artifact through a systematic literature review that supports the design and development of a model considered to be the artifact. The possible applications of the model are demonstrated through a use case, and its evaluation is done recurring to interviews to experts. The model presents an alternative to integrate process mining into ERP systems that can be adopted by organizations and works as base ground to incorporate other technologies such as machine learning and robotic process automation (RPAs).

KEYWORDS

Process Mining; ERP; Automation Technologies; Data Mining.

INDEX

1. Introduction.....	1
1.1. Context	1
1.2. Motivation	2
1.3. Objectives.....	3
1.4. Study Relevance and Importance.....	3
2. Literature review	4
2.1. Enterprise Resource Planning (ERP)	4
2.1.1. Concepts.....	4
2.1.2. Types / Taxonomy	6
2.1.3. Challenges and Opportunities	6
2.2. Process Mining	7
2.2.1. Concepts.....	7
2.2.2. Process Discovery	9
2.2.3. Process Conformance.....	11
2.2.4. Process Enhancement	12
2.2.5. Process Mining issues.....	13
2.2.6. Process Mining & ERPs	14
3. Methodology	16
3.1. Design Science Research	16
3.2. Research Strategy	18
4. Results and discussion	21
4.1. Assumptions	21
4.2. Model.....	23
4.3. Use case	28
4.4. Benefits.....	30
4.5. Evaluation & Discussion	31
5. Conclusions.....	33
5.1. Synthesis of the Developed Work	33
5.2. Limitations.....	34
5.3. Future Work.....	34
References.....	36
Annexes (optional)	39

LIST OF FIGURES

Figure 1: Example of a Workflow Net	10
Figure 2: Example of Alignment-based conformance checking.....	12
Figure 3: Design Science Research model	18
Figure 4: Proposed Model Outline	24
Figure 5: Detail of the Process Mining Architecture of the proposed model	25
Figure 6: Expected Process Model of a Purchasing Process	29

LIST OF TABLES

Tabela 1: Expertise Interviews	31
--------------------------------------	----

LIST OF ABBREVIATIONS AND ACRONYMS

BPEL	Business Process Execution Language
BPI	Business Process Information
BPM	Business Process Management
BPMN	Business Process Management Notation
DSR	Design Science Research
EPC's	Event Driven Process Chain
ERP	Enterprise Resource Planning
IS	Information Systems
RPA	Robotic Process Automation
UML	Unified Modeling Language

1. INTRODUCTION

1.1. CONTEXT

One of the critical objectives of enterprises is to make effective decisions that add value to the company and maximize their potential. These accurate decisions are dependent on reliable information sources and relevant data that is provided by information systems, such as ERP's (Kahraman, Kaya, & Çevikcan, 2011). They help the organization and users save time by aggregating massive data sets from the entire company, thereby increasing the ability of calculations that results in more informed decision-making (Jenab, Staub, Moslehpour, & Wu, 2019).

According to (Jenab et al., 2019), an ERP, is an information technology system that collects information from all the enterprise and allows managers to control and monitor their companies operations and processes. It's a package of software products that are combined to support business processes such as budgeting, order fulfilment, finance, human management, production, supply chain, logistics, sales, and customer service (Amalnick, Ansarinejad, Nargesi, & Taheri, 2011).

ERP systems help create management systems that optimize operation flows in short production periods, reduce costs, accelerate fund turnover, and increase production and service quality (Ma, 2009). The competitive position of an organization may be enhanced by implementing the ERP system, successfully, based on well-designed strategy and satisfaction of customer needs, which can assist company's in reducing costs and increasing income (Ahmad & Cuenca, 2013).

ERPs are an effective way to control company operations (Chapman & Kihn, 2009). Furthermore, machine learning and predictive analytics can assist ERPs in improving processes and predictions, enhancing planning for company operations via learning from experience, and adapting business rules (Rizza, 2016).

ERP can also be seen as one of the areas of expertise of process automation, a direct result of process mining. Process mining is therefore a method that allows a better understanding of the different phases of managing a particular business process. It uses data generated by information systems, event log data, to provide a more holistic view of the process (Der Aalst & Mining, 2011).

The main objective of process mining is to discover, monitor and improve business processes by extracting knowledge from event logs produced by information systems in real time and readily available (Mahendrawathi, Zayin, & Pamungkas, 2017).

This is a fast-growing field of search that combines other important data related fields of expertise, it is the result of a conjunction of data mining and machine learning with process modeling and analytic that can be used for three main purposes: process discovery, process conformance checking and process enhancement (Mahendrawathi et al., 2017).

The purpose of process discovery is to create a process model without any previous knowledge of the process itself based on event logs. The second purpose, process conformance checking has the goal of comparing the reality of already existing processes with the information generated by event logs, checking they conformance. Finally, the third type of process mining, tries to improve already existing

process using the information and knowledge extracted from the event logs (Mahendrawathi et al., 2017).

1.2. MOTIVATION

The growth of event data is currently a topic in information systems field. ERP systems manage large volumes of operations, and the proper analysis of event data can give a valuable knowledge of how business processes are carried out.

Analyzing this data help us gain insights into process execution and to identify the gaps between how the systems are planned to be used and how the employees carry out the operations. The goal is to discover from the transaction logs a model of how the business processes are actually carried out.

ERP systems are information technologies with a complex utilization and implementation in enterprises. It is not always clear if the best approach and strategy is used when dealing with this systems implementation as well as how users learn to use them in different scenarios and different enterprises with different business needs. Recent research has shown that ERP implementation rate of success is quite low and initial problems occur in the utilization of information systems after their implementation (Arnety, Musiega, & Juma).

On the other hand, Process mining is a new emerging type of data analysis techniques with the goal to discover a process model when an event log is given. According to (W. Van Der Aalst et al., 2011) it represents a group of techniques to “discover, monitor and improve real processes (not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems”.

Process mining offers a huge range of possibilities and techniques related to it have matured a lot in the last decade (W. Van Der Aalst, 2012b). The starting point of every process mining research and project is an event log. Each event log refers to an activity and is related to a particular process instance.

Information systems, have the particularity of producing huge amounts of data in the form of event logs, that can be analyzed as real time information. Therefore, ERP systems and the data they produce represent a rich supply of data to nourish a healthy and efficient process mining model, resulting in a smarter and more efficient ERP.

Process Mining is therefore a technique that is process-centric, learns from historic data and is based on event data instead of opinions. Functioning as a tool to redesign, control and support operational business processes (W. Van Der Aalst, 2012b). This search has the objective of applying this previously mentioned attributes of process mining into a very specific business process, ERP systems utilization. Process mining allows a more detailed analysis of business flows as well as to identify bottlenecks and understand deviations, more specifically it could use the event logs to improve the usage and implementation of ERP systems. Therefore, we can formulate the following research question that will be the focus of this study:

RQ: Can we use Process Mining techniques to improve ERP usability?

1.3. OBJECTIVES

The research objective is to propose a model that integrates Process Mining Techniques into ERP systems to improve its usability.

In order to reach the defined goal, the following intermediate objectives were created:

- Make an extensive and comprehensive literature review on Process Mining.
- Make an extensive and comprehensive literature review on ERP.
- Propose a model for the use on Process Mining techniques to improve ERP usability.
- Build a use case (demonstration).
- Validate the process.

1.4. STUDY RELEVANCE AND IMPORTANCE

This study aims to improve ERP in all their essence by incorporating process mining techniques within this information systems with the objective of improving them, their usability and making them more capable of adding value to enterprises that integrate them in their business processes and operations. It is expected to add value to enterprises in general by merging two components that are emerging as key factors in information and knowledge management, improving the utilization of ERPs as a fundamental information system in technological business.

In further detail, the possibilities of process mining techniques applied to ERPs to improve their usability, can be applied to areas such as:

- Design
- Implementation
- Deployment
- Usability
- Maintenance

Process Mining allows for a possible improved design of a new ERP or migration to a new one, as it gives a holistic view of the processes as well as a low-level detailed analysis. Allowing to a more efficient and suitable architecture to different business needs.

During implementation or migration, process mining techniques could be implemented to easily monitor progress and evaluate if the project is going towards the right direction as well as comparing it with the initial design and other control metrics, facilitating the process and increasing the success rate.

In the deployment phase, process mining ensures that system users learn the new processes and track how well they are using the ERP, improving the usability, and decreasing the learning curve of new users, as well as detecting the main difficulties and barriers the users find. While in the maintenance phase process mining provides a holistic view that allow to effectively detect the problems and understand their root causes, leading to a constant monitor and improvement of the ERP.

Finally, ERPs allow to the automation of core businesses processes. Process mining techniques could help identify activities that are best suited for automation and the impact their automation will have.

2. LITERATURE REVIEW

In order to support this research, an extensive literature review and research was performed with the goal of having a deep understanding of the components and full capabilities of ERP systems and how they can be improved. Furthermore, Process Mining techniques and best practices are also deeply studied and understood.

Therefore, in this next chapter, a literature review about these concepts was done, firstly focused on ERP systems, then focused on process mining, and finally combining the two of them.

2.1. ENTERPRISE RESOURCE PLANNING (ERP)

According to (Marnewick & Labuschagne, 2005) an ERP can be defined as a business software system that automates and integrates different business processes, shares common data and allows to produce and access information in real time while maintaining information integrity across the organization business units. “The ultimate goal of an ERP system is that information must only be entered once.” (Marnewick & Labuschagne, 2005)

ERPs are complex information systems, and to better understand them an extensive literature review about the most important concepts of ERP's was done in the next section.

2.1.1. Concepts

An ERP is constituted by a different number of modules, according to (Beheshti, 2006) “An ERP system is a set of business applications or modules, which links various business units of an organization” (Beheshti, 2006). Modules are the core of an ERP software; they represent the different areas that interact within each other and is the component of an ERP that is most visible and has closer interaction with the user. ERPs typically offer a process-oriented view of the company, so users often do not realize with which specific module its working with (Klaus, Rosemann, & Gable, 2000).

Modules in ERPs represent distinct business units and business areas that are linked between them. The definition of modules and the number of them is not equal across ERP systems; it varies according to the vendor, needs and objectives of the enterprise implementing the ERP and the implementation strategy (Parr & Shanks, 2000).

Although these variations, some models are mentioned by multiple authors (Marnewick & Labuschagne, 2005; Shehab, Sharp, Supramaniam, & Spedding, 2004; van Giessel, 2004) and can be considered the most generic modules;

Finance and Accounting: This module incorporates all the finances of the enterprise, it deals with all the capital flows of the organization and provides functions like accounting, budgeting, payroll and managing of other financial data.

Sales and Distribution: This module includes customer related activities like orders, product configurations, pricing, promotions, and availability as well as distribution requirements, shipping schedules, billing, and invoices.

Materials Management: This module contains functionalities like material requirements, purchases, inventory management, warehouse management and invoice verification.

Human Resources: This module automates processes related with personnel management including recruitment, business travels, vacations, and trainings. It can also include payroll although this later is usually integrated in the financial module. It can also help with employee information such as attendance and performance.

Supply Chain Management: module that give a holistic view of the materials and information as they move through the process from the supplier to the retailer and to the consumer.

Customer Relationship Management: This module helps the company have a database that allows it to organize in an efficient way the relationships with its customers, so sales and management can be informed about clients needs, products offerings and have a holistic view of the customers purchases.

Business Intelligence: This module provides tools that enable real-time analysis of critical corporate information, allowing users to leverage information to understand relationships and trends that support business decisions.

Project management: It is a module with the purpose of controlling the different phases of the project, providing tools to resource management, cost settlement and planning and activity management.

Quality Management: This module ensures the quality of other company processes such as supply chain quality, management, and customer service.

Besides the modules, an ERP is composed by a Central DataBase that interconnects all the modules. It is a single and common database that centralizes all the information and allows it to be freely shared and transferred between modules, supporting all the company activities and making information coherent and accessible by all of them (I. J. Chen, 2001).

The Client/Server system of the ERP is the technology that allows for the communication between the client and the sever, the sever is usually centralized, while the customers are dispersed through different locations among the company (Yen, Chou, & Chang, 2002).

(Marnewick & Labuschagne, 2005) propose there are four conceptual components that characterize an ERP; Software, Process Flow, Customer mindset and Change management.

The ERP Software consists of a number of different modules and is the part that users can see the most, which is consequently perceived as the ERP product (Marnewick & Labuschagne, 2005).

The Process Flow referring to how information flows among the various modules of an ERP system, it is critical in understanding how the ERP works and guaranteeing that information is only entered once in the ERP software. Business process must be modelled in order to ensure a smooth integration and process flow between modules and software components (Marnewick & Labuschagne, 2005).

The customer mindset which refers to how the users perceive the ERP system and the changes it entails. When implementing new ERP systems or making changes to existing ones, it's normal to have resistance from users that are used to an already established way or work. "A paradigm shift or customer mindset change must be achieved." (Marnewick & Labuschagne, 2005).

Finally, the change management is the approach needed by the managers in order to successfully implement ERP systems. Change must be managed at different levels (Marnewick & Labuschagne, 2005).

2.1.2. Types / Taxonomy

ERP's are not equal across all organizations, nor there is a specific definition for them, they vary according to the needs of the organization implementing them and according to the offer given by the vendor (Parr & Shanks, 2000). Therefore there are different types of ERP systems to consider.

According to (Gulledge, Sommer, & Vincent Jr, 2005), ERP systems have developed into two modular paradigms; Single-Module ERPs and Multi-Module ERPs.

Single-Module packages are built to support a single business process and are designed to provide a high degree of functionality to a specific aspect of the business. Typically, they support an activity that has a very limited flow across other business units and is very specific, they require a substantial investment to interact with other business units. A typical example would be a single financial module which interacts with a range of different other business units and business processes and would require to be interfaced with these other systems.

Multi-Module software packages support a range of different areas of the company. The architecture defines the core modules and how the processes flow across them, thus representing a more complete view of the enterprise value chain (Gulledge et al., 2005).

Integration can occur in different ways, originating different types of modular ERPs, either it can be done as a whole where the full ERP is implemented at once, or it can be done module-by-module, responding to the initial needs of the organization and cementing the base modules and in a later stage more modules are added until all modules are implemented (Parr & Shanks, 2000).

2.1.3. Challenges and Opportunities

Although ERP's have a lot of advantages and are certain to bring competitive advantages to the enterprises that adopt them when well implemented, there are also drawbacks and challenges to consider.

(Shehab et al., 2004) argue that one of the biggest challenges is the cost and implementation of this software, according to (Scheer & Habermann, 2000) providers like PeopleSoft and SAP calculate that the cost of implementing an ERP system is three to seven times more than the software itself. ERP systems and their implementation are therefore large projects, with high costs and require large investment not only in capital but also in management and staff.

(Yen et al., 2002) also points out the high cost of these systems as one of the main challenges, stating that the high cost of ERP systems themselves joined with the implementation costs make them out of reach for small businesses. Privacy concerns and lack of trained staff are also one of the challenges, as many times there isn't a clear answer to who has the access to change information within the system or responsibilities regarding the different functions of the ERP are not well defined. Finally, customization is also one of the main drawbacks, as it is costly and time consuming.

ERPs have on the other hand, a lot of benefits to enterprises that adopt them, according to (Yen et al., 2002) ERPs have low operational costs after implementation is complete and contribute to improve customer service as well as business operations in general in other areas. According to (Yen et al., 2002) the benefits of ERP systems are:

- ERPs update data across different business components automatically, improving communication and business processes company wide.
- ERPs allow for a global approach and deal with international businesses that use different languages, currency, and accounting standards.
- Data integration, which allows for data analysis from different sources and in real time, allowing for a more unified management of data.
- Utilization of the most recent information technologies, allowing businesses to quickly adapt.
- Enable process optimization, as one of the main goals of ERPs is that data only needs to be entered once, allowing for operations efficiency.

2.2. PROCESS MINING

Process mining combines data mining and computer intelligence with process mining and analysis, the objective is to discover, monitor and improve real processes by extracting knowledge from event logs (W. Van Der Aalst et al., 2012).

2.2.1. Concepts

One definition of Process Mining according to (W. Van Der Aalst, 2012b) is that it is a research field focused on analyzing end-to-end processes using event data. This event data results from activities executed by people, software, and machines that leave a trail. Process mining aims to discover, analyze, and improve these trails left by activities and real processes, by extracting the knowledge about them available in Information systems (W. Van Der Aalst, 2012b).

"Process mining results can thus be viewed as X-rays revealing what really goes on inside processes and can be used to diagnose problems and suggest proper treatment." (W. Van Der Aalst, 2012a).

Process mining represents an important link between data mining with process mining and analysis, whereas unlike data mining techniques, which are mostly data-centric, process mining is process-centric (W. Van Der Aalst, 2012b). Therefore it enables technology for fields that study businesses processes and analysis with different purposes that look at the processes in detail to see if further improvements are possible, some examples of this fields of study are; monitor in real time business

processes, Business Activity Monitoring (BAM), process large amounts of events and use them to optimize the business in real time and set directions, Complex Event Processing (CEP), it also is related and has an influence in management approaches like Continuous Process Improvement (CPI) and Business Process Improvement (BPI) (W. Van Der Aalst et al., 2012).

Another important field of study where process mining plays has become crucial is Business Process Management (BPM), which tends to rely on hand-made models, whereas process mining research aims to link big data with BPM. Business processes are therefore managed, supported, and improved using event data rather than with subjective opinions or outdated experience (W. Van Der Aalst, 2012a).

The starting point for process mining are event logs, all process mining techniques assume that is possible to collect events sequentially. An event log refers to an activity or well-defined step in a process that stores information about this event and is related to a particular process instance that can be seen as “one run of the process” (W. Van Der Aalst, 2012a). Events can have all kind of attributes such as timestamp, activity names, size, transaction, resource usage, and person or device executing the activity information (W. Van Der Aalst, 2012a).

In further detail, Process Instance/Case like previously mentioned can be seen as a “one run of the process”. It represents an observation of the process, and it’s by analyzing the different instances of a process, that process models are created. Process mining start by analyzing event logs, which are associated to an activity and this activity is part of a process instance or case (W. Van Der Aalst, 2012a).

When adapting process instances to business, there are business processes that can be described as a set of activities or tasks that take a variety of clearly defined inputs to produces a valuable and clearly defined output for the customer. These activities have a specific order, across time and place with a well-defined start and end. A business process has a set of attributes and flow of steps, to achieve a task.

Process mining have different objectives and covers different areas, one of them being the control-flow perspective that aims to order the activities and find all the possible paths for the process instances and produce a process model, usually the result is expressed as a Petri Net, Business Process Management Notation (BPMN) or other process notation (W. Van Der Aalst et al., 2012).

According to (W. Van Der Aalst et al., 2012) there are a few guidelines that should be followed in order to ensure and facilitate the implementation of process mining in real life;

- Event data quality should be the main concern, like previously stated it is the staring point for process mining and therefore the storage and especially the quality of this logs should be a priority as the results of process mining depend on these inputs, some of the criteria to evaluate the quality of the event data are:
 - Events should be trustworthy.
 - Events should be complete.
 - Events should have a well-defined semantic.
 - Events should be safe, and respect concerns such as privacy and security.
- Event data extraction should be motivated and driven by concrete questions, before implementing any technique it is necessary to define what questions are meant to be

addressed, otherwise it is way more complicated to properly define which data is necessary and where to retrieve it.

- The chosen process modelling language should support basic control-flow elements, to accurately describe the process model as the process flow description is one of the most important aspect of the process model. There are a variety of process modelling languages, such as Petri Nets or BPMN and they need to be able to support concurrency (And/Split/Join), choice (XOR-splits/Join) and loops.
- The relation between the event in the logs and the activity, which is referring in the model, should be clear and not allow for ambiguity.
- Models should be approached as “abstractions of reality” and should be selected according with a well-defined purpose and audience, as the level of granularity or activities represented vary according with this.
- Process mining should be a continuous process, processes are dynamic and constantly changing, process mining shouldn’t be approached as a one-time activity, but instead as recurrent task of analysis and improvement.

2.2.2. Process Discovery

Process Discovery is the section of process mining that learns and constructs models based on observed events using event data (W. Van Der Aalst, 2012b).

Process models are used by organizations to handle cases. Such processes are sometimes enforced by the information system. In most cases, however, processes are informal and may not have been documented at all. Furthermore, even when processes are documented, reality may differ from what is expected and originally designed. As a result, it is critical to use event data to discover actual process models that can be analyzed (W. Van Der Aalst, 2012b).

According to (W. Van Der Aalst, 2012b) process discovery may be used in the organization to;

- Problem-solving discussions among organizations stakeholders and help reach consensus, as it is critical to have a shared understanding of the actual processes,
- To generate ideas for process improvement, as seeing the actual process and its problems stimulates re-engineering efforts,
- Model enhancement,
- Configure business processes management systems, using discovered processes as a template.

Process discovery techniques produce process modules based on event logs using algorithms, whereas a process discovery algorithm is a function that maps a process model so that it is representative of the behavior observed in event logs. The biggest challenge is to define such an algorithm and decide which process model should be generated (W. Van Der Aalst, 2016).

Process discovery is a type of a-posteriori technique, mining on the recorded logs and assuming that is possible to record the events sequentially, and some mining techniques may use additional information such as timestamp, resources or user (W. M. van der Aalst, 2010).

The log data needed for this technique is however stored in many ways, it can be either stored in systems that provide dedicated storage like work-flow management systems or be scattered across many tables. Therefore, in many applications of process mining and process discovery it is necessary to gather, merge and convert this data into a suitable format. (W. M. van der Aalst, 2010) gives an example and defends the use of MXML format, which is a format that can be easily imported into tools like ProM and implement plug-ins that gather information from different systems.

The objective of process discovery is to mine, discover and produce a model from the recorded event logs, (W. M. van der Aalst, 2010) gives the example of Petri nets and more specifically a subclass known as workflow nets to represent the process model as exemplified in figure 1, however exist other process modeling languages such as: BPMN, EPC's, BPEL and UML activity diagrams.

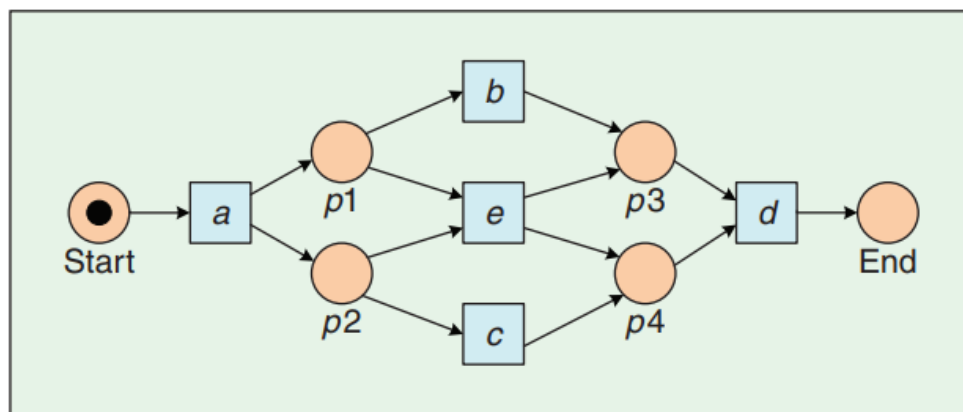


Figure 1: Example of a Workflow Net. Source: (W. M. van der Aalst, 2010)

One of the main algorithms in process discovery is the α -algorithm, which according to (W. M. van der Aalst, 2010) is not supposed to be used in practical applications, as it is very susceptible to noise, however it has been integrated in more complex and robust techniques and deals with the discovery of concurrency activities.

The α -algorithm searches the event logs for patterns and looks up for casual dependencies. In example, an activity a that is followed by b but where b is never followed by a reflects a casual dependency between a and b, in a petri net or other process model language this is represented by a connected to b (W. M. van der Aalst, 2010).

Although more robust algorithms using the α -algorithm already exist, according to (W. M. van der Aalst, 2010) there are still constraints and problems to process discovery like:

- Noise, which is incomplete or incorrect data in the logs, combined with a large number of possibilities to search for.
- Doesn't exist negative examples, meaning that the log data only represents what can happen, but never what can not happen.
- Due to loops and choices, the logs usually only have part of the total number of possibilities.
- There isn't a clear way to relate the size of the discovered model with its quality.

2.2.3. Process Conformance

Process Conformance or conformance checking takes an even log and a given model, whereas the model may have been built manually or through process discovery, and compares the observed behavior (log event) with the modeled behavior (W. Van Der Aalst, 2012b).

Conformance checking techniques relate activities in a model and relates them with the events in the log, making it possible to quantify differences and identify deviations, according to (W. Van Der Aalst, 2012b) process conformance can be used to:

- Assess the quality of documented processes and whether they accurately describe reality.
- Identify deviating cases and determine what they have in common.
- Identify process segments where most deviations occur.
- Auditing purposes.
- Evaluate the quality of a discovered process model.
- Starting point for model enhancement.

According to (W. Van Der Aalst, 2012b), there are four dimensions that should be considered when comparing a log event to a given model; fitness, simplicity, precision and generalization.

A model with high fitness can explain most of the behavior observed in the event log, therefore a model capable of replaying all traces in the log, from beginning to end is considered to have perfect fitness. Fitness is frequently described by a number between 0 (very poor fitness) and 1 (excellent fitness). Along with fitness, the model that can explain the observed behavior in the log events in the simplest way is considered the best model.

A precise model is one that doesn't allow for a too much scattered behavior. In a contrary way, a model that isn't precise is usually referred as "underfitting", meaning the model over-generalizes the behavior in the log event.

Although a model should be precise, it should also allow for generalization and not be restricted to only observed behavior in the log. Opposite to underfitting, a model that doesn't allow for generalization is "overfitting", meaning that a subsequent sample log of the same process should not yield an entirely different process model.

There are three approaches to conformance checking according to (W. Van Der Aalst, 2012b):

- Conformance Checking based on causal footprints, which is related to the α -algorithm used in process discovery and looks for casual dependencies. A footprint is a matrix that shows the casual dependencies between the different activities, during conformance checking this approach gets the footprint of a given log and compares it with the footprint of the proposed model, if any discrepancy between both footprints is detected, then it is possible to say there is some kind of deviation or underfitting from the model. It is a flexible approach, however it has some limitations, the frequency as deviation happen are not considered, the behavior is only considered indirectly and it captures precision, fitness and generalization in a single metric.
- Conformance Checking based on token-based replay, in this approach the event logs are replayed in the model and analyze how many completely parse the model. However, this wouldn't allow to distinguish between cases that almost fit the model with ones that are

completely unrelated. Therefore, in this approach event logs are continuously replayed through the given model, even when activities or steps are missing in the event log, these missing steps are replaced by tokens and allow for the model to be always completely parsed. In the end, the number of “borrowed tokens” indicate the fitness level. This approach allows for a very precise method to quantify fitness.

- Alignment-based conformance checking, computes an optimal alignment between each trace in log and the most similar behavior in the model. The log and the model are therefore aligned as much as possible and then the number of activities or situations where the log and the model can't be aligned are counted to quantify the fitness of the model. γ_1 , γ_2 , γ_3 present in figure 2, are examples of logs aligned with the model, in γ_2 and γ_3 there are two situations where the log can't mimic the behavior of the model and therefore we can conclude that the log and the model are not conforming.

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & c & d & e & h \\ \hline a & c & d & e & h \\ \hline \end{array} \quad \text{and} \quad \gamma_2 = \begin{array}{|c|c|c|c|c|} \hline a & d & c & e & h \\ \hline a & \gg & c & e & h \\ \hline \end{array} \quad \text{and} \quad \gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline a & d & c & e & f & d & b & e & h \\ \hline a & \gg & c & e & f_2 & \gg & b & e & h \\ \hline \end{array}.$$

Figure 2: Example of Alignment-based conformance checking. Source: (W. Van Der Aalst, 2012b)

When analyzing the conformance of a model, this can be done by looking at two perspectives. If the model is descriptive, the objective is to analyze if the given model is accurately when describing the reality and captures the real behavior. While when the model is normative, the objective is to evaluate if the reality deviates from the desired model (W. Van Der Aalst, 2012a).

2.2.4. Process Enhancement

It is also possible to enhance or extend a model using process mining techniques and information present on the event logs. An underfitting model can be corrected by observing the behavior and the results from process conformance and comparing the event logs with the model, for example it can make sense to skip some of the activities in the model (W. Van Der Aalst, 2012a).

Additionally, event logs can contain more data from each event, they can register information such as resources used, timestamps and case data (W. Van Der Aalst, 2012b). This way it is possible to replay the events and analyze this extra resources.

In example, when analyzing timestamps, by computing simple statistics it is possible to identify the main bottlenecks and points to improve. It is also possible to analyze resources, groups of people executing related activities and construct social networks that allow to identify the relation between workload and time spent to perform certain activities. Additionally standard classification techniques can identify the main decision points in the process model and help explain the different outcomes (W. Van Der Aalst, 2012b).

Process Enhancement is the stage where improvements are made to the model using previously found information and make the model more fit to the observed behavior so it describes the reality with better accuracy or make adjustments, so the observed behavior is closer to the expected model.

2.2.5. Process Mining issues

Process mining produces a range of interesting approaches when analyzing processes in different fields and offers different solutions and approaches to deal with challenges. In the same way, it also raises interesting questions and challenging problems, this section tries to enumerate and bring together the main challenges which may already have been answered or still require further research and attention.

1. *Hidden tasks*; in process mining one of the basic assumptions is that all the events are registered in the log, however it is possible that for some reason a task is not recorded when mining the process and therefore there are no information about their existence. In some cases, it might be possible to induce the existence of this tasks, however for more complicated processes it still poses as an interesting problem related to the observed behavior (W. M. Van der Aalst & Weijters, 2004).
2. *Duplicate tasks* refer to when in a process model there are two different nodes referring to the same task. The presence of duplicate tasks is linked to the presence of hidden tasks. Many processes with hidden tasks but no duplicate tasks can be converted into equivalent processes with duplicate tasks but no hidden tasks (W. M. Van der Aalst & Weijters, 2004).
3. *Mining Loops*, happen when in a process it's possible to execute the same task multiple times, thus referring to a loop in the model. Although easily identifiable, loops can be used to go back to any other part of the process. Another problem is that most mining techniques make assumptions about loops which may condition the classification of processes, however it is possible for a task to be executed multiple times but not be in a loop (W. M. Van der Aalst & Weijters, 2004).
4. *Dealing with complex event logs* having different characteristics, for example in many cases the log event also contains information about time in the form of a timestamp. This allows to model duration of each task and activity and have start events and end events which when compared to each other help determine processing times. This time information augments the model with more information and improves its quality. A proposed approach by (W. M. Van der Aalst & van Dongen, 2002) is to mine the process model first while ignoring timestamps and after that re-run the process with the timestamps, this way is easier to calculate averages, variance an minimum and maximum processing times for the different tasks.
5. *Noise*, one of the basic assumptions when analysing log data is this information to be correct. It may happen however the log to have incorrectly logged information, either because the event was only recorded with a delay or not recorded at all. A robust mining algorithm need to consider this incorrect information, or noise, and be able to distinguish between exceptions and the "normal flow", not making assumptions based on a single observation. A common approach would be to set a threshold to exclude exceptional behaviour (W. M. Van der Aalst & Weijters, 2004).
6. *Incompleteness*, like the noise problem is related with incorrect information, in this specific case information that is incomplete. "A log is incomplete if it does not contain enough information to derive the process." (W. M. Van der Aalst & Weijters, 2004). When mining a limited number of cases, it is possible that not all possible cases are recorded, and therefore resulting in hidden tasks or even incomplete models (W. M. Van der Aalst & Weijters, 2004).

7. *Gathering Data*, is a challenging task, information systems are increasingly complex and are composed by a variety of applications that support different parts of the process, resulting in scattered information across the different information systems of the enterprise.

2.2.6. Process Mining & ERPs

In this section a research about already existing approaches to combine process mining and ERPs are done, as well as what is necessary to integrate process mining into ERP systems.

2.2.6.1. Approaches

(W. M. Van der Aalst & Weijters, 2004) proposal to use process mining with ERPs is based on the fact that ERPs implementation require a specific process model to use as reference. Since these models are usually done by people and can be either normative or descriptive, process mining could be used to compare these models with the ones resulting from process mining. These comparisons could be done during implementation and use of ERPs. One of the techniques suggested is the usage of Delta Analysis algorithms to perform this comparison.

Business Process Intelligence (BPI) is also another approach, which was proposed by (Grigori et al., 2004), it provides a set of tools to apply business intelligence techniques to business processes that are analyzed using process mining techniques. BPI provides a wide set of features, with different automation levels to manage the quality of the processes, such as: analysis, prediction, monitoring, control, and optimization. The same authors propose an architecture to the BPI model that have a process mining engine at its base and can be applicable to other business process management systems or information systems.

2.2.6.2. ERP logs

Event logs are the starting point to process mining, and each event is related to a well-defined step in the process model (i.e. an activity), each event is also related to a process instance or case of the process and when ordered are considered one “run” of the process (W. Van Der Aalst, 2012b). Event logs can have all kind of attributes like timestamps or resources.

ERPs and other information systems record business events in what is considered an event log. This are usually referred as “history” and contain information related to an activity and a case (W. M. Van Der Aalst et al., 2007), some examples are transactions made in the ERP (i.e. filling out forms, changing documents), messages exchanges with other stakeholders or interactions with customers.

Therefore, according to (W. M. Van Der Aalst et al., 2007), ERP logs are expected to record events that:

- Each event refers to an activity or well-defined step of the process.
- Each event is referred to a case or process.
- It is possible to identify who performed the event.

- Events have timestamps and are ordered.

This event logs allow for auditing and diagnostic analysis and are used for process discovery, conformance and enhancement of the process model. After defining which business events are considered an activity and if events meet previous criteria, this are the event logs used in process mining.

(Lucas & Babaian, 2012) proposes a framework to more accurately capture user interactions with the ERP, the TIL framework consisting of three modules, the Task module, the Interface module and the Logging module. The task module consists of predefined tasks according to the system at use, where a task instance is the performance of a task by one or more users (i.e. a purchase requisition that is started by one user and then completed by another one), a process instance is the performance of the set of tasks, performed by one or more users. The Interface module defines components that populate the interface pages of the system (i.e. text fields, buttons and menus) and the Logging module captures the user-system interaction with the pre-defined components.

This framework proposes a separate approach to analyze the process model composed by the different tasks and process instances, and the user-system interaction. Requiring for these two distinct types of log data produced by the ERP system:

- Log Data, related to the performance of the tasks.
 - Case
 - Activity
 - Timestamp
 - Resources (i.e. user)
- Log Data, related to the interaction of the user with the interface components.

3. METHODOLOGY

This study aims to create a model that allows the application of process mining techniques to ERPs to make their usage easier. It is the typical problem of Information Systems (IS) that aim to create, apply, evaluate and improve Information Technology artifacts (Hevner & Chatterjee, 2010), where artifacts are human-machine systems that act in areas such as “support operations, management, analysis, and decision making functions in an organization” (Davis & Olson, 1984).

More specifically, and according to (A. R. Hevner, S. T. March, J. Park, & S. Ram, 2004) artifacts can be broadly defined as “constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems)”.

The aim of this study is to create a model which falls into the definition of artifact previously given, therefore Design Science Research (DSR) Methodology was chosen to conduct this study.

3.1. DESIGN SCIENCE RESEARCH

DSR is a relatively recent research approach that aims to construct and solve existing problems (Reubens, 2016). DSR main purposes are to; use previous knowledge to solve problems, to change and improve solutions already existing and to generate new insights and theoretical explanations.

There are a variety of approaches and the number of steps suggested to carry out this methodology vary according to different authors. According to (A. R. Hevner et al., 2004) there are seven guidelines to be followed, while in another study by the same authors (A. Hevner, S. T. March, J. Park, & S. Ram, 2004), they only suggest six guidelines.

The research made by (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007) aimed to define and design a methodology that would work as a framework for carrying out Design Search Research, and instead of comparing different views of different researchers they focused on a consensus-building approach to ensure it was based on well accepted elements. Therefore, they suggest six activities in a nominal order based on the contribution of seven papers where the authors substantially agree on common elements.

The activities following a nominal order are:

1. Problem Identification and Motivation.
2. Define the objectives for a Solution.
3. Design and Development.
4. Demonstration.
5. Evaluation.
6. Communication.

Each one of these steps was adapted to the context of this study. A brief explanation of each of this steps is done based on the research made by (Peffer et al., 2007).

1. Problem Identification and Motivation

In this first step the aim is to define a specific research problem and justify how the solution to this problem can add value. The solution is expected to be an artifact that provides an effective solution to the initial problem. By accurately justifying the value of the solution it helps the audience understand the reasoning behind the researchers approach to the problem.

Identified problems, however, are not necessarily objectives, as Design Research is also focused on incremental solutions. Therefore, after the problem is identified it is necessary a second part where performance objectives are defined.

2. Define the objectives for a Solution

"Infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible" (Peffer et al., 2007). These objectives can either be quantitative, where terms or thresholds are defined so the new solution is better than the already existing ones, or qualitative, where a description of how the new artifact is expected to improve and support the solution of the problem previously addressed. The objectives should be rationally inferred from the specified problem.

3. Design and Development

In this step of the methodology the artifact is created, which can be a construct, a model, a instantiation or *"new properties of technical, social, and/or informational resources"* (Järvinen, 2007). A design artifact can therefore be any object in which a research contribution is embedded in the design.

This activity includes determining the artifacts architecture and desired functionality, which requires a strong knowledge of theory that can originate a solution and an actual artifact.

4. Demonstration

This step is characterized by the use of the artifact to solve one or more instances of the problem, and demonstrate its efficiency. This could be accomplished using, simulation, proof, case study, experimentation, or other appropriate activity. It is required knowledge of how to effectively use the artifact.

5. Evaluation

This activity measures the efficiency of the artifact in solving the problem previously established. It involves comparing the objectives of a solution with the one's actually observed in the demonstration phase when the artifact was used. It requires the use of relevant metrics and analysis. Depending on the nature of the problem and artifact created to solve it, evaluation can take many forms. It could compare the functionality of the artifact with the solution objectives defines in activity 2, or it could evaluate quantitative performances and different metrics. It could also set and evaluate quantifiable measure of performance, in short evaluation *"could include any appropriate empirical evidence or logical proof"* (Peffer et al., 2007).

By the end of this activity, researchers can decide if they either go back to *"Design and Development"* to try to reach further improvements on the artifact, or if they continue to the next step.

6. Communication

This last activity consists in communicate the problem and its importance as well as the utility and novelty of the artifact and the rigor of its design. It's about informing other researchers and other relevant audiences about the importance and effectiveness of the new artifact.

Although this process is structured in a nominal order, it is not expected that the researchers always proceed in a sequential order or always start in activity 1. In fact they can start in different activities according to the nature of the research and move onwards.

A problem-centered approach would start in activity 1, while an objective-centered solution would start in activity 2. On the other hand, if the artifact already exists but hasn't been formally thought through as a solution for an explicit problem a design and development centered approach would start in activity 3, while a more client/context-initiated approach would start in activity 4. In figure 3 it is possible to understand the general model of DSR.

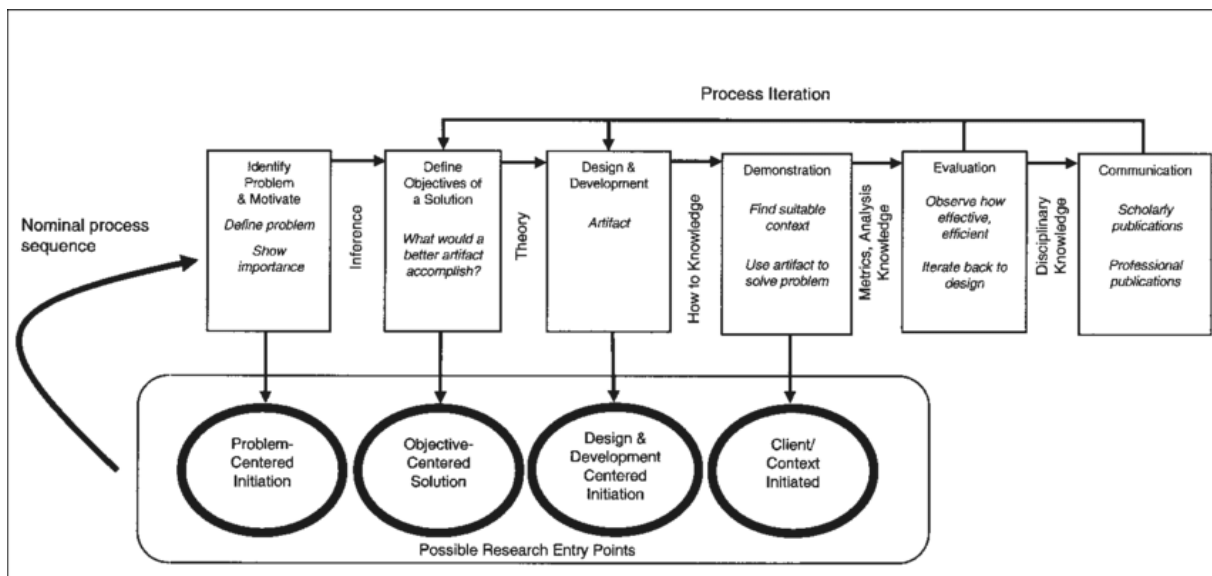


Figure 3: Design Science Research model. Source: (Peffers et al., 2007).

3.2. RESEARCH STRATEGY

The need for a model that integrates process mining techniques in Intelligent ERP systems and how to do it, represent the main purpose of this study. Therefore, this study will take a **Problem-Centered Approach** using the DSR methodology.

1. Problem Identification and Motivation

As more companies and businesses implement ERPs in their operations, the competitive advantage of owning one is gradually decreasing while the inadequacy of traditional decision support systems increases, originating the need of transforming information and knowledge in intelligence and broadening the range of applications of ERPs (R.-S. Chen & Tsai, 2005).

Enterprises that implement ERP systems accumulate massive amounts of data and log data (R.-S. Chen & Tsai, 2005), creating the possibilities to improve this systems through analysis of this data. Therefore, the need for improvement allied with promising process mining techniques based on log data generated by ERP systems originated the need for a model that can integrated process mining techniques in an ERP system.

2. Objectives of a Solution

The proposed artifact should be a model that allows the integration process mining techniques in an ERP. This model should accurately describe and suggest an architecture of how this implementation will be done as well as what are the main needs and restraints.

The main objective is to improve the utilization of log data generated by ERP systems and make their design, implementation, deployment, usability, and maintenance smarter, resulting in a competitive advantage for enterprises that adopt this artifact.

3. Design and Development

A deep understanding of the concepts involved in the creation of the artifact such as process mining, ERP systems and event log from an extensive literature review is the main knowledge requirements to develop the artifact.

The process mining project approach proposed by (W. Van Der Aalst, 2012b), allowed to make an initial approach, following a number of steps previously defined that conducted the initial part of the research, such as process discovery and process modelling.

The starting point to develop the artifact was to extract process mining data that is later transformed in event log data, which allows for visualization of the process. This allows a better understanding of the main processes as well as shadow processes that will be the main input and decision-making influence when designing the final model which will consist of the artifact.

4. Demonstration

In the demonstration phase, it is expected that the artifact is applied in different situations and specific scenarios which require the use ERP systems and allow for a practical use of the proposed model. This phase requires a bigger time period to collect different demonstrations in different situations.

The usage and implementation of the artifact and the consequent outputs and feedback generated by it, will be the main form of demonstration throughout this research.

5. Evaluation

In the evaluation phase, and based on the research made by (Fleig, Augenstein, & Maedche, 2018), some aspects were considered of major importance to be evaluated and decisive for the success of this research.

Given the characteristics of the artifact and the purpose of the study to develop a model, the feasibility and general appliance of the artifact to a broad number of business situations is one of the main evaluation topics. The number of incremental benefits of applying process mining techniques to ERPs,

comparing to the already existing ERP systems that do not incorporate these techniques is also subject of assessment as it will judge the efficiency of the artifact.

Given the artifact success is directly related with a smarter and easier interaction between the ERP and its users, the feedback given by them is another important topic of evaluation.

6. Communication

The divulgation of this artifact as a scientific paper can be a possible way to communicate it. Another possibility is making this research public so further improvements and suggestions can be done to the artifact itself.

4. RESULTS AND DISCUSSION

The purpose of this chapter is to give the bases to support the model proposed and justify why and how decisions were made. It is based on the knowledge gathered previously in Literature review about ERP systems and process mining techniques.

Therefore, it can be concluded that process mining is an emerging field of studies with growing importance and with opportunity for further developments. ERP systems on the other hand although already well established and well documented, have always faced great challenges in implementation and complexity, which represents an opportunity to integrate process mining techniques into ERP systems.

4.1. ASSUMPTIONS

The assumption made in this chapter connect concepts about ERP systems with Process mining techniques that culminate in a proposed model and architecture that uses process mining to improve ERP systems usability.

According to (Bahssas, AlBar, & Hoque, 2015), ERP has evolved deeply since its beginnings and with every technological advances, ERP providers develop new architectures and designs based on this technology, improving the implementation and accessibility of the ERP systems.

ERP systems future improvements are deeply connected with technological advances and have now the capability to process millions of transactions per second, creating opportunities to further and more quantitative improvements, “it is required to conduct more quantitative research for deeper understanding of users’ needs and thoughts” (Bahssas et al., 2015). Process mining is one of the logical paths to address these developments, as it requires large sets of log data that are produced and processed by recent ERP systems and allow for a better and quantitative understanding of user and how they interact with the ERP systems.

The definition and number of modules is not equal across all ERP systems and varies according to the vendor, to the needs and objectives of the enterprise implementing the ERP (Parr & Shanks, 2000). When applying process mining techniques, the number and type of modules are also not fixed and will vary for the same reasons.

The core modules however are frequently the same and are commonly mentioned by different authors (Marnewick & Labuschagne, 2005; (Shehab et al., 2004). The common ERP modules are therefore:

- Finance an Accounting,
- Sales and Distribution,
- Materials Management,
- Human Resources,
- Supply Chain Management,
- Customer Relationship Management,
- Business Intelligence,
- Project management,

- Quality Management.

The workflow of Enterprise information systems, including ERP systems, and how they are used are considered to be a modellable process and the life-cycle of a single case and utilization to be defined as a process instance (W. M. Van der Aalst & Weijters, 2004). It is assumed that it is possible to collect log data from this processes that allow for an adequate modulization of the observed behavior.

It is also assumed, that when implementing an ERP system for the first time, it is required a process model to use as reference. These models are usually hand-made previously to the implementation and can be either normative, trying to define a desired workflow or descriptive trying to describe how the users will interact with the ERP system. In either way, most of the times ERP's implementation is a human based process that can be improved by analyzing and controlling during the early stages of the implementation.

Process mining is the field that analyses processes using event data with well-defined objectives of discovering, analyzing and improving activities and process (W. Van Der Aalst, 2012b). Process mining allows for a deep understanding of processes, including the ones generated by ERP systems and its techniques are suitable to improve ERP's flexibility and problem diagnosis.

According to literature, it is assumed that the tasks of process mining are well defined and are as follows:

- Process Discovery; with the objective of finding, discover and construct models based on observed events using event data (W. Van Der Aalst, 2012b). It uses algorithms to produce process modules based event logs that represent the observed behavior (W. Van Der Aalst, 2016).
- Process Conformance; is the task of process mining responsible for comparing an event log with a given model, comparing the observed behavior with the modeled behavior (W. Van Der Aalst, 2012b), allowing to identify and quantify deviations. This comparison can be done either in a perspective to see if the model captures the real behavior or in another perspective to evaluate if the reality deviates from the desired model (W. Van Der Aalst, 2012a).
- Process Enhancement, the task of process mining with the objective of extending or improve an existing model based on information present in event logs. Models can be corrected or enhanced by comparing the observed behavior resulting from process conformance with the original model and adjusting. Additionally, event logs can contain extra data that by replaying past events allow to understand the model better, helping identify bottlenecks, points to improve and the main decision points (W. Van Der Aalst, 2012b).
Process Enhancement is therefore the stage where previously found information is used to adjust the model and make it more fit to reality or improve it, so the observed behavior is closer to the expected model.

4.2. MODEL

Based on the previous defines assumptions, the following section elaborates on the proposed model which has the following main objectives:

- Improve early stages of ERP Systems implementation to decrease the chance of implementation failure; the objective is to analyse, control and improve the process model defined when implementing the ERP system, using process conformance techniques to identify necessary changes and adjustments in the architecture and design of the model that increase the fitness of the model to the observed behaviour, giving the ERP system more flexibility to adapt to the challenges and needs of each company.
- Improve usability of the ERP system from the user perspective; by discovering the correct model that best describes the normal behaviour of the users when interacting with the ERP system, compare how a certain user is interacting with the ERP system and provide in real time useful tips in case the ERP user is deviating to much from the expected behaviour, improving time consumption and solving common questions and bottlenecks easier and more efficiently.
- Constantly improve the ERP architecture and ERP usability from the ERP manager perspective; by constantly analyze, monitor and control how users interact with the ERP system, providing constant insights to the ERP manager, so he can act and implement changes in the ERP design as to make it more efficient and closer to the observed behavior as well as a more accurate process model. It has also the objective of providing the ERP manager with insights that help improving the user experience, such as automation of certain tasks or decreasing common bottlenecks.

Model Description:

The model is composed by the ERP in the middle with its different modules that generate the log data, the Log Data Warehouse which collects and stores the log data generated by the different modules, the Process Mining Architecture where process mining techniques are applied to the stored log data and generate outputs, the users that utilize the ERP and the ERP manager that can implement any changes to the ERP itself.

In figure 4 it is possible to see the proposed model, and although it is circular as it is supposed to be a continuous process of improving the ERP, the model starts in the ERP modules and the log data they generate. This log data produced by the user usage, is then collected, and stored in the Log Data Warehouse for further analysis. The usage of this log data has two distinct objectives, first objective is to improve the ERP through inputs given to the ERP manager that can then apply changes and improvements to the ERP and a second objective of improving the user usability of the model.

Between the Log staging area and these two objectives and respective participants, the ERP manager and the ERP Users, there is the Process Mining Architecture which applies the different process mining techniques to the log data and allows for insights and conclusions.

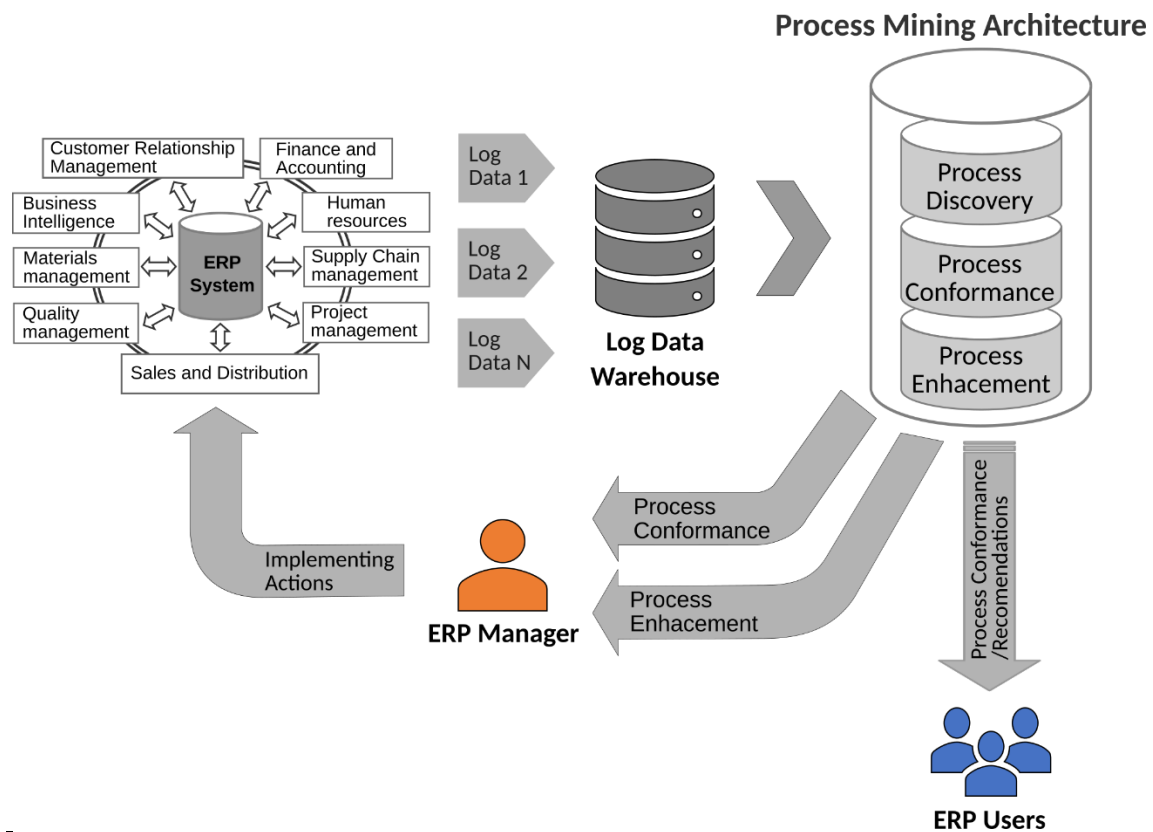


Figure 4: Proposed Model Outline. Source: Own Illustration.

Process Mining Architecture is the area of the model where the different process mining techniques are applied and can be seen in further detail on figure 5. It receives log data from the log data warehouse and performs different process mining techniques in a defined order that end with outputs given to the users and to the ERP manager.

The architecture proceeds to perform data preparation, the model uses the TIL framework proposed by (Lucas & Babaian, 2012) to separate the log data referring to the process model composed by the different tasks and process instances, and the user-system interaction, accounting for two separate types of log data and two separate approaches when applying process mining techniques. Log data, referring to the activities performed by one or more users and the business process is extracted from different tables related to the processes, which have been previously defined as well as the activities themselves, this type of log data ignores the user-system interaction. The other type of log data extracted, accounts for the user-system interactions. The different components that integrate the interface pages of the ERP system are previously mapped and the log data of their utilization is extracted from the respective tables.

After data preparation, the architecture then applies process discovery techniques to the log data, this technique is a process that runs parallel to the model and ERP utilization as a circular process that is constantly improving the described process model. It aims to capture as accurately as possible the real behaviour and keep the described process model as updated as possible, it is the base line for any further improvements, changes and process mining techniques applied.

Process Conformance techniques are then applied to the discovered model to compare how the ERP is being used, with how it was expected to be used, allowing to quantify and qualify any kind of deviations from the desired model, and is a constant process as the discovery process changes as well. Process Conformance is performed from two optics, one to improve the ERP usability from the user perspective; for this it creates a list of common deviations and their root causes and, as the user interacts with the ERP it performs real time process conformance. In case it identifies deviations, the architecture purposes real time recommendations and tips to the user as how to improve ERP usage, these recommendations are continuously improving. On another optic, it gives insights to the ERP manager identifying deviations from the expected behaviour or bottlenecks, this is combined with Process Enhancement techniques that are also applied, to correct and extend the process model and therefore improve how the EPR is being used, allowing the ERP manager to implement actions and changes.

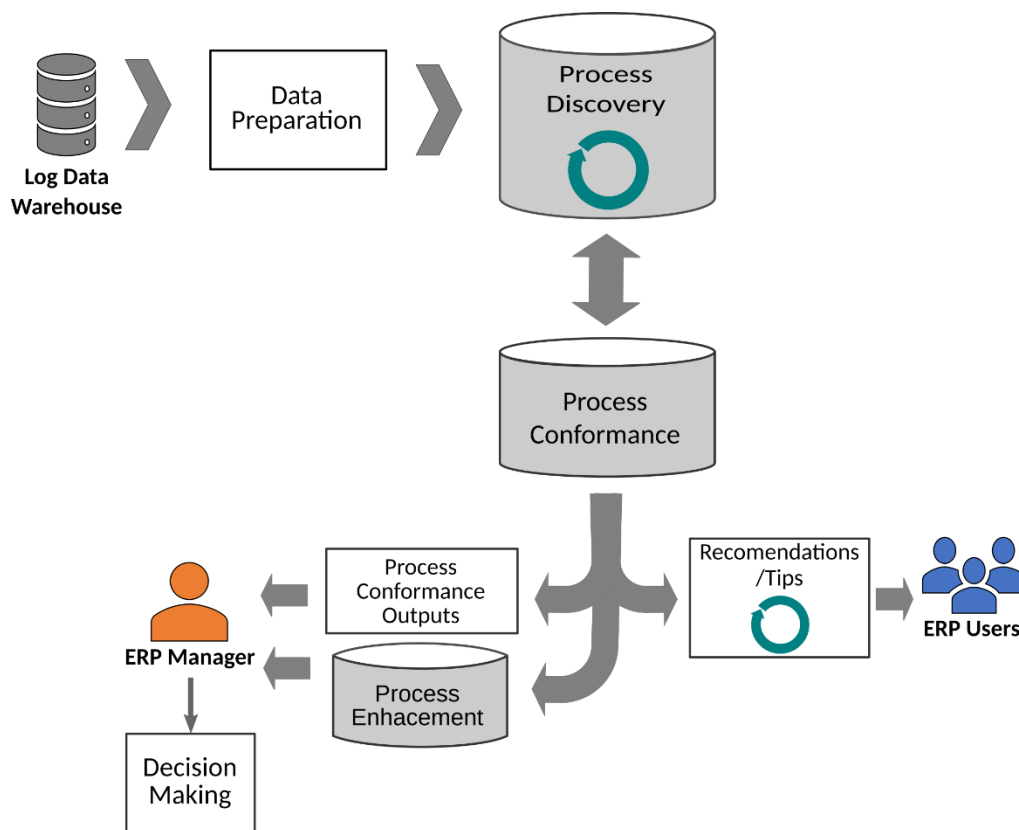


Figure 5: Detail of the Process Mining Architecture of the proposed model. Source: Own Illustration.

The model gives different importance levels to the different paths according to the maturity phase in which the ERP system is. In an early stage of the ERP implementation, the model focus mainly in process discovery and process conformance, it aims at finding critical breaking points that could compromise the ERP implementation and improving its flexibility to adapt to particular cases and situations, giving inputs to the ERP manager so he can act is the number one priority, as user usability can be improved in later stages. After the early stages of ERP implementation, and once it is well established, the normal workflow and process instances will be more normalized and less scattered,

allowing to better identify and quantify deviations from the user perspective, in this stage the model aims at improving the ERP system as a whole and in steady and consistent way, focusing in both process conformance and enhancement providing inputs both to the ERP manager and the ERP users combining improvements to the ERP system itself and to the user usability.

Model Components:

The architecture of the model is constituted by different components that act together and are related between them;

- The ERP system, with its different models, assuming that the most common ones are Finance, Human Resources, Supply Chain Management, Supplier Relationship Management, Customer Relationship Management, Business Intelligence, Sales and Distribution, Project management, Quality Management. And each of these modules produce their own log data event. The usage of the ERP by the different users is what constitutes our model when applying process mining techniques and therefore the main goal is to improve its usability.
- Log Data produced by the different modules, which refers to an activity or well-defined step in a process and stores the information about this event (W. Van Der Aalst, 2012a), in this case it's the outputs of the user interactions with the ERP and it's different modules, they are the starting point for the process mining analysis.
- Log Data Warehouse is where all the log event data is stored, it receives the log data generated by the ERP utilization and is an intermediate step to store and organize data until its ready to be mined.
- Process Mining Architecture, is the area where process mining techniques are applied to the log data stored in the Warehouse, it is constituted by the following components:
 - Data Preparation, where the log data is prepared to be able to be mined and processed by the algorithms.
 - Process Discovery, it is a constant process that is constantly running and improving the described process model, its preformed on the log data generated by the ERP system and provides a described model of the real behavior.
 - Process Conformance, it compares de observed behaviour with the one described in process discovery, it is a back-and-forth process as this comparison is constant. It provides outputs to the ERP manager by quantifying and qualifying deviations, it also performs real time recommendations to the users.
 - Recommendations, is a list of identified deviations and their root causes, when process conformance identifies that the user is deviating from the model and it finds a correspondence to this list, it gives recommendations and tips in real time on how to improve the ERP utilization. Recommendations are continuously improving as more deviations are identified and more log data is processed.
 - Process Enhancement, has the objective of extending and improving the process model, by identifying bottlenecks and main points to improve, it analyses resources, groups of people executing related activities and identifies relationships between workload and time spent to perform certain activities. It provides the ERP manager

with outputs on how to improve the model, as well as to better understand the business itself and how the different actors interact between them.

In example, when analyzing timestamps, by computing simple statistics it is possible to identify the main bottlenecks and points to improve. It is also possible to analyze resources, groups of people executing related activities and construct social networks that allow to identify the relation between workload and time spent to perform certain activities. Additionally standard classification techniques can identify the main decision points in the process model and help explain the different outcomes (W. Van Der Aalst, 2012b).

- Users, representing everyone that interacts with the ERP from a user perspective and therefore outputs log events data and process instances. Users not only are the actors that output the logs to be mined and analyzed in the scope of process mining, but they are also in the receiving end of the conclusions generated by applying process mining techniques.

In our architecture, process discovery is constantly being applied to understand how users are using the ERP and produce a model that accurately describes the real behavior of the users. In parallel, process conformance techniques are applied in real time to improve user experience and usability, the objective is to compare the normal use of the ERP found in process discovery and help the user with relevant tips when his usage of the ERP system deviates from the expected behavior.

- ERP Manager represents everyone that oversees the ERP systems or have any influence in its structure or other decisions related to the ERP system. The ERP manager receives the outputs generated by mining the log data and is responsible to act and implement the needed changes and improvements in the ERP system. Process discovery techniques are also constantly being applied, enabling the ERP manager to always have a deep understanding of how the ERP system is being used and its many variants, producing a model that accurately describes the user's behavior when using the ERP model. Process Conformance and Process Enhancement techniques are running in parallel.

Process Conformance with the objective of identifying and quantify the deviations from the described model found in the discovery process as well as comparing the real behavior of the users with the expected behavior when the ERP was designed, in our architecture has the main purpose of identifying the process segments where most deviation occur and determine what they have in common, it is the starting point for process enhancement.

Process Enhancement conclusions are meant to be used by the ERP manager to allow him to act and implement improvements in the ERP system, enhancements in the ERP systems aim to make it as efficient as possible, reducing time spent by the users in daily and repetitive tasks and make it as user friendly as possible, therefore the main objectives of our architecture when applying process enhancement techniques are:

- Analyse timestamps and identify main bottlenecks.
- Identify the process segments where most deviations occur and implement solutions to improve the ERP usability.

- Identify the process segments where it makes sense to implement RPA's and other i-ERP functionalities to automate process segments avoiding human mistake and decreasing the workload of the ERP users.
- Identify the user workflow and analyse the differences in the ERP usage with the expected behavior of the users, improving design, user interface and usability in general, i.e., page layout, buttons functionality and different ways to reach and use the different modules.

4.3. USE CASE

For this section, a hypothetical ERP, the ERPX that implements all the standard ERP functionalities and characteristics was evoked to demonstrate one of the possible use cases for the previously proposed model.

In particular, using a purchasing process integrated in the module of material management, which contains the following activities:

- Purchase Requisition,
- Approval of Purchase Requisition,
- Placement of purchase order,
- Contract Agreement,
- Without Agreement,
- Invoicing from a supplier,
- Payment.

The proposed model uses the TIL framework proposed by (Lucas & Babaian, 2012) to distinguish between the process model composed by the different tasks and process instances, and the user-system interaction, consisting in two separate types of log data and two separate approaches when applying process mining techniques.

To analyze the activities performed by one or more users, where the objective is to understand the business process while ignoring the user-system interaction, the process starts with the extraction of the log events from the tables related to the process in analyses, which have been previously defined as well as the activities themselves which in this case constitute the purchase process.

As the users keep using the ERP and creating purchase orders, the log data generated by the performance of the activities related to the process, is extracted from the different tables, and imported into the data warehouse, from here process discovery techniques are continuously applied to discover how users are completing their purchase process and translate it into a process model using BPMN that reflects the real behavior of the users. Process Discovery is constant and the process model it generates keeps getting more complete as the user interact more with the ERP, generating a more log data and covering more instances of the process.

Process Conformance techniques like alignment-based conformance checking are then applied to compare the process model discovered with the one describing the behaviour expected from the users when completing a purchasing process shown in figure 6. Process Conformance is also a constant

process, that keeps comparing the real behaviour of the users with the one expected when the ERP was implemented.

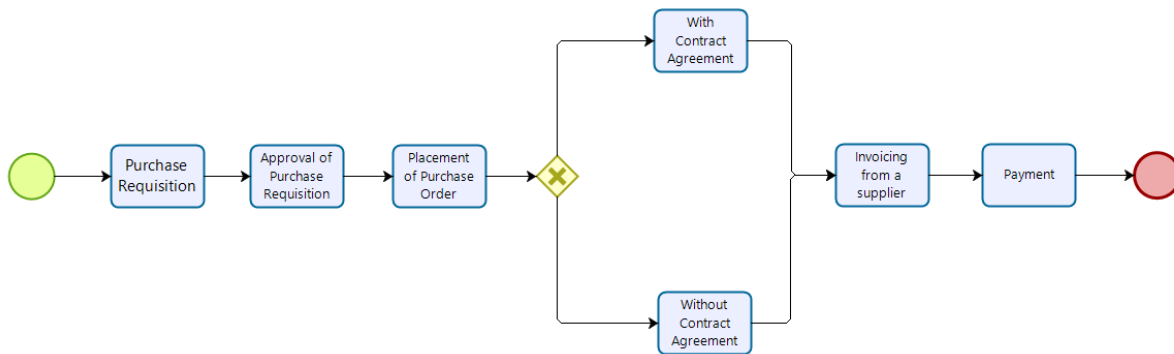


Figure 6: Expected Process Model of a Purchasing Process. Source: Own Illustration.

In this case, process Conformance found out that there was an unnecessary loop, where a big number of users would complete the activity “Purchase requisition” and then in the next activity “*Approval of Purchase Requisition*”, the purchase requisition wouldn’t be approved, making the user have to repeat the first activity, originating a loop.

This outputs from Process Conformance are analysed by the ERP manager, and together with Process Enhancement techniques which analyses further attributes of the Event Log, the ERP manager realizes that most of the Purchase requisitions not approved are due to mandatory fields that are not filled in. With this information in mind, the ERP manager implements changes into the ERP, that after implementation don’t allow the “Purchase requisition” to be placed unless all the mandatory fields are filled in, ending with most of the loops generated by the rejection of the purchase requisition.

On another perspective, focusing on the user-system interaction with the objective of improving user usability of the ERP and according with the TIL framework proposed by (Lucas & Babaian, 2012) the different components that integrate the interface pages of the system are mapped and log data of their utilization is extracted.

Process discovery is constantly applied to identify how the users are interacting with the system interface, and process conformance performed in real time allow to understand when the user is deviating from the normal user-system interaction expected pattern, when a deviance is identified and it has a known root cause, useful tips are suggested to the user in order to improve the usability. It could also tip the ERP manager about main bottlenecks that could indicate a need to change the interface design.

In this case, it was noticed that one of the main deviations were users that clicked on the submission button, before clicking in all the mandatory fields, which after ERP manager alterations to the ERP, didn’t allow the user to precede to the next activity. With this a new recommendation was added to the recommendations and tips list, whenever this deviation is identified, a pop up appears to the user which informs him that all the mandatory fields need to be filled in before moving to the next activity, the field that haven’t been filled in yet are also highlighted.

4.4. BENEFITS

In this section, the benefits of the proposed model will be discussed and presented. The proposed model had the objective of solving common problems found in the ERP systems and enhance their capabilities by making use of the most recent technologies.

On the other hand, Process mining is a relatively recent field of research that aims at improving processes and provide enterprises with insights and knowledge about their processes (W. Van Der Aalst et al., 2011). Therefore, this model has also the objective of demonstrating the possible benefits of applying this type of techniques and technology to ERP systems in order to improve them and increase their added value to the organizations.

The main benefits of the proposed model to organizations that adopt it are:

- Improved flexibility when implementing a new ERP system in the organization or facilitate the adaptation phase to a new ERP or changes to an existing one. The model allows for a better understanding of the processes that characterize the ERP and therefore allow for more accurate and faster adjustments to its components when required, allowing for a smother implementation process that is able to adjust the ERP system to the real behaviour of the users and needs of the organization.
- Improved personalization of the ERP modules and adaptation to each organizations needs and requirements, by applying process mining techniques such as process discovery and conformance checking, the model allows for a bigger personalization of the ERP system that can adapt to the processes of the company and to the users themselves taking into consideration the real behaviour of the organization.
- Improved user usability of the ERP system and its interface, giving personalized assistance and guidance to users as they interact with the ERP interface, improving their user experience, time consumption and dealing with common bottlenecks easier and more efficiently. The model is user friendly and reduces the learning curve of the users.
- Constant improvement and maintenance of the ERP model, not only in a first stage when the model is firstly implemented but also during its life cycle and constantly over time, allowing the ERP to adapt to the different challenge and needs the organization might feel in the future. The model, allows for a continuous analysis of the processes and provides the ERP Manager with insights when necessary, allowing him to understand in real time and adapt to the users and the organization changes and maintain a healthy and useful ERP.

In general, the model provides a solution to take advantage of the huge amount of log data information systems like ERPs produce, reducing the gap between how the systems are planned to be used with how the organization carry out the operations. Providing the organization with a useful tool to increase the value added by the ERP system.

4.5. EVALUATION & DISCUSSION

In this chapter an evaluation is done with the purpose of validating the proposed model and understand how efficiently it responds to the research question. The result of this validation will determine if the research should precede to the next steps or if it should focus on improving the proposed model.

Since there wasn't an implementation of the proposed model in a real ERP system, and the use case is hypothetical, it is not possible to measure evaluation criteria's such as user satisfaction when applying the model or other quantitative benefits of implementing the model in a real-life scenario. Therefore, the validation process of the proposed model was done recurring to interviews to experts in the field, with the main objective of understanding the feasibility and general appliance of the model to a broad number of organizations as well as measuring qualitatively the benefits of adopting it.

Regarding the interviews themselves, an interview guide was developed with the purpose of providing structure to the interview and help evaluate the knowledge from the experts. The questions were formulated in order to provide maximum knowledge gain from the interviewed and approach different angles of the proposed model. It was also taken in consideration the importance of the interviewers impartiality and openness to criticism.

The interviews were individually conducted during the month of November of 2023 and approached experts from the academic field with specializations in process mining and ERPs, all of them accepted to be recorded and have their interviews transcript, the interviews were conducted in Portuguese and can be consulted in the annex section.

Expert #	Fiel of Expertise	Field of Employent	Specialization
1	Academic	Assistant Professor at Nova IMS	Process Mining
2	Academic	Assistant Professor at Nova IMS, PhD in Information Systems	ERPs

Tabel 1: Expertise Interviews.

Both interviewees believed the model would be useful if implemented, underlying that it would be especially benefic to improve productivity, decrease human error, would help enhance the ERP systems allowing to make better use of the full potential of the ERP and help in the implementation phase. Although one of the interviewees expressed some concerns regarding the feasibility of the part of the model directed to the ERP user where suggestions and process mining techniques would have to be performed in real time, stating that it doesn't have doubts it works in a predictive manner, but not so sure if it can be done in real time. The interviewees also gave important suggestions on how to improve the model and make it more robust.

Interviewee 1 stated that the model is well thought and would be very interesting to validate it further with a real implementation, as it would be interesting to see how it deals with real life situations. The suggestions on improvement defended are to be clearer and more standardized on how the outputs

from process conformance are passed to the ERP manager, as different managers may have different interpretations of the results.

Interviewee 2 also found that the model would be useful, especially from the point of view of ERP implementation and extracting the full potential from it and would allow the ERP manager to optimize the system, he also underlines the usability in reducing human errors and increasing productivity in general. The biggest criticism regarding the model according to the interviewee 2, is related to the feasibility of implementing real time process mining techniques such as process discovery and conformance and recommendations to the users, stating that it would be more feasible to apply this processes predictively and not in real time.

After conducting the interviews, it can be extracted that the feeling from the interviewees is positive regarding the usability of the model and implementation, although some concerns regarding the user recommendations ramification of the model. The criticism is limited to the previously mentioned ramification and recommendations are incremental to the model and not structural.

5. CONCLUSIONS

After doing a literature review, it can be concluded that ERP Systems are complex Information Systems widely used by many organizations. However, they come with many challenges (Scheer & Habermann, 2000; Shehab et al., 2004; Yen et al., 2002) that organizations struggle to deal with so they can optimize all the advantages and functionalities ERPs provide. They also perform many business activities and processes, generating data which can be used to extract further business knowledge. Likewise, the literature review allowed to conclude that process mining is a more recent field (W. Van Der Aalst et al., 2011) with great potential to explore business processes based on data rather than opinions.

From what was discovered in the literature review, a model architecture that incorporates both concepts, more specifically using process mining to solve some of the ERP challenges, is possible and a viable solution.

Therefore, this research has produced a model that incorporates process mining into ERP systems to improve their usability, this model has been demonstrated through a use case and validated recurring to interviews to experts.

5.1. SYNTHESIS OF THE DEVELOPED WORK

This research was done taking a structured approach that took well defined steps and guidelines to achieve a coherent work with trustworthy results. Initially, the motivation and objectives of this study were outlined and culminated in the generation of the research question that served as a guide for the rest of the research and which this study aims to answer:

RQ: Can we use Process Mining techniques to improve ERP usability?

To achieve the goal of answering the research question, a systematic literature review was carried out on the most relevant articles, firstly on ERP systems with the goal of understanding the different components and characteristics of the ERPs and their capacities. Secondly on process mining techniques, most common algorithms, concepts, best practices and most common drawbacks. Finally, and based on the initial findings, a literature review on the approaches and the necessary requirements to integrate process mining with ERP systems was also conducted.

After this, a Design Science Research methodology was adopted to create an artifact capable of giving response to the research question, which resulted in a proposed model that integrates process mining techniques into ERP systems. The purpose and possible uses of the proposed model was demonstrated recurring to a hypothetical ERP that adopts all the standard ERP functionalities and characteristics, and that recreates a possible use case for the model. The final validation of the model was done qualitatively recurring to interviews to experts. A group of experts was chosen and individually interviewed in order to collect their feedback on the usability, quality and recommendations to the model.

Therefore, the objective of creating an artifact that can improve the usability of ERP systems recurring to the integration of process mining techniques was achieved, answering to the research question initially conceived.

5.2. LIMITATIONS

One of the limitations that this research faced was a lack of studies and concrete applications of how to correlate process mining with ERP systems. Although there is a lot of studies and deep research on both ERP systems and process mining separately, there are few guidelines or potential methodology frameworks that can be applied to implement process mining in ERPs and what are the requirements and constraints of correlating these two fields of research.

Another limitation that this study faced, was the complexity of ERP systems and the variety of vendors that provide different types of ERP systems, all with particular characteristics and variations which makes it especially difficult to make generalizations and take general conclusions and applications. The different ERP systems that exist, and the fact that most ERPs are customized to the needs of the organization using it, with different types and number of modules which generate different types of data makes it increasingly difficult to take general conclusions and propose general frameworks that applies to all these types of existing ERPs.

This study focuses mainly on integrating process mining techniques into ERP systems, however once this integration is successful, it allows for more possibilities as to what do with the outputs given by the process mining which haven't been further explored, such as integrating machine learning and RPA's (Robotic Process Automations).

Finally, one of the main limitations of this research, is that due to the complexity and difficulty of implementing the proposed model in a real ERP used by an organization in their daily operations, the Use Case is hypothetical, and the proposed model hasn't been tested in a real-life situation in a real ERP. Therefore, although it has been demonstrated how the proposed model can be used to improve the ERPs usability using process mining techniques, the main problems and limitations that the model may face when dealing with real life scenarios haven't been tested and studied yet.

5.3. FUTURE WORK

Therefore, related to the limitations, some of the future work proposed in this research, would be to:

- Implement the proposed model into a real ERP system utilized by real organizations in their daily operations and evaluate its performance.
- Implement the proposed model into ERP systems from different vendors, with different modules and with different organizational needs. In order to evaluate the model performance across different ERP systems and how it adapts to the different ERP vendors and special characteristics of each one.

- The proposed model uses process mining techniques such as process discovery and process conformance to identify when users are deviating from the expected behaviour and might be having problems with the ERP usage and then proceeds to give tips and recommendations to the user. This functionality could be extended in order to integrate other technologies such as machine learning and process automation that would increase the value added to the user with more complex and personalized recommendations.
- Further research on how to use the outputs given by the model, such as integrating machine learning and RPAs into the ERP systems and research on how the model can also be implemented into Intelligent ERPs (I-ERPs).

REFERENCES

- Ahmad, M. M., & Cuenca, R. P. (2013). Critical success factors for ERP implementation in SMEs. *Robotics and computer-integrated manufacturing*, 29(3), 104-111.
- Amalnick, M. S., Ansarinejad, A., Nargesi, S.-M., & Taheri, S. (2011). New perspective to ERP critical success factors: Priorities and causal relations under fuzzy environment. *The Journal of Mathematics and Computer Science*, 2(1), 160-170.
- Arnetty, N., Musiega, D., & Juma, S. Implementation of Enterprise Resource Planning Systems in Kenyan Public Universities, A Case of Masinde Muliro University of Science and Technology.
- Bahssas, D. M., AlBar, A. M., & Hoque, M. R. (2015). Enterprise resource planning (ERP) systems: design, trends and deployment. *The International Technology Management Review*, 5(2), 72-81.
- Beheshti, H. M. (2006). What managers should know about ERP/ERP II. *Management Research News*, 29(4), 184-193.
- Chapman, C. S., & Kihn, L.-A. (2009). Information system integration, enabling control and performance. *Accounting, organizations and society*, 34(2), 151-169.
- Chen, I. J. (2001). Planning for ERP systems: analysis and future trend. *Business process management journal*.
- Chen, R.-S., & Tsai, Y.-S. (2005). The Application of Data Mining Technology for Intelligent Enterprise Resource Planning System.
- Davis, G. B., & Olson, M. H. (1984). *Management information systems: Conceptual foundations, structure, and development*: McGraw-Hill, Inc.
- Der Aalst, V., & Mining, W. P. (2011). Discovery, Conformance and Enhancement of Business Processes. In: Springer: Berlin, Germany Heidelberg, Germany.
- Fleig, C., Augenstein, D., & Maedche, A. (2018). Designing a process mining-enabled decision support system for business process standardization in ERP implementation projects. Paper presented at the International Conference on Business Process Management.
- Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., & Shan, M.-C. (2004). Business process intelligence. *Computers in industry*, 53(3), 321-343.
- Gulledge, T. R., Sommer, R. A., & Vincent Jr, J. P. (2005). An introduction to basic enterprise resource planning concepts. *International Journal of Management and Enterprise Development*, 2(2), 204-218.
- Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In *Design research in information systems* (pp. 9-22): Springer.
- Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science research in information systems. *MIS quarterly*, 28(1), 75-105.

- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.
- Järvinen, P. (2007). Action research is similar to design science. *Quality & Quantity*, 41(1), 37-54.
- Jenab, K., Staub, S., Moslehpour, S., & Wu, C. (2019). Company performance improvement by quality based intelligent-ERP.
- Kahraman, C., Kaya, I., & Çevikcan, E. (2011). Intelligence decision systems in enterprise information management. *Journal of Enterprise Information Management*.
- Klaus, H., Rosemann, M., & Gable, G. G. (2000). What is ERP? *Information systems frontiers*, 2(2), 141-162.
- Lucas, W., & Babaian, T. (2012). Implementing design principles for collaborative ERP systems. Paper presented at the International Conference on Design Science Research in Information Systems.
- Mahendrawathi, E., Zayin, S. O., & Pamungkas, F. J. (2017). ERP post implementation review with process mining: A case of procurement process. *Procedia Computer Science*, 124, 216-223.
- Marnewick, C., & Labuschagne, L. (2005). A conceptual model for enterprise resource planning (ERP). *Information management & computer security*.
- Parr, A. N., & Shanks, G. (2000). A taxonomy of ERP implementation approaches. Paper presented at the Proceedings of the 33rd Annual Hawaii International Conference on System Sciences.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.
- Reubens, R. (2016). To craft, by design, for sustainability: Towards holistic sustainability design for developing-country enterprises.
- Rizza, M. (2016). The Rise of Intelligent ERP (i-ERP). *IDC Community*, 19.
- Scheer, A.-W., & Habermann, F. (2000). Enterprise resource planning: making ERP a success. *Communications of the ACM*, 43(4), 57-61.
- Shehab, E., Sharp, M., Supramaniam, L., & Spedding, T. A. (2004). Enterprise resource planning: An integrative review. *Business process management journal*.
- Van Der Aalst, W. (2012a). Process mining. *Communications of the ACM*, 55(8), 76-83.
- Van Der Aalst, W. (2012b). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2), 1-17.
- Van Der Aalst, W. (2016). *Process mining: data science in action (Vol. 2)*: Springer.
- Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., . . . Buijs, J. (2011). Process mining manifesto. Paper presented at the International conference on business process management.

- Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., . . . Buijs, J. (2012). Process mining manifesto. Paper presented at the Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9.
- van der Aalst, W. M. (2010). Process discovery: Capturing the invisible. *IEEE Computational Intelligence Magazine*, 5(1), 28-41.
- Van Der Aalst, W. M., Reijers, H. A., Weijters, A. J., van Dongen, B. F., De Medeiros, A. A., Song, M., & Verbeek, H. (2007). Business process mining: An industrial application. *Information systems*, 32(5), 713-732.
- Van der Aalst, W. M., & van Dongen, B. F. (2002). Discovering workflow performance models from timed logs. Paper presented at the International Conference on Engineering and Employment of Cooperative Information Systems.
- Van der Aalst, W. M., & Weijters, A. J. (2004). Process mining: a research agenda. *Computers in industry*, 53(3), 231-244.
- van Giessel, M. (2004). Process mining in SAP R/3: A method for applying process mining to SAP R/3. Eindhoven University of Technology, Eindhoven, The Netherlands.
- Yen, D. C., Chou, D. C., & Chang, J. (2002). A synergic analysis for Web-based enterprise resources planning systems. *Computer Standards & Interfaces*, 24(4), 337-346.

ANNEXES (OPTIONAL)

Expert Interviews:

Interview 1: Expert 1, 14-11-2023

Q1: Do you consider the proposed to be useful and why? If not, why do you believe it is not?

Olhe, acho que é um modelo útil e que seria interessante de ser testado, essencialmente. Porque com um log, um layout específico para se poder testar, acaba por se calhar o resultado validar de uma forma mais efetiva o que está a propor. Penso que a arquitetura que aqui está está bem pensada, talvez quando sai do Process Conformance para os outputs, eventualmente linkar aqui um ficheiro, de um conjunto de tópicos. Na perspectiva do manager e a nível da usabilidade. Um ficheiro para cada lado, para se perceber qual é que é o resultado, o que é que realmente tira daqui. É uma forma de suportar melhor, por exemplo em UML, representamos notas que associamos a objetos para eventualmente haver aqui alguma contextualização escrita que não tem um valor sintático, mas que ajuda um developer, software architect, etc., a entender, o que ficou pouco entendido, tanto no desenho como na modelação. Porque cada modelador tem uma perspectiva muito pessoal e a imagem pode, às vezes não dar a visão, ou então a orientação, ou passar a mensagem que é correta. O UML, realmente, como é um standard, por isso é que os standards são importantes, quando eu ponho um determinado objeto, aquilo significa aquilo, não há um segundo significado.

Q2: Do you have any criticism towards the proposed Model? Please explain.

Relativamente ao criticismo, acho que foi mais ou menos também explicado.

Q3: Would you consider implementing the proposed Model? Please clarify why/ why not.

Tal como já tinha dito é interessante experimentar. Implementar. Ou seja, eu acho que os modelos são todos credíveis do ponto de vista visual e teórico. Não há propriamente... Epá, isto está mal! Quer dizer, podes cometer erros, não é? O que é que é interessante é pegar naquilo, olhar, dar um feedback ou ter uma percepção do que é que realmente lá está. Os Data Marts pode ser um conceito que pode aqui explorar, pode ser útil.

Q4: Do you have any recommendation or suggestions for further improvements of the proposed Model?

Os Data Marts pode ser um conceito que pode aqui explorar, pode ser útil. Por exemplo, em um módulo de custo, em um financeiro, ou um departamento de recursos humanos, ou o que que seja, o Data Mart pode ser só a extração da informação que é útil ao financeiro.

Tem um modelo de dados para a financeira, tem um modelo de dados para os recursos humanos. Depois há um modelo de dados que suporta isto tudo, portanto, os Data Martes, estamos a tirar pequeninhas partes só para estudar aquilo. Porque o que é que é Data Mart? É um bocado da informação do Data Warehouse que vai trabalhar e do qual depois extrai a informação que precisa.

Interview 2: Expert 2, 21-11-2023

Q1: Do you consider the proposed to be useful and why? If not, why do you believe it is not?

Sim, ser útil é. Será útil, se permitir por exemplo, melhorar o processo, a forma como o processo está implementado no ERP, por um lado, e isso é uma função do ERP Manager à partida, ou seja, aquilo que eu estou a perceber é que isto vai dar inputs ao ERP Manager, ao gestor do sistema, para que ele possa fazer alterações no ERP, de forma a que o processo seja otimizado. Exato. Assim, sem dúvida é útil. Como é que o faz, já é outra questão. Se é ele que o faz, ou se tem que recorrer a quem programou e implementou o ERP, mas isso já é outra questão. Do ponto de vista dos utilizadores, se eu considero que é útil, sim, se esta arquitetura, permitir dar sugestões aos utilizadores, para eles, para os ajudar, por exemplo, a preencher os ecrãs, não é? Quando estão a fazer determinadas transações, claro que sim, que é útil. Vejo que pode ajudar em termos de produtividade, redução de erros. Isto do ponto de vista do operador, do ponto de vista do gestor do sistema, pode ajudar a melhorar o processo de maneira a que ele esteja mais automatizado, ou que tenha um processo que demore menos tempo a executar, que tire mais potencial, consiga ir extrair mais das potencialidades do ERP. Porque, às vezes, a implementação não está a tirar todo o potencial do ERP. Porque na implementação uma coisa típica é que vamos implementar isto da maneira como nós temos feito as coisas e não setira todo o potencial. Por exemplo pode ter lá algumas funcionalidades de robot process, mas que não estamos a usar, porque a implementação foi feita como faziam as coisas antes. Esta tecnologia vai permitir uma nova versão do ERP, se calhar vai permitir algumas coisas que a anterior não permitia.

Q2: Do you have any criticism towards the proposed Model? Please explain

Não é bem criticismo. A minha questão é se, será que as sugestões para os utilizadores não vão para além do Process Mining e possam ser implementadas com outros tipos de tecnologia? Por exemplo com bots, com AI, que lhe vai dar sugestões de como é que ele deve preencher. Isso é uma das questões, que eu não vou chamar de criticismo, mas ponho em dúvida se será com esta arquitectura a melhor forma de o fazer. Se for de uma forma preditiva e que depois permita programar esses mecanismos de automação, ok. Agora, a minha dúvida é se hoje, só com esta arquitetura, se pode dar em tempo real informação aos utilizadores, não é? É só isso.

Q3: Would you consider to implement the proposed Model? Please clarify why/ why not.

Pois, aqui tenho dúvidas, porque as dúvidas prendem-se com a questão anterior, não é? Será que este modelo proposto, ou seja, se conseguiria implementar? Não sei. Não sei porque, não sei se existem outras formas de fazer alguma coisa. Tem mais a ver com isso. Portanto, sem saber isso, eu não consigo responder esta questão. Se consideraria implementar, eu teria que explorar mais.

Q4: Do you have any recommendation or suggestions for further improvements of the proposed Model?

Neste momento não, porque isto deriva também da anterior, se eu não consigo dizer se iria implementar. Também não posso dizer se tenho sugestões ou recomendações. Ou seja, tenho dúvidas, mas isso é do meu conhecimento, atenção. Tenho dúvidas se não há outras formas de, por exemplo, providenciar algumas das funcionalidades. Sobre a questão de alimentar o gestor do ERP com sugestões, isso daí não me causa tanta dúvida. Acho que sim. E hoje em dia há soluções tais como Celonis e no Signavio e outras que fazem essa análise dos processos do ERP que foram executados, e

vêm dar indicações de erros que acontecem. E dar, portanto, sugestões para alterar esses processos e a maneira de implementar esses processos no ERP. Portanto, aí sim, acho que sim, acho que isso pode ajudar sem dúvida. As minhas dúvidas são mais do ponto de vista de dar informação em tempo real ou dar sugestões em tempo real aos utilizadores. Daí é que tenho mais dúvidas.



NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa