NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação

Master Program in Information Management

Use of Audit Data to Improve the Supply Chain Performance

Beatriz Maria Azevedo Fernandes

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management

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 AUDIT DATA TO IMPROVE THE SUPPLY CHAIN PERFORMANCE

By Beatriz Maria Azevedo Fernandes

Dissertation presented as a partial requirement of the degree of Master of Information Management, with a specialization in Business Intelligence and Knowledge Management

Supervisor: Prof. Vitor Santos, PhD.

July 2023

ABSTRACT

In the last decades, globalization and digitalization were two of the main reasons for the increase of complexity in supply chains, altering the industries due to the massive amount of information available. This complexity started to become harmful for the companies that do not understand how to use data and information as their competitive advantage, increasing the risk and costs associated with their processes, and decreasing effectiveness and efficiency. We look for the concept and area of internal auditing and process mining techniques as a solution to revert this situation. While research has focused on different and mostly narrow aspects in these areas and solution-oriented and more practical approaches can be found and applied to a broader environment, a practical solution that incorporates these areas into the supply chain are hard to find. Therefore, following a design science research methodology, this study proposes an iterative framework that consists of a guide for an organization that wants to incorporate new technologies into their processes in the supply chain while making the best out of the massive amount of information available using internal auditing and focus on process mining techniques. The framework provides a chain of steps that can be adapted by the company during the transformational process, guaranteeing a smooth transition away from the traditional systems to a more modern and flexible architecture.

KEYWORDS

Design Science Research; Process Mining; Supply Chain; Complexity: Audit Data, Maturity Model

INDEX

INDEX OF FIGURES

INDEX OF TABLES

List of Abbreviations and Acronyms

ADA Audit Data Analytics

AI Artificial Intelligence

BDA Big Data Analytics

BPM Business Process Management

BPMN Business Process Model and Notation

BS Behavioral Science

CMM Capacity Maturity Management

DSR Design Science Research

ERP Enterprise Resource Planning

IBM International Business Machines

IoT Internet of Things

IS Information Systems

IT Information Technology

KPI Key Performance Indicators

LT Literature Review

ML Machine Learning

MM Maturity Model

SCM Supply Chain Management

SCOR Supply Chain Operations Reference

VTA Value Through Analytics

1 Introduction

1.1 Context

In the last decades, globalization together with mass customization and other variants, altered the manufacturing industry all over the world, making organizations the target of technical, organizational, and operational changes. This phenomenon happened due to the massive information and communication technologies available that make processes and companies more complex (Beinhocker, 2007).

Hand in hand with this, we can see the growing usage of information systems as an answer to the need to improve the efficiency and performance of business processes. This happens once a single functionality approach is no longer enough, and we are starting to see a Business Process Management (BPM) approach, meaning that they require the interaction of several people and functions throughout the organization instead of dealing with single functionalities (Lamghari, 2022).

This leads to a new side effect of supply chains: high complexity, justifiable by the complexity of processes, products, infrastructures, and required usage of Information Technology (IT) and data. This complexity will affect the supply chain performance, which is a key competitive strategy element that increases the productivity of a company and the profitability of a business (Gunasekaran, 2004). By adding unnecessary complexity to the supply chain, and reducing its performance, therefore, what was a competitive advantage transforms into a liability (van der Aalst W. , 2011).

It is also revealed in different studies the potential of managing the complexity of a business, and consequently, the need to integrate this management into the supply chain management. In manufacturing companies, 25% of the total expenditure is on account of the products and processes' complexity over the supply chain (Aelker, Bauernhansl, & Ehm, 2013). Another study, states if companies manage their complexity regarding the processes and other variables, that they can increase the EBIT up by to 5% (Scheiter & Gerjets, 2007).

We cannot deny the importance of using complexity management within the organization, especially in supply chain management. However, the tools needed to quantify the supply chain complexity have not been fully developed, making it a challenge to identify the value and costs of complexity.

In conclusion, several studies discuss the supply chain performance and how it is influenced by several factors (for example, (Divyaranjani, 2018)). However, there are a few regarding how we can use audit practices on data to improve the supply chain performance - which has been declining over the past few years.

1.2 Motivation

As companies expand and become more complex, we might believe that all complexity is harmful but that is not entirely true. There are specific cases where the complexity of a business adds value to itself as an outcome of the interaction between drivers, processes, activities, and resources.

Based on this, it is necessary to understand what part of the beneficial complexity is and what is not. Process mining has the potential and goal to, by using the massive amount of data available, identify and understand the data across different information systems.

Internal auditing of the supply chain management performance can be the beginning of a solution that provides the company with a way to reduce costs and have a competitive advantage. Auditing can decrease the cost of the organization's financial statements and risk associated. It can also affect the supply chain performance as a good audit practice maximizes the enterprise risk management, which enhances the supply chain performance itself (Hameed, 2017) (Hameed, 2017).

The problem arises when a proper model is not designed accordingly to today's needs, which embraces a new set of tools and techniques that look forward to auditing the available data in the most efficient way possible and finding the optimal supply chain based on the data provided and the maturity level of the business.

1.3 Objectives

The goal of the presented paper would be to purpose a model that contains a new set of tools and techniques that allow the appliance of process mining techniques in supply chain analysis based on audit data analytics.

To achieve this goal, the following intermediate objectives were defined:

- Perform a deep literature review on the topics of process mining and audit data analytics
- Investigate the role of audit in supply chain performance
    - Understand the supply chain complexity in the present – causes, and effects in the business
    - Use theory to further explain the correlation between audit and supply chain complexity
    - Use theory to further understand the correlation between process mining and supply chain complexity
- Provide insights into how we can use process mining techniques and metrics to simplify the supply chain complexity and audit data at the same time.
- Build a model that will improve the efficiency of the supply chains through a better data audit

- Demonstrate the model utility through a use case
- Construction of the possible model
- Validate the purposed model

1.4 Study Relevance and Importance

The study aims to contribute to the appliance of process mining techniques in supply chain analysis based on audit data analytics. All these different concepts contribute somehow to the company by enabling analytics of the processes, improving operations, and identifying bottlenecks.

Using data audit can help improve the efficiency of the company's processes while providing data evidence that identifies anomalies and trends. It is then possible to correct both the anomalies and the bottlenecks identified to improve the supply chain complexity in a way that brings value to the process.

On the other hand, finding a model that works to audit the company's data, using the right metrics, allows one to extract, manipulate, and process large amounts of data from the informatic systems that characterize today's business, especially the companies that have a global supply chain.

Regarding the role that audit data plays in the economy, it is the base to build confidence, reduce financial costs and contribute to an efficient allocation of capital (Doty, 2013). On the other hand, auditing data can identify, correct, and implement products, services, and processes, while providing clear reporting and greater assurance.

2 Methodology

The choice of methodology comes from looking at the desired output of this study and the desired process to achieve so. Keeping this in mind and having as the desired output the modulation of how we audit data to improve the supply chain performance and efficiency, it is believed that a Design Science Research (DSR) methodology is the most suited to obtain the best quality data and outcomes.

Our main purpose is to improve the efficiency and effectiveness of how we audit data and gain knowledge throughout that process, only possible by using implemented information systems within the organization. Although the capabilities of the information systems, and the capabilities of the organization itself, limit the extent to which the purpose of the usage of information systems can be achieved, it is still important to note that it is only possible to advance and create knowledge about both the management of information technology and its application in organizational and managerial contexts (Zmud, 1997).

This knowledge can only be obtained when involving two distinct yet complementary paradigms: behavioral science and design science (March & Smith, 1995). Banathy goes further by drawing a line between three archetypical approaches to organizational studies from where one can acquire knowledge: the sciences, the humanities, and the design sciences (Banathy, 1996) (Romme, 2003).

The foundation of the behavioral science paradigm lies in the natural science research methods, meaning that it looks to establish and validate theories that forecast human and organizational phenomena rounding the analysis, design, implementation, management, and use of information systems. As the main point is to have the information system achieve its stated purpose of improving the effectiveness and efficiency of an organization, then these theories would enlighten on the interplay among people, technology, and organizations needed to reach this goal (Hevner & Chatterjee, 2010).

On the other hand, the design science paradigm has its foundation in engineering and the sciences of the artificial (Simon & Laird, 2019) (Simon, 1996), being a problem-solving paradigm as it aims to create innovative solutions that enhance the analysis, design, implementations, management, and use of information systems effectively and efficiently (Denning, 1997) (Tsichritzis, 1998) (Hevner & Chatterjee, 2010). The artifacts developed throughout this process rely on existing theories that are applied, tested, modified, and extended (Markus, Majchrzak, & Gasser, 2002) (Walls, Widmeyer, & El Sawy, 1992).

On the contrary to both the science-based research and the humanities approach, - which aim to understand and interpret the human experience in and around organizations (Zald, 1993) – the design science approach combines these with organizations as artifacts. "Artifact" aims to describe something that can be transformed into an object or process, artificial or materialistic, while they answering the main purpose which is to solve an existent research problem (Hevner, Park, & Ram, 2004). This means that it provides additional understanding of organizational phenomena which translates to providing the various stakeholders of organizations with the kind of knowledge that is fundamental for designing, changing, and redirecting their organizations. This goes accordingly to the DSR characteristic of emphasizing solution-oriented knowledge by being the bridge between systems and outcomes to solve field problems. (Pigneur, 2013).

If the researcher is conducting explorative research addressing ill-structured problems, he will find many methodological challenges that he would not find if he focused on evaluative research instead. The solution to facing these problems is to base the research on design science, which specifically focuses on systematically tackling ill-structured problems (Dasgupta, 2003) (Niiniluoto, 1999).

Moreover, as the design of the artifacts is based on a problem-solving approach, it further promotes the DSR methodology when identifying the business needs early on in the process and finds a solution to the organizational problem (Gregor & Hevner, 2013), also meaning that the research questions are driven by field real problems (Pigneur, 2013). As little behavioral research was made on evaluating models, we look to develop and implement IT artifacts to make it possible for design researchers to grasp the target problem tackled by the artifact and understand the viability of the proposed solution.

In essence, DSR and Behavioral Science (BS) research establish a symbiotic research cycle that enables the resolution of underlying challenges encountered when implementing information technology.

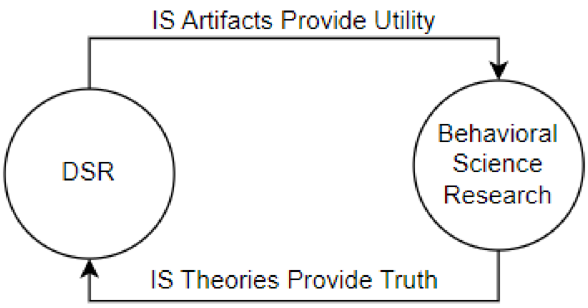

Figure 1 - Design Science and Behavioral Science Research Cycle (Hevner & Chatterjee, 2010)

The following sections will unfold, in detail, the concept of Design Science Research methodology followed by the process used for this research which will be consulted during the formulation of the final artifact as the solution for the problem encountered. The last subsection will explain in detail which will be the research strategy applied for this work based on the points already made.

2.1 Research Design Science

Design Science is, by definition, the approach that has its main goal of discovery and problem-solving instead of focusing on theoretical knowledge. According to (Simon, 1973) design science is the research that seeks to explore new solutions and alternatives to solve problems, to explain this explorative process, and to improve the problem-solving process.

The methodology used in this paper has the singularity that what is worth giving attention to would be the systems that do not exist yet, or the improved performance of given systems, which exist by creating new practices from scratch or by changing present social practices and situations into desired ones.

According to (Gregor & Hevner, 2013), the DRS knowledge contribution framework self divides into four quadrants and it is necessary to place our research process into one. The DSR project has the potential to vary in its research contributions depending on the level of the solution maturity and the level of the application domain maturity.

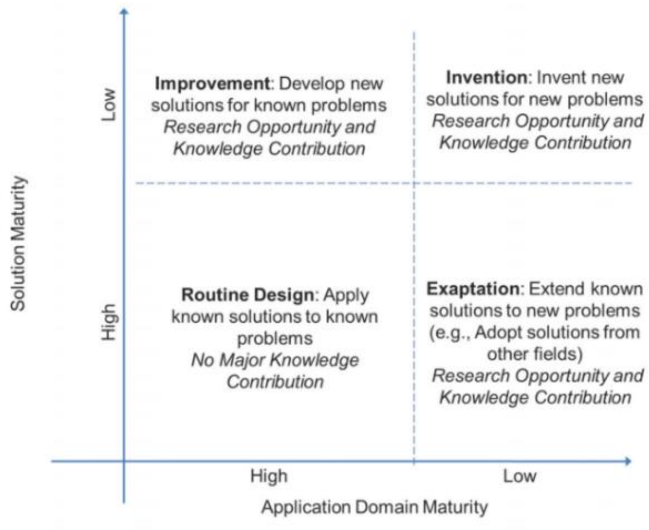

Figure 2 - Matrix Solution Maturity and Application Domain Maturity (Gregor & Hevner, 2013)

We can place the problem with the quadrant of *Improvement: develop new solutions for known problems.* The goal of the quadrant is to create better solutions, new solutions for already known problems to make the products, processes, services, technologies, ideas, etc., more efficient, and effective.

This master's study has as its main deliverable and objective the proposal of a new model – based on how the audition of data happens in the supply chain nowadays – that contains a new set of tools and techniques that allow the appliance of process mining techniques in the supply chain analysis. As so, we are focusing on the creation of a new solution for a problem that already exists, and the actual solution is not good enough to tackle the new reality. Focusing on answering how we can better audit data in the supply chain process, we are also making sure that the improved solution advances on the previous solution.

When we review how we position the knowledge contribution of DSR and to the scientific community of this specific research study, we refer to the table in the MIS Quarterly article, based on the framework developed by (Purao, 2002), which shows how to differentiate the DSR as research deliverables according to the maturity levels of the artifacts used.

Design Science Research Contribution Types		
	Contribution Types	**Example Artifacts**
More abstract, complete, and mature knowledge	Level 3 – Well-developed design theory about embedded phenomena	Design theories (mid-range and grand theories
↕	Level 2 – Nascent design theory – knowledge as operational principles/architecture	Constructs, methods, models, design, principles, technological rules
More specific, limited, and less mature knowledge	Level 1 – Situated implementation of artifact	Instantiations (software products or implements processes

Table 1 - Design Science Research Contribution Types (Purao, 2002)

Furthermore, the knowledge contribution in DSR is difficult to identify as it depends on several aspects such as the nature of the designed artifacts, the audience to whom it is communicated, the state of the field of knowledge it is inserted into, and the publication outlet - meaning that depending on the problem and solution maturity the DSR study can make different types and levels of research contributions to the different communities (Gregor & Hevner, 2013).

As we situated the problem within the *improvement* quadrant previously, we are proposing a model and methods that are already in use today as a research improvement, making these artifacts of level 2, according to the table above. Each of these is also a result of artifacts of level 1 that were developed throughout time. Our objective requires formulating an artifact of level 3 because of an improvement of the artifact of level 2 based on the understanding of the current problem and possible and found solutions. This new form of knowledge aims to contribute to the development of the descriptive knowledge of the field of study based on the studied prescriptive knowledge.

The DSR approach was first proposed by a team of researchers (Hevner, Park, & Ram, 2004), having as a main objective to offer guidance to Information Systems (IS) researchers and business managers on the methods of conducting, evaluating, and presenting DSR, fostering a comprehensive understanding of the process.

To conduct and evaluate an exhaustive Design Science Research, we must also look at the set of guidelines that also elaborate on research activities to consider in a conceptual framework – these guidelines being overlayed with the three research cycles described next.

Guideline	**Description**
Guideline 1: Design and artifact	DRS should create a functional artifact in the form of a model, method, construct, or installation

Guideline 2: Problem Relevance	The primary aim of DRS is to develop technology-based solutions and address important and pertinent business questions
Guideline 3: Design Evaluation	A design artifact's usefulness, quality, and effectiveness must be demonstrated convincingly through well-executed evaluation methods
Guideline 4: Research Contributions	Successful DSR should offer verifiable contributions to the areas of design artifacts, design foundations, and/or design methodologies
Guideline 5: Research Rigor	DRS depends on using rigorous methods in both the development and evaluation of the design artifact
Guideline 6: Design as a Search Process	Finding an effective artifact necessitates using available resources to achieve desired outcomes while adhering to the laws in the problem domain.
Guideline 7: Communication of Research	DSR must be communicated effectively to audiences from both technology and management backgrounds

Table 2 - DSR Guidelines (Hevner & Chatterjee, 2010)

When evaluating IS research combining both behavioral and design science paradigms, we need to take into consideration the below conceptual framework. According to (Hevner, Park, & Ram, 2004), we can get knowledge by applying the three inherent research cycles within the DSR research framework: the relevance cycle, the rigor cycle, and the design cycle.

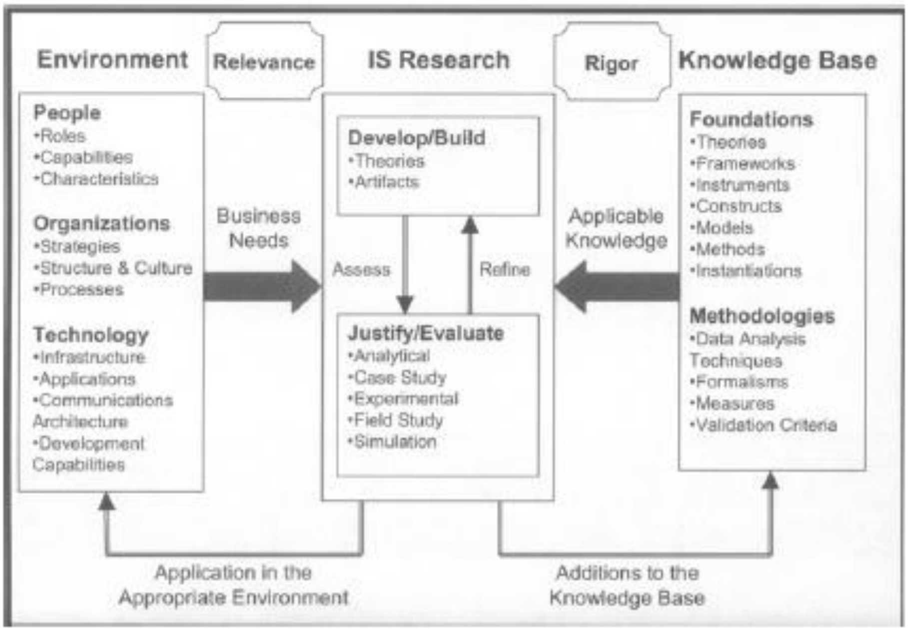

Figure 3 - Information Systems Framework (Hevner, Park, & Ram, 2004)

The Relevance Cycle: This cycle serves as the connection between the contextual environment of the research and the design science activities (Simon, 1996). The problem of the business in DSR generally comes from a need to improve a preexisting environment, by identifying opportunities for it. The environment is composed of people, organizations, and technologies, also including the goals, tasks, problems, and opportunities that define the business needs.

By framing research activities to address these needs, the relevance of the research is ensured based on the “problem” of the business. By doing this cycle, not it is only initiated the design research within the application context but also established the necessary research requirements and acceptance criteria for evaluating the research results, as the output needs to return to the application domain for further evaluation and analysis (Hevner & Chatterjee, 2010).

The Rigor Cycle: To conduct rigorous Design Science Research, it is crucial to have a vast knowledge base that encompasses scientific theories and engineering methods. Additionally, the knowledge base includes two additional types of knowledge: the experiences and expertise that define the current state-of-art of the application domain and the existing artifacts and processes that are present within the application domain (Hevner & Chatterjee, 2010).

Based on this, the rigor cycle provides the knowledge base that is composed of foundational theories, methodologies, frameworks, instruments, models, and methods used in the development/build phase of a research study. (Hevner, Park, & Ram, 2004). The outcomes of design research are threefold: firstly, they can extend existing theories and methods. Secondly, they can result in the creation of new design products and processes. Lastly, they can provide valuable insights gained through iterative design cycles and field testing in real-world settings (Hevner & Chatterjee, 2010).

The Design Cycle: Generates and evaluates alternative internal designs against requirements until a satisfactory one is achieved. The design cycle is then for inputted requirements from the relevance cycle and methods from the rigor cycle. While understanding the interdependence of the other two previous cycles, the Design Cycle also it is characterized by having relative independence during research execution. It is then important to find and maintain the balance between the efforts for constructing and evaluating the design artifact to have a successful performance of the design cycle (Hevner & Chatterjee, 2010).

2.2 The Design Science Research Process

In 2006 there was the creation of the remodel of the 2004 approach made by Hevner’s Design Science Research. This remodels organized the DSR approach into six steps that need can be followed in any order rather than sequentially. This development was made by Ken Peffers and was called the “Model for producing and presenting information systems research”, making it possible to begin at any step of the process and work the way outward. In this section, we will apply these six steps of the DSR Process Model to the practical example of this dissertation.

1. **Identify the problem and motivation**

This step defines the research problem and justifies the value of the proposed solution to motivate the researcher and the audience to pursue the solution and accept the results, making it also easier to understand the reasoning behind the researcher's logic and knowledge of the problem (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

2. **Define Objectives and a solution**

In this step we identify the problem to infer the objectives that need to be met to solve the problem, clarifying the specific goals that need to be set to work towards a solution. We also look to specify the objectives and requirements, quantitative or qualitative ones, to set the basis for the proposed solution based on the problem underlined and what is feasible to be done (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

3. **Design and development**

This step focused on the creation of the artifact that will be used as the solution for the problem identified in the first described step. According to (Hevner, Park, & Ram, 2004), the artifacts can come as models, methods or instantiations, or "new properties of technical, social, and/or informational resources" (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

By developing the artifact itself, it aims to create knowledge (Gregor & Hevner, 2013) by splitting the identified problem into simpler and smaller ones, aiming to meet the business needs – in this case the improvement of the supply chain of certain companies (Hevner, Park, & Ram, 2004).

4. **Demonstration**

Enables showing the efficacy in solving a problem of the artifact through experimentations, simulations, case studies, proof, et. However, this requires effective knowledge of how to use the artifact to solve a problem (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

5. **Evaluation**

By comparing the objectives of the proposed solution to the results observed we can observe and measure how well the artifact supports a solution to the identified problem in the first step (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

At the end of this step, the researcher can always go back to the design and development step if the solution does not correspond to the desired effectiveness or continue to the next step and leave further improvement to subsequent studies (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

6. Communication

In the last step, it is expected to communicate everything regarding the study. By communicating how the artifact was built and which was the evaluation process, we can collect some feedback to improve the solution for future implementations and following studies (Hevner, Park, & Ram, 2004).

2.3 Research Strategy of Investigation

In this subchapter, we look forward to explaining how all the steps in the previous subchapter will be used during this research and how we are going to approach each one of them, and what will be the subject in each step.

This is based on the following model based on the article (Peffers, Rothenberger, & Tuunanen, 2007).

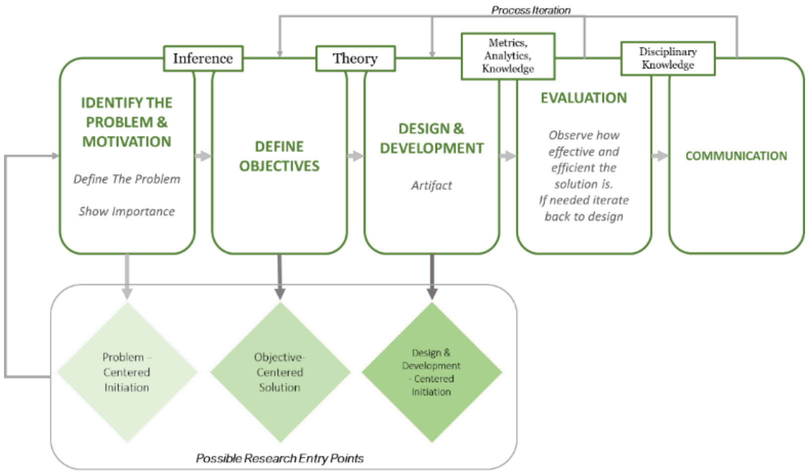

Figure 4 - Research Strategy of Investigation I

This thesis, as said previously, is driven by an objective-centered solution – to build a model that will improve the efficiency of the supply chain of different organizations through better use of data analysis – starting at design objectives and solution stages.

To be able to achieve this, we need to specify and **define those objectives and a solution** around the problem that we encountered in the previous chapter – the inexistence of enough studies on how we can rely on data audit practices for supply chains to make them the more efficient, taking in consideration the new tools and techniques available and the maturity level of the business. This will enable us to know which areas will serve as a base for this study according to the final solution.

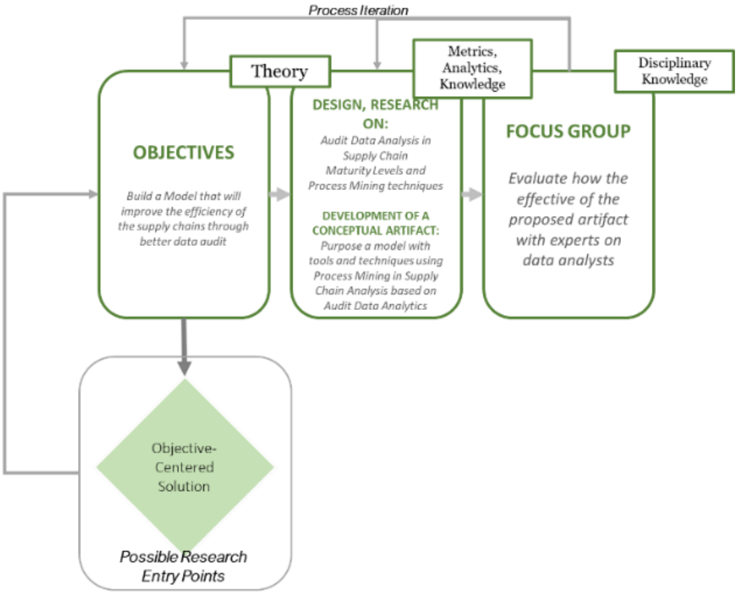

Figure 5 - Research Strategy of Investigation II

The next step consists of the **Design and Development stage** of the study around two main topics of research that are highly correlated: the main concepts of Audit Data Analysis, Process Mining - techniques, and tools around the concepts; and how we can relate the main concepts to achieve efficiency. This is based on the second step which involves researching the best tools and techniques that are already in use throughout the market and organizations that have a highly complex supply chain.

Following the initial testing in a hypothetical use case, the artifact will undergo another careful examination by engaging experts from relevant scientific and practical disciplines. This evaluation is crucial to ensure that the research output effectively achieves the objectives outlined in phase two (Collatto, Dresh, Lacerda, & Bentz, 2018) and enables necessary enhancements while identifying imitations.

Conducting interviews as an evaluation technique is a commonly used approach to validate the design of the artifact and gather feedback on its efficiency and effectiveness (Cleven, Gubler, & Hüner, 2009). Since the artifact aims to be usable by non-experts, the interviewee selection will prioritize participants from expert fields and potential users in companies. This approach will aid in validating the artifact's accuracy and usability (Sari, Hidayatno, Suzianti, Hartono, & Susanto, 2021).

If possible, within the scope of this project, the feedback obtained from the interviews will be incorporated into a revised version of the initial artifact. Subsequently, the final artifact will meet all necessary criteria to be tested in real companies, thereby contributing to both practical applications and society at large. However, it is important to note that the implementation of the framework in a real-world context falls beyond the scope of this thesis.

Ultimately, the outcomes of this research and its contributions will be disseminated to the public through publication and presentation to an academic committee for assessment. Sharing the research with other scholars is essential to fulfilling the intended theoretical contribution of the work and encourages further progress in the field, building upon the findings presented in this paper.

As we are skipping the demonstration step in this dissertation – as our focus is not to simulate the artifact – this step will be based on mere speculation on how we believe the new solution would impact certain metrics and analysis techniques in the future through a fictional use case.

To sum up the above sections 2.1, 2.2, and 2.3, and giving focus to the Design and Development phase, the figure below looks forward to representing how each part will be approached throughout this thesis and which chapter corresponds to the previously explained framework. In this way, the introductory chapter of the thesis will answer the first and second steps in the DSR Model – Problem and objectives identification – in the same way that answers the same problem relative to the research strategy.

The Design and Development phase in the DSR Model, corresponding to chapters 3 (environment), 4 (literature review), and part of chapter 5 (subsections related to the framework and the use case) in the layout of this paper will look forward to defining the relevant cycle, the rigor cycle, and the design cycle of the research strategy, followed by a fictional use case.

At this stage it is required to have a clear understanding of the desired solutions outcome to design an effective and efficient assessment tool that will be the basis of our artifact – that will be created in this phase. For that, a thorough examination of the state-of-the-art research on Audit Data Analysis and Process Mining and the maturity levels of the process in scope is necessary.

Therefore, a literature review (LR) will be conducted to summarize the existing knowledge in this scientific area. This means that the focus will be on collecting characteristics rather than formulating a hypothesis, making the research qualitative in nature (Kohlegger, Maier, & Thalmann, 2009). This will be our rigor cycle.

For the Relevant Cycle, we will look at today's environment of supply chains and understand which are the current opportunities and challenges that the organizations are facing. By understanding and collecting this information, it is possible to evaluate the tools and techniques that are already in use, especially when focusing on technologically focused process mining techniques, in the market and in organizations that have a highly complex supply chain.

Following a deep detailed search that ensures a thorough review of the appropriate literature and therefore enhances the credibility of the scientific value of the work performed, we can, based on the learnings of this review, develop a model that combines the opportunities and challenges understood in the rigor cycle with the need within the industry environment in the relevant cycle.

The fictional use case anticipates showcasing the validity and usability of the framework developed. This use case will be purely a theoretical one, as a real use case is not in the scope of this work as it would be necessary for a detailed adoption of the framework by a company. Therefore, the illustrative use case chosen only has the usefulness of being an example of how the framework would be applicable in a company, to test and validate its functionality.

Lastly, step 4 of the DSR Model, or the evaluation will take part in the last subsection of chapter 5 of the thesis.

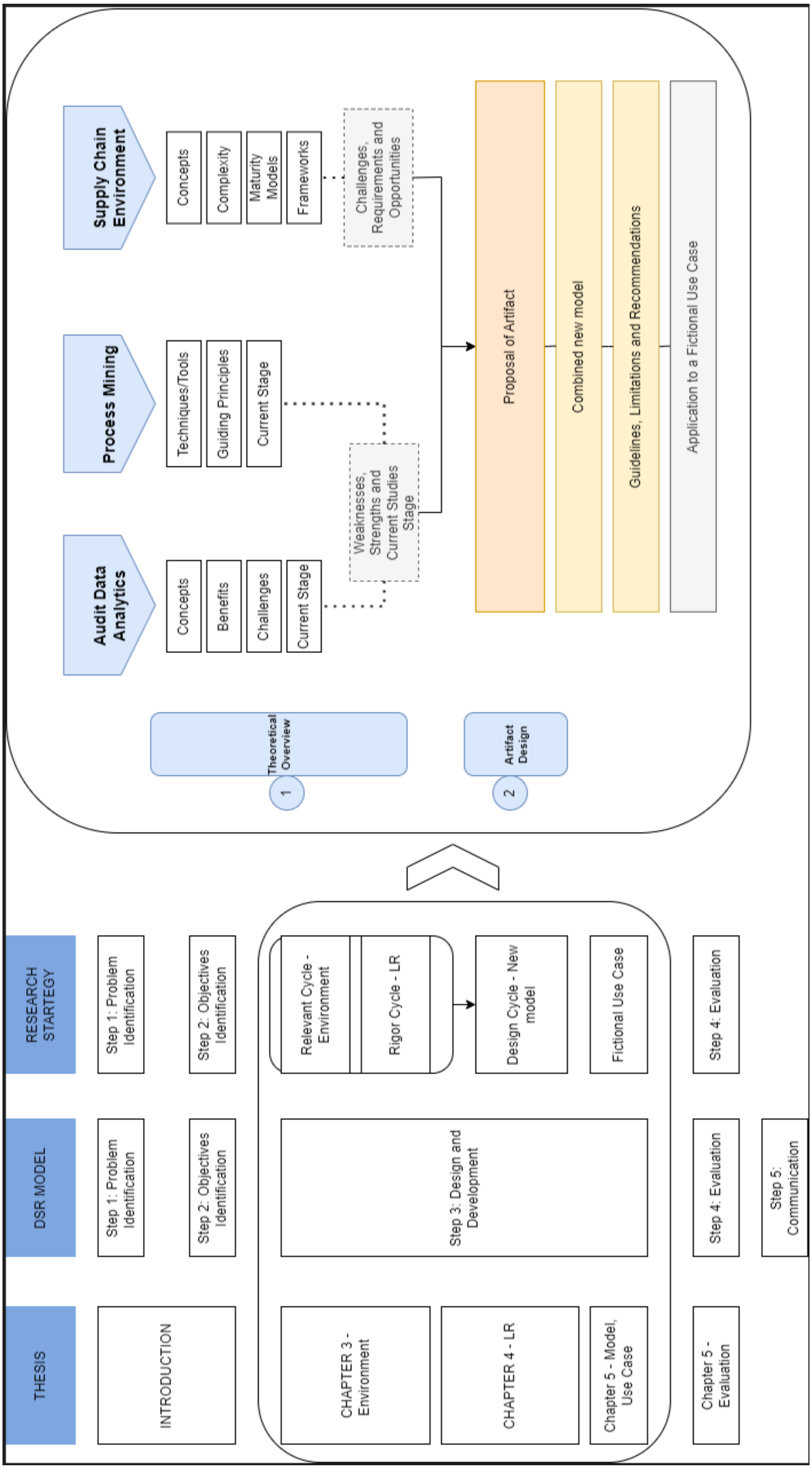

Figure 6 – Research Strategy of Investigation III

3 Supply Chain Management - Environment

This section provides an overview of the literature on Supply Chain Management (SCM), with a focus on the key themes, theories, and frameworks that have emerged in the field. The chapter begins by defining SCM and discussing its evolution over time and its benefits. Finally, it ends with the maturity models in supply chain management that will help us understand how an organization is positioned regarding its supply chain's process maturity.

3.1 Concepts

According to the Council of Supply Chain Management Professionals (CSCMP), supply chain management is defined as "the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. Supply chain management integrates supply and demand management within and across companies. Supply chain management is an integrating function with primary responsibility for linking major business functions and business processes within and across companies into a cohesive and high-performing business model. It includes all the logistics management noted above, as well as manufacturing operations, and it drives coordination of processes and activities with and across marketing, sales, product design, finance, and information technology." ((CSCMP), 2023).

The definition of supply chain management can go deeper when we add the flow of information that each supply chain process possesses. Several authors produced another definition for SCM that also incorporates this second intangible part: "the synchronization of a firm's processes with those of its suppliers and customers to match the flow of materials, services, and information with customer demand." (Simchi-Levi, Kaminsky, & Simchi-Levi, 2018).

There are several more definitions of supply chain management, overall conducting to the same principle of what is SCM: an essential aspect of modern operations that includes everything from the planning phase up until the monitorization of all the activities involved in the flow of the products, services, and information from the suppliers until the final consumer/client. According to the CSCMP ((CSCMP), 2023), ensuring the satisfaction of the seven rights of fulfillment – a guide for the companies to meet customer expectations – requires an efficient movement of products from the source to the destination making it crucial to establish a two-way flow of information and data that enhances the visibility of demand and enables supply chain managers to promptly identify any issues within their supply chain network. This ability is regarded as a critical factor in the decision-making process of supply chain managers.

An effective SCM enables organizations to optimize their operations, reducing the overall costs of the business while maintaining customer satisfaction and, consequently, gaining a competitive advantage in the marketplace. SCM plays a vital part in achieving company success and meeting customer requirements by ensuring efficient and effective execution of the key activities that constitute a supply chain. To comprehensively analyze how the supply chain plays a big role in this thesis we need to look that the fundamental elements of the end-to-end process: planning, sourcing, manufacturing, warehousing, distribution, and customer interface. The flow of materials moves through the supply chain from suppliers to customers, while the flow of information goes in the opposite direction.

As stated before, in today's highly competitive market a company that possesses a more effective and efficient supply chain holds a competitive edge over other organizations. Consequently, supply chain management has emerged as a significant challenge for companies operating in several diverse industries as it is the source of competitive advantage (Mentzer, 2004). Based on this, SCM has forced many companies to rethink their competitive strategies, a subject which we see a trend to move towards collaboration and an exchange of a lot of data while shifting the focus to "win with data" (Hopkins, LaValle, & Balbone, 2010) (Zacharia, Sanders, & Nix, 2011).

3.1.1 Hierarchical Levels in SCM

The activities in SCM can be categorized into three hierarchical levels, namely strategic level, tactical level, and operational level. These levels are characterized by different time horizons, ranging from several years to a few hours (Larson, 2001)

At the strategic level, we look to make long-term decisions that set the overall direction and objectives of the supply chain. It involves determining the supply chain network design (Tsiakis, Shah, & Pantelides, 2001) (Santoso, Ahmed, Goetschalckx, & Shapiro, 2005) (Georgiadis, Tsialis, Longinidis, & Sofioglou, 2011), such as the configuration of facilities and locations (Owen & Daskin, 1998) (Snyder, 2004) (Tsiakis & Papageorgiou, 2008), to achieve the desired targets and objectives (Shapiro, 2004). These decision-making points have a lasting impact on the supply chain and consequently on the organization, typically spanning several years or even decades.

On the tactical level, we focus on medium-term decisions that translate strategic decisions into actions. By taking these decisions we make sure that we efficiently utilize the resources. Typically, tactical decisions include production and distribution planning (Timpe & Kallrath, 2000; Lee & Kim, 2002; Mula, Poler, Garcia-Sabater, & Lario, 2006) or inventory policies (Gupta, Maana, & McDonald, 2000; Disney & Towill, 2003), being updated periodically ranging from a few weeks to a few years.

Lastly, on an operational level, we take short-term decisions that concern the implementation of tasks and operations necessary to fulfill the objectives set at the tactical level. On this level, we take more time and effort to look at the details taken into consideration in these decisions, and, because of that, they are updated frequently, on a daily or weekly basis (Çetinkaya & Lee, 2000; Higgins, Beashel, & Harrison, 2006).

3.1.2 Key Elements in SCM

From the above two subsections, we can now understand better that SCM encompasses a wide range of issues related to various stages within the supply chain. Extensive research has been conducted on the six elements, identified by (Capello, Lösch, & Schmitz, 2008) and the coordination and integration between them:

- Service level management including customer segmentation (Chen, 2001) and service level management (Yoo, Hong, & Kim, 2009);
- Order and demand management including sales demand planning and forecasting (Liang & Huang, 2006), inventory management (Lee & Billington, 1992) and order entry and fulfillment (Akhil & Sharman, 1992);
- Production Management that encompasses network configuration/rationalization (Pyke & Cohen, 1994), and production execution (Dickersbach, 2009);
- Supply management including procurement planning and supplier performance management, distribution management that includes network configuration/rationalization, (Chopra, 2003) warehousing (Landers, Cole, Walker, & Kirk, 2000), and transportation (Morash & Clinton, 197);
- Integrated SCM planning and execution that encompasses the integration of SCM processes, IT systems, organization, and performance measurement (Power, 2005).

These areas of SCM have received considerable research attention, highlighting their significance in achieving effective supply chain operations and overall business success.

3.2 Maturity Models in Supply Chain Management

The Maturity Model (MM) that is widely known among researchers is called the Capability Maturity Model (CMM), which is a framework that describes the key elements of a software process, outlining the evolution of the improvement path from an ad hoc, immature process to a mature, disciplined process. For that, assess a process on a scale of five process maturity levels. Kohlegger (Kohlegger, Maier, & Thalmann, 2009) affirms that a maturity model “represents phases of increasing quantitative or qualitative capability changes of a maturing element to assess its advantages concerning defined focus areas”. Each level of a maturity model ranks the organization according to its standardization of processes in different and distinct areas of the business. Apart from level 1, each maturity level is composed of several key process areas and one of them is organized into five sections called common features. Each maturity level provides a layer in the foundation for continuous process improvement, compromising a set of goals that, when satisfied, stabilize an important component of the software process. Achieving a higher level in the maturity model increases the process capability of the organization.

The 5 previously mentioned maturity levels can be explained below (Paulk, Weber, Garcia, Chrissis, & Bush, 1993):

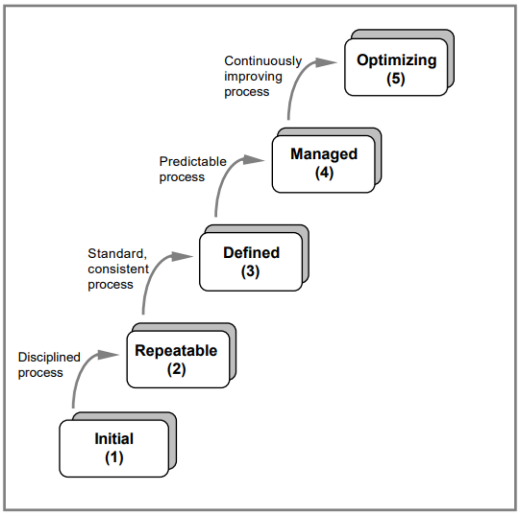

Figure 7 - The Five Levels of Software Process Maturity (Paulk, Weber, Garcia, Chrissis, & Bush, 1993)

Initial Level: At this level, organizations typically lack a stable environment for software development and maintenance. The software process capability is unpredictable due to constantly changing or ad-hoc processes. Success largely depends on individual capabilities rather than organizational capability.

Repeatable Level: Policies and procedures for managing a software project are established, allowing organizations to replicate successful practices from previous projects. The software process capability of organizations can be described as disciplined. The process is under the effective control of a project management system, and earlier successes can be repeated.

Defined Level: At this level, both software engineering and management processes are documented and integrated into a standard process for the organization. These processes are used to help managers and technical staff perform more effectively. The software process capability of organizations at this level is standard and consistent.

Managed Level: This level involves setting quantitative quality goals for software products and processes and measuring productivity and quality for important software process activities across all projects. The software process capability at this level can be described as predictable because the process is measured and operates within measurable limits.

Optimizing Level: The organization focuses on continuous process improvement. The aim is to identify weaknesses, strengthen the process proactively, and prevent defects. Innovations that exploit the best software engineering practices are identified and adopted throughout the organization. The software process capability of organizations at this level is continuously improving.

But how can we guarantee that the business survives in times of high velocity? The answer lies in making supply chain management sustainable. Beske and Seuring (Beske & Seuring, 2014) defend that this is achieved by keeping in mind that supply chain management is considered "the systematic, strategic coordination of the traditional business functions and the tactics across business functions within a particular company and across businesses within the supply chain, to improve the long-term performance of the individual companies and the supply chain as a whole".

As always, this drives some challenges concerning the management of the resources: the drive for a sustainable supply chain (SC) requires the best coordination between activities and information sharing which might increase the level of complexity – the root cause evaluated in the thesis. This makes it necessary to dive also decision-making processes at various levels within the SC (Reefke, Ahmed, & David Sundaram, 2014).

It is also important to evaluate how the maturity models are presented. We can denote the following phases: (i) a descriptive tool for the evaluation of strengths and weaknesses (Neuhauser, 2004) ; (ii) a prescriptive instrument to help develop a guide (roadmap) for performance improvement (McCormack, Ladeira, & Oliveira, 2008); or (iii) a comparative tool to evaluate the processes/organization and compare it with standards and best practices from other organizations, thus enabling internal and external benchmarking (Röglinger, Pöppelbuß, & Becker, 2012).

In the end, it enables decision-makers to evaluate their organization's efforts regarding sustainability. If we apply MM to the SCM it is, then possible to assess the level of maturity of the SC.

3.3 Complexity in SCM

The complexity in supply chains arises from the interconnected material and information flow between various partners involved in the chain. Traditionally, these flows were organized sequentially starting from the supplier and ending with the final customer. However, with the existing modern technologies, the nature of information flows has undergone a significant transformation where information flows are no longer limited to a linear progression. Instead, they are dynamic and simultaneous enabling real-time data sharing and collaboration among all supply chain partners, regardless of their position in the sequence.

This allows for enhanced coordination, visibility, and responsiveness across the entire supply chain here partners can have access to information such as inventory levels, demand forecast, and production schedules, enabling better decision-making and more efficient operation which is the end goal.

Overall, the shift towards simultaneous information exchange in supply chains has significantly contributed to the complexity of the modern SCM. It requires supply chain partners to adapt to their processes and technologies to leverage the benefits of real-time data sharing and collaboration. However, when effectively managed, those complex information flows can lead to improves efficiency, flexibility, and overall performance. To obtain the necessary tools for modeling, evaluating, and enhancing the entire supply chain, it is essential to gain a comprehensive understanding of the business processes. For this, we look at the Supply Chain Operations Reference (SCOR) model.

3.3.1 SCOR Model

There is a highly competitive nature between businesses, emphasizing the necessity for companies to enhance their supply chain efficiency and reduce costs to stay profitable. To do this, firms need to integrate their processes and compare them with those of other companies for analysis, improvement, and benchmarking purposes. Performance measurements are crucial in this context and can be easily evaluated. The misalignment can lead to poor service, high inventory, unexpected costs, limited growth and profits, and loss of market share.

Supply chains have grown more efficient over time due to the use of advanced tools and technologies. In their aim to improve service levels, supply chain sellers felt the need to use a standard model to base their operations and measure their performance. This led to the development of the SCOR model by the Supply Chain Council (SCC). The SCOR model is used to measure chain performance, allowing companies to gain a competitive advantage and enhance their organizational performance.

The use of the SCOR model it is created a shared language and promotes the synchronization of practices among stakeholders in the supply chain by modeling logistics systems. It serves as a valuable tool for diagnosing the flow of business between a company's primary and secondary customers and suppliers. Users of the model can benefit from performance-standard metrics that measure process performance and establish strategic objectives. It also offers process-standard descriptions of management processes and their interrelationships. It is structured around five fundamental management processes:

- Planning: his process involves gathering customer requirements, collecting information on available resources, and balancing requirements and resources to determine planned capabilities and identify any resource gaps.
- Sourcing: These processes encompass activities related to the ordering and receipt of goods and services from suppliers.
- Making: his process category encompasses activities involved in the conversion of materials or the creation of content for services.
- Delivering: These processes encompass activities associated with the creation, maintenance, and fulfillment of customer orders.
- Returning: his process category covers activities related to the reverse flow of goods from the customer, including product returns and reverse logistics.

Organizations can establish a common framework for understanding and improving their supply chain operations. It facilitates effective communication, performance measurement, and process improvement across the entire supply chain network. The article (Delipinar & Kocaoglu, 2016) looks over several studies, some of them combining the use of the SCOR model with information systems within supply chains. It is worth highlighting from the article two conclusions to take into consideration moving forward.

It was studied that using a standard model, which connects information, features, and practices with knowledge and skills asserts the alignment of business processes, information systems, and ERP implementation has become the main subject of their study. It is believed that this SCOR-based alignment reference model bolsters a more efficient "multi-view" for Enterprise Resource Planning (ERP) projects. This alignment approach can aid business process management in an operational environment and can direct a continuous alignment approach including process management in a re-engineering life cycle.

Another study proposed the use of the SCOR model as a connector between business objectives with the operation of logistics by giving a systematic approach to finding the performance of the firm in supply chain management. Integration with it leads all activities' logistics process indicators (KPI)'s to suitable for providing successful strategies for the firm.

3.3.2 5VS

The 5 V's of supply chain refer to the key characteristics or dimensions that are essential for understanding and managing modern supply chains. These dimensions are often used to describe the challenges and complexities that arise in SCM (Ishwarappa & Anuradha, 2015).

Volume: Refers to the quality or scale of goods and information that flow through a supply chain. It involves understanding the magnitude of demand, production levels, inventory levels, and transportation capacities. Managing volume effectively is crucial to ensure the smooth flow of goods and minimize bottlenecks or overstocking.

Velocity: Velocity represents the speed at which goods, information, and money move through a supply chain. It involves managing lead times, transportation schedules, and response times to meet customer expectations. In today's fast-paced business environment, managing velocity is crucial to meet short lead times, responding to changing customer demands quickly, and reducing time-to-market.

Variety: Variety refers to the diversity or range of products, suppliers, customers, and channels in a supply chain. It involves managing the complexity of different product variants, supplier capabilities, customer preferences, and distribution channels. Effectively managing variety requires flexible processes, robust supplier networks, and adaptable logistics capabilities.

Variability: Variability relates to the uncertainties and fluctuations that occur within a supply chain. It includes demand variability, supply disruptions, market dynamics, and other sources of uncertainty. Managing variability involves developing strategies to accurately forecast demand, buffer against uncertainties, and build resilience within the supply chain.

Visibility: Visibility refers to the ability to track and monitor the movement of goods, information, and funds across the supply chain. It involves having real-time access to relevant data and insights at various stages, from sourcing raw materials to final delivery. Improved visibility enables better decision-making, coordination, and collaboration among supply chain partners.

Companies are actively exploring ways to utilize large volumes of data to gain a competitive advantage. They are leveraging data to achieve precise market demand forecasting, customize services, and develop new business models. The question stands when we try to understand how we can address the complexity of supply chains with the availability of large sets of data and the challenges that may arise with that. In summary, we can understand that data plays a key role in several parts of the supply chain:

- Planning: Big Data helps reduce the risk associated with infrastructure investments and external capacity contracts.
- Supplier networks: Big data revolutionizes how supplier networks are formed, expanded, enter new markets, and mature over time.
- Production: Utilizing a combination of analytical techniques optimizes manufacturing processes, shop-floor management, and manufacturing logistics. This enables the production of new products more efficiently and reduces logistics costs.

- Distribution: Big data analytics aid in forecasting demand changes, allowing for better supply matching. This benefits the manufacturing, retail, transportation, and logistics industries.
- Returns: Big data analytics enable understanding customers' perceptions of products and services, uncovering their unobservable characteristics, and anticipating future consumer demands. This knowledge facilitates the development of more customized products and services, improving customer satisfaction.

By leveraging analytics on big data, organizations can extract valuable insights that enhance decision-making and enable informed choices. It is defined that to succeed in the world of Big Data, organizations should recognize data as a strategic asset rather than a mere source of information.

This perspective allows supply chain management to realize the inherent economic value of data and capitalize on it through revenue-generating activities in conjunction with Big Data Analytics (BDA). Real-time analysis of markets, production, and sales data using simulation, statistics, and visualization techniques facilitates the computation of key performance indicators relevant to supply chain decision-making.

3.4 Challenges

The utilization of Big Data offers logistics participants a fresh avenue for gaining a competitive edge in conducting supply chain management. This enables them to achieve heightened visibility, adapt to demand and capacity fluctuations in real time, and acquire valuable insights into customer behaviors and patterns. These advantages allow for smarter pricing strategies and improved product offerings.

Despite the points mentioned earlier, this review underscores several avenues for future research in this area. The challenges can be summarized as follows:

- Identifying relevant theories for studying big data in supply chain contexts.
- Developing metrics to measure supply chain performance in the context of big data.
- Determining how to integrate supply chain management initiatives with big data analytics programs.
- Investigating the impact of big data on the external supply chain.

Several research works have been conducted over time to try to group the different challenges present in the supply chains into several categories that would help us identify them and try to tackle them at the same time.

(Papageorgiou, 2009) categorized key issues in SCM into three categories: supply chain design, supply chain planning and scheduling, and supply control. The review also identified future challenges in the field, including optimization under uncertainties, multiscale optimization, development of efficient solution procedures, multi-objective optimization considering environmental impacts, and the emergence of sustainable and healthcare-related supply chains.

With this division in mind, we also can say that the supply chain industry has a potential for development with the development of new technologies. However, it also suffers from a wide range of issues in SCM that can impact its efficiency, effectiveness, and resilience. Alongside a few challenges briefly mentioned previously in this chapter (supply chain complexity and regulatory compliance), we also encounter a lack of transparency and information sharing that delay data retrieval that then affects every stage of a logistics network.

We also find several risks while managing the different parties within the logistics flow, that being inventory management (as we find it challenging to keep the inventory balance while meeting the customer demands and minimizing costs), supplier management (where one company by depending on its supplier has the risk of quality issues, capacity constraints, and supplier failures; In this point is also crucial to highlight the risk of the supplier-organization relationship), cost pressures (fluctuations around raw material prices, transportation costs, and labor expenses while maintaining quality and service levels) and demand volatility (that ties up with the inventory management risk where fluctuations in customer demand can disrupt the supply chain, possibly leading to stockouts or excess inventory which is highly costly to the organization).

Although adopting and integrating new technologies such as the Internet of Things (IoT) and automation can enhance supply chain efficiency, their management also requires significant investments, skills, and change management efforts that not all companies are in the right position to do.

Insufficient knowledge is often cited as a drawback of the recent influx of voluminous data. Consequently, data scientists may encounter difficulties in enhancing decision-making without possessing expertise in SCM activities (Waller & Fawcett, 2013). Many scholars have expressed skepticism toward the exaggerated hype surrounding data analytics. This bias is, however, in the significant awareness of big data techniques and their effectiveness, despite limited empirical research on the factors enabling their success and the potential obstacles they may encounter (Schoenherr & Speier-Pero, 2015).

Summing up, although the great potential around the development of supply chains, the evolution of SCM relies on efficient and reliable data management in which data collected from supply chains are supposed to be stored, integrated, and retrieved with reliability and high efficiency.

Knowing this, this thesis also looks for the opportunity to think about how we can find a better way to manage the large data sets that companies nowadays own and use that to drive decisions in the long-term perspective. The main idea is to identify bottlenecks and inefficiencies within the processes of the supply chain that can lead to any of the previously named risks within the processes.

Once identified we could be able to use different and new techniques that help organizations investigate the data flow and inputs that they own and improve their operational processes by creating clear process models, by discovering, monitoring, and improving the real process of the supply chain. For that, we look at what is happening in the Process Mining area and how organizations are auditing big data for analytics.

4 Literature Review

The purpose of this chapter is to give an overview of the key components of the study, establish the background and support the significance of the research by reviewing related literature.

This chapter focuses on three main topics: Audit Data Analytics (ADA), Supply Chain, and Process Mining, underlining the correlation between the topics and each topic's techniques and metrics. The chapter concludes with the presentation of the hypotheses aimed to study throughout this study.

4.1 Audit Data Analytics

The first main topic that we abord in this chapter is the concept of Audit Data Analytics. In this section, it will be introduced the concept of ADA and its application. Afterward, it is provided a discussion on the benefits and challenges of ADA and the current state of usage of it nowadays and how it is applied in Supply Chain so far. Lastly, there is also an introduction to the Value Through Analytics (VTA) model.

4.1.1 The Concept of Audit Data Analytics

ADA is frequently associated with big data (Earley, 2015), making them concepts that are highly discussed together. For instance, in accounting literature, "big data" is defined by the types of analysis that can be performed with it, such as data analytics or predictive analytics, rather than the type of data source, as acknowledged by (Alles & Gray, 2014). So, understanding big data can provide insight into the concept of ADA as it is starting to be widely utilized in current business practices.

Big data has been defined as ''high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making'' (Gartner, 2013). It is the data that results from a process that is called datafication, where everything can be recorded and turned into data (Cao, Stewart, & Chychyla, 2015). The value of big data lies in the ability to extract useful knowledge from it, which can provide a competitive edge for a business (Cao, Stewart, & Chychyla, 2015). The concept of Big Data can be divided into three coexisting dimensions: Volume (the amount of data an organization collects), Velocity (the speed at which the data is created and processed), and Variety (the different types of data) (Appelbaum, 2016).

In addition to the initial three dimensions, (Kaur & Sood, 2017), (Lee I. , 2017), (Alaoui & Gahi, 2019) introduced four more aspects of big data: Veracity, Variability, Value, and Visualization. Veracity, a dimension added by the International Business Machines Corporation (IBM), encapsulates the unreliability and uncertainty inherent in data sources due to factors like inaccuracy, incompleteness, inconsistency, subjectivity, latency, and deception.

The dimension of Value, introduced by Oracle, signifies the importance of useful knowledge from big data through data analytics (Alaoui & Gahi, 2019) (Kaur & Sood, 2017) (Lee I. , 2017). Businesses must understand the potential value of big data. While raw big data may have low intrinsic value, proper analysis can transform it into a high-value strategic asset (Lee I. , 2017).

Lastly, Visualization, the seventh dimension, involves the presentation of big data in a comprehensible format, typically via graphs and charts (Alaoui & Gahi, 2019) (Kaur & Sood, 2017). This visualization aids in the interpretation and understanding of the data.

Big data analytics is a process used to extract intelligence from both structured and unstructured data. While structured data is essential in big data, an increasing amount of data is unstructured, which cannot be processed using traditional data management technologies (Lee I. , 2017). This process, known as big data analytics, involves inspecting, cleaning, transforming, and modeling the data (Cao, Stewart, & Chychyla, 2015).

Although these are two different concepts – big data and data analytics – we can relate them: based on the presentations in the 2015 and 2016 meetings (Alles M. G., 2015) - we correlate the two concepts in the following table. We see a shift from cell A, where traditional techniques and sources were used such as Excel to analyze samples of data, especially accounting data, to cells B and C. The first is used extended data analytics techniques such as visualizations and predictive analytics, while in the latter what changes is the data source shifting from traditional to trying to incorporate also non-financial data – big data, while maintaining the traditional techniques in use. We can consider that cell D is where we want to evolve in the next steps regarding auditioning data, using big data and predictive analytics, and visualization tools to drive more data-driven decisions within the business.

For the scope of this study, we must focus our attention on the movement to cell C as applying audit data analytics in the supply chain environment is based on non-financial data, and, in most cases, we will be dealing with big data – the bigger the supply chain, the more data we must audit.

		Data Analytic Techniques	
		Traditional (Excel, ACL, Idea)	**Extended** (Visualization, Predictive analytics)
Data Sources	**Traditional** (Accounting & Financial)	A	B
	Extended (Non-Financial Data → Big Data)	C	D

Figure 8 - Paths to expand data analytics in financial statement audits (Alles & Gray, 2014)

The use of big data analytics can be defined as ADA, when talking about the audit profession, as: “the science and art of discovering and analyzing patterns, identifying anomalies, and extracting other useful information in data underlying or related to the subject matter of an audit through analysis, modeling, and visualization for planning or performing the audit” (Eilifsen, Kinserdal, & Messier, 2020).

Data analytics has been utilized in various forms for a long period, but companies are now discovering more advanced and timely ways to use data analytics to improve their operations. By using data analytics, businesses can identify new opportunities, reduce costs, and make quicker, more efficient decisions. From identifying the potential for new products and services to detecting the possibility of losing clients and taking action to retain them, data analytics is now a fundamental part of modern business (ACCA, 2021). The data used by companies typically includes both internal and external data and can be both quantitative and qualitative. This process is often supported by specialized software that is designed to combine and analyze data from various sources and formats, including audio and visual data in addition to traditional text and numerical data (ACCA, 2021).

In recent years, we see several researchers diving into the direction of proving the number of potential insights that applying ADA tools to audit would have. For instance, the International Auditing and Assurance Standards Board acknowledges that “In an increasingly complex and high-volume data environment, the use of technology and data analytics offers opportunities for the auditor to obtain a more effective and robust understanding of the entity and its environment, enhancing the quality of the auditor’s risk assessment and response.’’ ((IAASB), 2016).

Following this, we also notice that major international public accounting firms have invested in ADA to transform how they practice their services “Data analytics is changing both the way we conduct our audits and what those audits deliver. It allows us to extract and analyze ever-larger data sets. Further use of data analytics will allow us to deliver effective audits more efficiently’’ ((PwC), 2017).

4.1.2 Benefits and Challenges of ADA

As we are looking to understand the concept of Audit Data Analytics to correctly use it further in this study, we should consider its benefits and challenges. In this section, we discuss first the benefits, followed by the challenges.

Benefits

“99% of respondents noted that data and analytics is at least somewhat important to their business strategy, and 96% expressed that they could make better use of big data within their organizations” (KPMG, 2014).

- Auditors can evaluate a greater number of transactions than they do now, providing greater insights into clients' processes

  By increasing the appropriate amount of audit evidence, it will be possible to automatically evaluate the transactions, having the capacity to examine 100% of the clients' transactions using high-powered analytics. This means that is also easier to focus on the main points that can have potential concerns and risks, learning from historical data and previous audits (Earley, 2015).

  According to the American Institute of Certified Public Accountants (AICPA, 2020) audit evidence should be sufficient and appropriate. Big data and the use of ADA contribute to the sufficiency of audit evidence due to the large volume and variety of data (Appelbaum, Kogan, & Vasarhelyi, 2017): (Brown-Liburd & Vasarhelyi, 2015) (Yoon, 2015).

- Fraud will be easier to detect due to tools and technology

  Based on the literature of several authors, (ACCA, 2021) (Appelbaum, Kogan, & Vasarhelyi, Big data and analytics in the modern audit engagement: Research needs, 2017) (Cao, Stewart, & Chychyla, 2015) (Earley, 2015), as we have a better understanding of the business environment and a bigger amount of available data to audit and analyze, we can detect events that are correlated.

- Auditors can solve problems that are beyond current capabilities by utilizing external data to inform audits

  By analyzing nonfinancial and external data, high-risk areas can be identified, leading to a more efficient audit (ACCA, 2021) (Appelbaum, Kogan, & Vasarhelyi, 2017) (Cao, Stewart, & Chychyla, 2015) (Dowling & Leech, 2007) (Earley, 2015). The use of computer programs for certain parts of the audit also improves efficiency Additionally, the use of Automated Data Analysis leads to more accurate and relevant audit evidence (Yoon, 2015).

- Reduce the risk

  According to several authors, (ACCA, 2021) (Appelbaum, Kogan, & Vasarhelyi, 2017) (Cao, Stewart, & Chychyla, 2015) (Earley, 2015) and considering the first bullet point, ADA makes it possible to analyze a larger, or even 100% of the data instead of just a sample, reducing the detection risk for auditors.

  The risk is also reduced when considering the previous bullet point, where we can use the benchmarks to know where to focus the organization's resources – either on the most risker parts of the business to make them less risker or on the part of the business that makes it stand out (Cao, Stewart, & Chychyla, 2015).

Overall, considering that we are in the presence of non-financial data as the scope for this study, we can compare the current practice with the potential future one, meaning that currently we only seem to use big data with non-financial data marginally on audits or used with significant auditor judgment required to interpret, whilst we look forward to being able to develop more tools to run models that give us a predictive analytics perspective to identify business risks and areas of focus to help evaluate and assess going concerned (Earley, 2015).

Challenges

Although there is a high potential for the usage of Data Analytics in the practice of data auditing, there are still a few challenges that make it more difficult to expand the utilization of data analytics techniques.

- Data privacy, confidentiality, and access

  While auditing an organization or a process/activity, the internal auditors have access to sensitive information such as emails, transactions, etc. that can be used for several objectives such as to detect fraudulent employee behavior or to optimize efficiency within one internal process. If external auditors have access to that same data, employees might feel that their privacy is not being respected. This is the reason why while auditing, auditors must make sure that they cooperate with clients and have transparent communication (Yoon, 2015).
- Information overload

  As we know, too little information is bad, but too much information is equally bad as well. If auditors must examine an entire data set of transactions, meaning a larger and more complete scale of data evidence, there is the possibility that many outliers within the information retrieved require the attention of the auditor's resources that are limited (No, 2019). On the other hand, there is also a possibility of the data being noisy which leads to lower reliability (Yoon, 2015). When talking about reliability, another thing that must be taken into consideration is the relevance of the data that is gathered which depends on the auditor himself (PCAOB, 2010).
- Integration of evidence generated with ADA and traditional audit evidence – expertise of the auditors

  Hand in hand with the need for the right hard and software, it is needed the right expertise to be possible to process and analyze the data evidence, meaning a further investment for the companies, that could be a constraint (Cao, Stewart, & Chychyla, 2015) (Earley, 2015). Also, the integrity of the data can be compromised as there are several limitations to its extraction whether is from the company's side or the client's side, making it only possible for the client to only make some data available.

4.1.3 Current stage of usage of ADA

Due to the newness of the topic, as said previously, there are just a few studies regarding the topic of Audit Data Analytics, especially when we search for its appliance to a specific topic such as supply chain. However, we shall look to a study that is quite relevant for this study: Kend and Nguyen conducted a study to gather stakeholders' perspectives on the impact of big data analytics, Artificial Intelligence (AI) and Robotics on the Australian audit and ended up finding that it has generally positive effects (Kend & Nguyen, 2020). Although this study does not examine the actual implementation of ADA within an audit process nor it is applied to a supply chain, it gives a good perspective to shift the focus to other areas and study the possible effects of big data analytics.

The use of ADA is relatively limited, and the use of more advanced ADA is rare (Eilifsen, Kinserdal, & Messier, 2020).

4.1.4 The Value Through Analytics

The Value Through Analytics (VTA), developed by Zoet, is designed to make the concept of data analysis more understandable for auditors, helping in overcoming the knowledge gap by answering three different questions to derive the specific data analysis an auditor wants to perform (Mantelaers & Zoet, 2018).

The model consists of three rings, and it should be read from the inside out (see Appendix A Figure 3). The first ring answers the question: "What do I want to analyze?". The possible answers to this question include a process, such as the purchase to pay; a decision, such as a going concern assessment; or an object, such as customers or revenue. This helps to identify the specific area of focus of the analysis (Mantelaers E. &., 2018).

The second ring answers the question: "Why do I want to analyze it?" There are also three answers to this question that go hand in hand with subchapter 4.2.1 of this dissertation: 1. Discover, 2. Check and 3. Improve. If we decide to go with point number 1 then no analysis or model or dashboards is yet made so the goal is to create a first one. When checking, data is evaluated against a standard or an idealized output, and deviations are reported. The last point of the decision is if we want to improve an already existing model, analysis, or dashboard – process, decision, or object.

This question is based on the idea that there are two types of data analysis (van der Aalst W. , 2011): creating a new analysis, model, or dashboard: checking against a norm: or improving an existing analysis, model, or dashboard. This helps to establish the purpose or goal of the analysis (Mantelaers E. &., 2018).

Further in the model, the third and final ring answers the question: "To what level do I want to analyze it?". This question is based on the idea that there are six types of data meaning six possible answers to this question. This is based on the article of (Leek & Peng, 2015). The authors found that the most frequent failure in data analysis is mistaking the type of question considered and the extent of the analysis needed. The six possible answers are the following: (1) descriptive, (2) explanatory (3) inferential, (4) predictive, (5) causal, and (6) mechanistic (Mantelaers & Zoet, 2018).

The easiest type is descriptive analysis, in which you give a summary of the data without interpreting it (Leek & Peng, 2015) (Mantelaers & Zoet, 2018). The explanatory analysis builds upon the descriptive analysis. In this analysis, you interpret the summary of the data to find correlations and form hypotheses. The inferential analysis consists of testing whether the trends or correlations hold in a different sample. The predictive analysis consists of predicting a value in the future based on data. The fifth type of analysis, the causal data analysis, seeks to measure what happens to one measurement on average if you change another measurement. Causal and mechanistic analysis both consist of analyzing the effect of change of one measurement on the other. However, mechanistic analysis seeks to show that a change in one measurement exclusively and always leads to a certain, deterministic behavior in another (Leek & Peng, 2015) (Mantelaers & Zoet, 2018). Each of these types of analysis builds upon the previous one, so that someone who can perform one, would be able to perform the other types of analysis as well. This helps to determine the level of detail or complexity of the analysis.

With these three rings, the VTA model allows auditors to identify and understand the specific type of data analysis they want to perform (Mantelaers E. &., 2021).

By answering these three questions, the model can help auditors identify the area of focus, the purpose or goal, and the level of detail or complexity of the analysis. The model can also be illustrated in a periodic table, which helps to visualize the capabilities of an organization in a clear and easy-to-understand way (Mantelaers E. &., 2021).

4.2 Process Mining

To study Process Mining we must first look at the concept of Data Mining. Data Mining is “the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner” (Coosmans, Smyth, Lee, Hancock, & Yang, 2009) meaning that the primary aim of data mining is to discover new and unexpected patterns useful for the data owner, taking in consideration that its usage it is not restricted to one field of expertise.

Taking this definition into account, we can now look at the purpose and definition of Process Mining. Process Mining looks to be the bridge between data mining and model-based process modeling but with the use of more advanced techniques to perform the analysis. The strength of the concept comes from making it possible to manage large datasets while producing clear models as they occur (Aalst, 2016). Process Mining techniques can help organizations improve their operational processes by identifying bottlenecks and inefficiencies within their processes as it creates clear process models (Aalst, 2016).

When we refer to the concept of Process Mining, we refer to the discovery, monitorization, and improvement of the real processes – making it a process-centric concept instead of a data-centric one –

by comparing how the processes take place in practice versus how they were meant to operate – fact-based concept.

This becomes very crucial as we start inserting the concept of Process Mining into the audit world as it enables the auditor – internal or external – to understand the unintended outcomes that may arise from the relaxation of enterprise resource planning control settings, which is done to manage unexpected situations – for example within the supply chain. This way, the concept can add value when applied to auditing: 1. By analyzing the entire dataset instead of a sample, giving a more extended and comprehensive view of the process audited; 2. The data audited is entered independently of the auditee's actions, reducing the risk of bias; 3. It enables analysis that is not possible through traditional audit tools as it looks for the practical part.

The final output of the Process Mining algorithm would be a model of a "process" of how a certain operation is happening in the organization. As the process is executed, it deals with a lot of inputs and outputs of information that then are stored within the information system used by the company. The record of all the operations that are performed, we named it as "log files" (thesis applicability). We can represent the iterations among the main components involved in process mining in the image below. To begin with, we look at the incarnation part in the diagram that pertains to the information system that enables the practical and real implementation of the process within the organization. This system maintains a record of all the operations stored in event logs.

An event log serves as the base and starting point of process mining, making the purpose of the concept to get knowledge from these event logs that are recorded within the software system. As they are the base of process mining, their quality will highly influence the quality of the outcome of the process as well – and depending on their attributes, the researcher/auditor can drive conclusions regarding the whole process (Van Der Aalst W. A., 2021).

The attributes are taken from several data analysis techniques and algorithms to make it possible for the researcher/auditor to discover, understand, monitor, and improve the targeted processes (Van Der Aalst W. A., 2021)). We can conclude that the event logs are, the collection of events generated and captured and information systems in several circumstances (Reinkemeyer, 2020) referring to a certain activity in a certain time – meta-data - and case.

Then, these event logs are the prerequisite for conducting the process mining analysis by relating the event logs to the analytical model of the process. After the identification of the event logs, according to (Reinkemeyer, 2020) we can extract and visualize the real process flow.

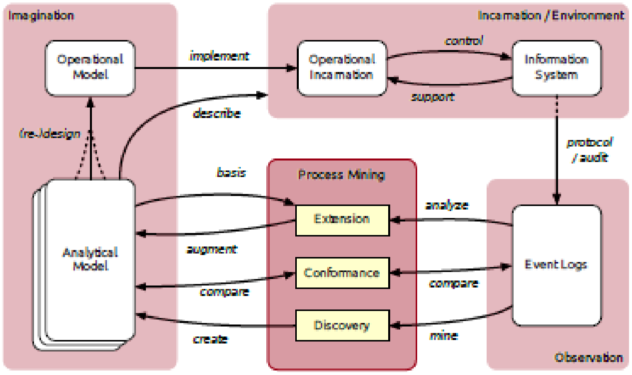

Figure 9 - Representation of the three main perspectives of Process Mining (Akçetin, Celik, & Yaldir, 2016)

Hand in hand with the three possible ways to perform the mining, there are three perspectives that we can take: process discovery, process conformance, and process enhancement, or extension.

4.2.1 Process Mining Techniques – Process Mining as an audit tool

We are now aware that the strength of Process Mining comes from the ability to produce clear and real, based on the organization's real practice, models based on large datasets of information. However, we also must be aware that using process mining does not make our process better automatically, as it is only the starting point for further analysis. Process Mining techniques refer to a set of methods/algorithms used to discover, analyze, and improve business processes based on event log data. These will leverage the quality of the data captured in the event logs to have insights into the process behavior, performance, and compliance.

The first step for a company that has not formally modeled its processes is taking an event log as input to produce a model without additional information. Then the conformance step will give a great insight into the models and bottlenecks that need to be adjusted as we have the theoretical model we can align it with the reality, getting as an outcome the deviations and the adjustments needed. Finally, we can take the deviations as a learning opportunity and improve the process model (Aalst, 2016).

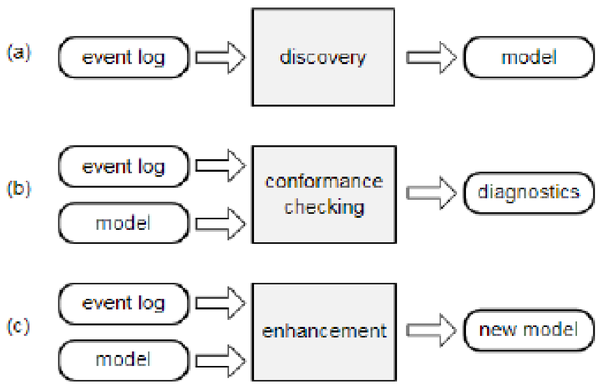

Figure 10 - Three Types of Process Mining (van der Aalst W. , 2011)

4.2.1.1 Process Discovery

The purpose of process discovery is to acquire knowledge from the recorded event logs. This is possible as we can find patterns within the recorded event logs and draw a process model that comprehensively explains and captures the behavior in the data collected, without any prior information, solely based on findings based on the practical process of the organization.

This way, it is possible to understand the practical and actual situation of the process instead of the one idealized. This brings the benefit of accelerating the processes' understanding in an organization with complex processes (Aalst, 2016).

4.2.1.2 Process Conformance

The second technique involves comparing the previous process model to an even log associated with the same process. This makes it possible to detect any discrepancies between the real process and the modeled process, meaning an event log and an a priori constructed model as an input (Van Der Aalst W. A., 2021). We measure the output using four metrics: fitness – ratio of traces that a discovered process model can reproduce from the event log (Dunzer, 2019)-simplicity – complexity of a model (Real, Pimentel, Oliveira, Braga, & Stiubiener, 2020) -, precision – how accurately the model represents the pattern captured in the event log (Munoz-Gama & Carmona, 2011)- and generalization – degree of abstraction (Dunzer, 2019).

4.2.1.3 Process Enhancement / Extension

Based on the discrepancies found previously, we can improve or expand the existing process model based on the information got from the event logs of the actual process from the first step (Aalst, 2016). As an output we are looking to get an updated model based on two types of possible improvements: 1. Change the model to better reflect the actual process; 2. Expand the model for it to include new perspectives that are correlated to the event log (Aalst, 2016).

When done, we hope to get an optimized process that is closer to the idealized one through constant improvement based on the bottlenecks found and different attributes of the event logs. (van der Aalst, 2016).

4.2.2 Guiding Principles

The IEEE Task Force on Process Mining released a manifesto describing the guiding principles and the challenges that engage in Process Mining. This manifesto aimed to increase process mining's visibility as a new tool for improving the design, control, and support of operational business processes. (Van Der Aalst W. A., 2021).

Six guiding principles prevent users from making mistakes when applying process mining in their processes.

GP1. Event data should be given equal importance and priority. Events should be reliable, ensuring that they occurred, and their attributes are accurate. The event logs should be comprehensive, leaving no room for missing events within a specific scope. Each recorded event should have a clear and well-defined meaning. Furthermore, privacy and security concerns should be addressed when recording event data.

GP2. Extraction of logs should be guided by specific questions. Without specific inquiries, extracting valuable event data becomes challenging. Consider the vast number of tables in a database, such as an enterprise resource planning system like SAP. Without guiding questions, it becomes difficult to determine the starting point.

GP3.Support for concurrency, choice, and other fundamental control-flow structures is necessary. Basic workflow patterns supported by mainstream languages (e.g., Business Process Modeling Notation, event-driven process chains, Petri nets, Business Process Execution Language, UML activity diagrams) include sequence, parallel routing (AND splits/joins), choice (XOR splits/joins), and loops. Process mining techniques should encompass these patterns.

GP4. Events should be associated with elements in the model. Conformance checking and enhancement heavily rely on the relationship between elements in the model and events in the log. This relationship facilitates the "replay" of the event log on the model. Using replay, discrepancies between the event log and the model can be identified (e.g., events in the log that are incompatible with the model), and the model can be enriched with additional information extracted from the log (e.g., identifying bottlenecks using the event log's timestamps).

GP5. Models should be regarded as purposeful abstractions reflecting reality. A model derived from event data offers a purposeful perspective on reality, capturing the behavior depicted in the event log. Considering a single event log, multiple views may prove valuable.

GP6. Process mining should be viewed as an ongoing and continuous process. Given the dynamic nature of processes, treating process mining as a one-time activity is ill-advised. The objective should be to infuse vitality into process models, encouraging users and analysts to engage with them daily.

4.2.3 Further Types of Process Mining

This section wants to give a few examples of how Process mining can be used, and evolved, together with other areas of expertise to take better results out of a process that already exists within today's organizations, especially when looking at the predictive analysis of data. As it is out of the scope of this thesis, there will not be an elaborative description of the examples.

Root Cause Analysis (RCA) and Predictive Process Monitoring are other examples of new techniques. These, together with process mining and other tools, such as Machine Learning, enable the prediction of future flows and operational decision support (Tax, Verenich, La Rosa, & Dumas, 2016). Process Mining can also be used to identify process automation opportunities, which overlap with the scientific field of Robotic Process Automation (RPA). This RPA consists in a new technology that can automate repetitive tasks by replacing the human workforce in front-end repetitive tasks with an automated system (Van der Aalst, 2020).

4.2.4 The Four Main Quality Dimensions

Is also important to take into consideration that when comparing observed and modeled behavior, we typically consider four main quality dimensions (van der Aalst W. M., 2011).

- Recall (also called replay fitness): the discovered model should allow for the behavior seen in the event log. This can be quantified by the minimal number of edit operations needed to make all traces in the event log fit into the model (or simply the fraction of perfectly fitting traces). A model has perfect fitness if all traces in the log can be replayed by the model from beginning to end.
- Precision: the discovered model should not allow for behavior completely unrelated to what was seen in the event log. This can be quantified by the number of possible continuations in the model never observed in the event log.
- Generalization: the discovered model should generalize the example behavior seen in the event log to avoid overfitting This can only be tested on “fresh unseen” event data. To evaluate a process discovery algorithm, standard cross-validation can be used to detect overfitting problems (although this is less clear when evaluating a process model rather than a discovery algorithm). Overfitting happens when we have a very specific model being generated whereas it is obvious that the log holds example behavior.
- Simplicity: the discovered model should be as simple as possible. The complexity of a model is defined by the number of nodes and arcs in the underlying graph.

Fitness and Simplicity alone are not adequate, meaning that they are necessary but not sufficient. We can say then, based on this, that a model is only precise when it does not allow “too much” behavior, not falling into being considered “underfitting”, which happens when a model over-generalizes the example behavior in the log.

The challenge starts when we try to balance the four quality criteria, as an event log generally contains only a fraction of the possible behavior. Moreover, we know as a fact that one is typically primarily interested in frequent behavior and not all possible behavior.

4.2.5 Modelling: Techniques and Practices, Tools, and Algorithms

Creating a process model is the main objective of process science and represents the outcome of process discovery, as stated in subsection 4.2.1. of this paper. Process models graphically represent the flow of a process, with the design in Process Mining derived from the event log (Greco, Guzzo, Pontieri, & Saccà, 2006). These models are simplified representations of reality.

The main principle of creating a model revolves around tracing the pathway of a case or a sequence of activities (Cook & Wolf, 1998). Regardless of the representation used, understanding the logic behind the model's creation is crucial for proper interpretation and utilization.

BPMN

In Business Process Model and Notation (BPMN), various symbols represent different nodes and flows. Rectangles denote activity nodes, circles symbolize event nodes, rhombus symbols represent control nodes, and arrows illustrate flows. These unidirectional arrows connect decision points and activities, signifying the direction followed by process instances.

Decision points can incorporate different logical operators. If multiple options are defined, the produced token awaits the completion of all activities. Tokens represent the state of an instance in the process. These decision points, also known as gateways, are of two types: split or join, depending on token behavior. To enrich the model with additional information, information artifacts like data objects and data stores can be used. Text annotations may be added to improve model readability.

When modeling a process, it's vital to minimize handovers, which are points in the process where the instance waits for an activity outside of its control, causing a waste of time. Also, an effective model should prioritize readability, often by maintaining simplicity and smoothness.

Petri Nets

Petri Nets, established is a representation in process modeling, influencing Business Process Model and Notation (BPMN) and forming the base for various other process model visualizations. For instance, Celonis employs Petri Nets in its root for conformance checking.

In Petri Nets, the notation comprises squares and circles. Squares symbolize transactions, while circles depict states. Constructing a Petri Net involves either a bottom-up or top-down approach (Aalst, 2016). The bottom-up method, such as the alpha algorithm, incorporates all transactions into the net before attempting to make the relationships between them. On the other hand, the top-down approach is used and represented by the inductive mining approach (which Celonis uses), which yields a Process Tree as the output.

Process Tress

Process Trees, a product of the inductive mining approach, aim to create more structured models. This approach involves successive cuts to the event log, based on specific criteria.

4.2.5.1 Process Mining Tools and Algorithms

Process Mining tools and algorithms are crucial components of process mining once they help extract valuable insights, identify patterns and make help make informed decisions for process optimization.

Process mining software tools analyze event logs to uncover hidden patterns, anomalies, and business processes. These encompass process discovery, design, analysis, and recovery functionalities while providing recommendations for process design, improvement, and development. When talking about process mining software we must mention ProM (open source), By using a process mining software tool, we should be able to get from it the necessary features to implement priority use cases and processes, while being able to process commonly used event logs in selecting software. This way it is guaranteed that it does not increase the user's workload to obtain insights.

And when speaking about algorithms, they can be classified into three main categories:

- Deterministic algorithms: the output is constant and produces repeatable models by also accepting all types of variable data as an input (Dadik, Stefanović, Sladojevic, & Vučković, 2019);
- Heuristic Algorithms used to get a solution based on insights when the algorithm cannot conclude (Dadik, Stefanović, Sladojevic, & Vučković, 2019);
- Genetic algorithms: Powerful and complex approach that yields many models by combining features and adding random variations in operations conducted with an arbitrary starting point (Mahendrawathi, Arsad, Astuti, Kusumawardani, & Utami, 2018).

In the process of Discovery, we can have two situations: the first one is when the pattern recorded by the input event log is simple and noise free. Then basic algorithms such as alpha miner can output accurate and effortless process models. However, when the complexity increases, the quality of the data can get worse quite quickly as they often refer to data unstructured, incomplete unreliable repetitions, and infrequent traces. In these cases, although there is no perfect automated process discovery technique, we often can use the Alpha algorithm, the region-based approach, inductive mining techniques, and the split miner.

On coherence checking, there are two different approaches that one can take: token-based replay or alignments. By employing techniques like token-based replay and alignments, event data can be related to a process model, allowing for the annotation of the process model with frequency and time information, thus facilitating analysis and improvement. On the first one, the process model is represented as a Petri net, and traces in the event log are replayed on the model. If the trace indicates that an activity needs to take place, the corresponding transition is executed. If this is not possible because an input place is empty, a so-called missing token is added. The numbers of missing and remaining tokens relative to the numbers of consumed and produced tokens indicate the severity of the conformance problem. This is efficient but does not always produce valid paths through the process model.

The (Costa, 2022) paper, compares the results of the application of several available process mining technologies. Firstly, the author compares the process mining tools – Python, ProM, and PowerBI and concluded that ProM verified most features under evaluation, although it is not a tool that is easily accessible for beginners. As the scope of the paper is to study three different algorithms – Alpha Algorithm, Heuristic Miner Algorithm, and Inductive Miner Algorithm – and they only can be applied in two different tools, the author ended up focusing on Python and ProM.

Ultimately, the Alpha Miner Algorithm was not recommended as it did not perform well with real-life logs, having as an output a model with separate activities – meaning that it had no links between activities.

The Heuristic Miner Algorithm allows the specification of certain parameters, such as a threshold for the dependency factor, which indicates the relationship between different activities in a process – either a positive dependency or a negative dependency. According to (Costa, 2022), the algorithm generally performs well, generating process models that represent the event logs. However, some errors may arise as some process models fail to include all events from the data and certain models contain isolated activities that are disconnected from the workflow. The issue arises due to the algorithm's difficulty in handling low-frequency data. An issue that does not exist if more data is incorporated.

The Inductive Miner Algorithm differs from the Heuristic Miner one as it does not require the specification of a dependency threshold. Additionally, the models generated provide a more comprehensive representation of each process's behavior as it includes all the activities in the model, offering a more precise temporal understanding compared to the previous algorithm.

Concerning the algorithms studied – Alpha Miner, Heuristic Miner, and Inductive Miner – the one that showed the best results was the Inductive Miner algorithm as it did not exclude low-frequency activities. Alignments are often seen as the gold standard for conformance checking because they provide paths through the process model that are as close to the observed behavior as possible.

4.2.6 Process Mining in Supply Chains Today and Challenges

Process Mining, a concept pioneered by Wil van der Aalst, has gained widespread recognition in academic circles since the beginning of the century. Over the following decade, businesses started adopting Process Mining to increase transparency and derive insights from their actual processes for discovery, understanding, and enhancement.

Companies such as such as BMW, Siemens, and Uber, have implemented Process Mining across their entire value chain, in areas such as Procurement and Order Management where it has been utilized to examine order processing and identify issues like duplicate payments, deviations in payment terms, and unauthorized purchasing. Logistic Experts employ Process Mining to investigate causes of delayed deliveries, bolster supply chain resilience, and ensure punctual order fulfillment. Order Managers utilize Process Mining to reveal inefficiencies in customer order processing, address problems arising from rework, and enhance customer satisfaction. Plant Managers rely on Process Mining to pinpoint bottlenecks in manufacturing processes, perform value stream analysis, and enhance operational efficiency. Sustainability Managers employ Process Mining to identify the operational root causes of waste and the CO2 impact of individual process decisions.

In each of these cases, Process Mining exposes the actual process flows, allowing human intelligence to interpret the transparency it provides, determine the causes of process inefficiencies, and translate these findings into business value.

Nevertheless, since process mining is an emerging discipline, it has its own set of challenges, particularly in data preparation and processing. There are difficulties with event data such as non-process-oriented, incomplete, scattered, noisy, and mismatched timestamps. Moreover, this data can change due to alterations in process models or information systems. Infrequent behaviors or patterns that differ from the mainstream behavior; absence of negative examples in event logs in the process discovery alongside the frequent incompleteness of observed behaviors compared to the entire range of possibilities, delayed case completion, recurring quality problems, excessive rework, make it challenging to use process mining in every single flow of a supply chain

In addressing these challenges, the IEEE Task Force on Process Mining has released a manifesto that outlines the guiding principles and challenges of Process Mining, including the issues encountered when it is used in Supply Chain Management (SCM). The manifesto highlights its potential applicability within the studied industry despite these challenges.

Challenges	Characteristics
Findings, merging, and cleaning event data	Data mining can be distributed over several sources, with the possibility of being incomplete, having outliers, or different levels of granularity
Complex event logs with diverse characteristics	Event logs can have different characteristics making it either difficult to handle them or not reliable to drive any conclusions

Creating representative benchmarks	There is a need to have good benchmarks to be possible to compare various process mining tools and algorithms
Dealing with concept drift	Need to understand that the process might change as it is being analyzed
Improving the representation bias used for process discovery	To ensure high-quality process mining results, it is needed a careful and refined selection of representational bias
Balancing between quality criteria	There are 4 competing quality dimensions. It is difficult to find models that do well in all four
Cross-organizational mining	Typically process mining techniques consider one event log in one organization which does not reflect anymore the total reality
Providing operational support	Process mining can provide both offline and online operational support through different activities: detect, predict, and recommend
Combining process mining with other types of analysis	The challenge is to combine automated process mining techniques with other analysis approaches to extract more insights from the event data
Improve usability for nonexperts	The challenge is to hide sophisticated process mining algorithms behind user-friendly interfaces
Improve understandability for nonexperts	Users might have problems understanding the output or drive incorrect conclusions. Process mining results should provide a suitable representation

Table 3 - Process Mining Challenges Identified in the Manifesto (Van Der Aalst W. A., 2021)

The processing of invoices from multiple suppliers is a common back-office task in supply chains that are known for being time-consuming and prone to errors. Kryon Systems, a firm specializing in automation technology, has addressed this issue by developing a solution that integrates robotic process automation software with computer vision and natural language processing (NLP).

This powerful combination allows Kryon's technology to efficiently scan and verify invoice data, even when it is in unstructured formats. With the implementation of Kryon's pioneering technology, businesses can anticipate considerable time savings, accelerate supplier payments, evade penalties for tardy or inaccurate payments, and free up their employees to concentrate on more intricate tasks such as decision-making.

We start seeing the effort from the companies to undergo a digital transformation to introduce digital technologies and further integrate them. However, the implementation of new digital tools and systems requires a certain preparation within the organization. It has influenced supply chain management through the integration of systems designed to optimize operations and activities across an entire enterprise.

To make use of process mining in supply chains, enterprises need to share event logs and transactional data which can come as a challenge as well: Data can be manipulated, jeopardizing its authenticity, and system security may be compromised. Furthermore, data retrieval can be time-consuming. Therefore, to promote supply chain data management, we consider the following guidelines: (1) enhance the coordination and information sharing within SC; (2) protect the data authenticity; (3) speed up the data retrieval.

5 Model and Proposed Framework

5.1 Assumptions

The purpose of this thesis came about while debating the issue of the appearance of innovative technologies that are being applied in other industries, such as healthcare and finance quite extensively but not in supply chains, especially global supply chains that have a vast quantity of data available.

The question was how we can embrace the new tools and techniques while auditing the available data in supply chains in a manner that makes it possible to find the optimal process/flow of supply chain information based on the data, processes provided, today's company's needs and the level of maturity of the business. Since this new age of digital information is characterized by being faster, we are able to provide more insights and drive better and faster decisions, allowing the reduction of costs, a much better competitive position in the market, and a higher level of automation through the efficient mapping of the process and the smart use of information.

The advantages inherent in building this architecture are greater productivity, efficiency, and effectiveness of supply chain processes and data flow management through the combination of Process Mining techniques in Audit data analytics in a supply chain process. This can be achieved because Process Mining serves as a bridge between data mining and model-based process modeling, meaning being able to handle large datasets, that characterize today's supply chain's data flow, while producing clear models, by identifying actual bottlenecks within the process and inefficiencies and possible future risks.

Based on what has been studied in the literature review the exact information we have for this analysis is:

- Implementing and maturing processes in one supply chain is an ongoing process that requires constant iteration and evaluation to appropriately reflect the rapid development of new technologies and knowledge.
- Each company's attitude towards the improvement framework's work has a significant impact on their intention to apply it and therefore on the overall success of the application of the techniques. A collective effort of all employees is indispensable. Hence, the awareness component should be highly promoted.
- To succeed in the improvement of the supply chain's processes and to get the maximum benefits, new capabilities and investments are needed. This requires organizational change, which should be actively managed and guided.

- A maturity model is a tool to assess the current level of maturity as well as provide guidelines to determine targets for each indicator while improving performance.
- At the time of writing, there is no complete framework that combines process mining techniques and algorithms with the auditing of big data to improve the supply chain's processes.
- Each company that uses this framework should have the necessary expertise to adapt it to its own needs and situation. The process mining algorithms and the techniques used to resolve bottlenecks are out of the scope of this work and should be defined by each company's needs and available tools and resources.
- The companies that apply this framework have already some knowledge about analytics and computer science and are ready to invest further in the improvement of the supply chain processes.
- The scope of the framework is solely supply chain processes and the processes need to have been at least planned, monitored, performed, and controlled on an individual level. This means that the data/processes retrieved are not poorly controlled or unpredictable.

5.2 Implementation Framework

Despite the continuous growth and importance, there are advantages to the use of new process mining techniques and innovative ways to audit data. Supply chains can still be a relatively new and complex research area yet to be thoroughly explored. Chapter 4 provided an overview of the concepts of process mining and its techniques, the concepts around audit data analytics, and compared them to the existing models that are already applied within the industry, including drawbacks and opportunities in their design and usability.

None of the current tools and techniques were fully integrated into a framework that satisfies the need for an appropriate step-by-step view that integrates such new and crucial areas that drive so many benefits in the long-term perspective for the supply chain of global companies, independently of its industry of operation. Additionally, none of them considers the pre-implementation stage of the organization and the scoped process ahead of massive changes while allowing to make recommendations for future improvements based on the event logs and the collection of data acquired, providing the space for continuous improvement even if new techniques come to be.

To develop a holistic framework, the key insights from the literature review will be combined with the analyzed environment needed to provide a useful framework set for practitioners.

To accommodate all findings from the literature review in only one holistic framework, it is useful to distinguish between three main phases: the pre-implementation phase (A), the implementation phase (B), and the continuous improvement phase (C).

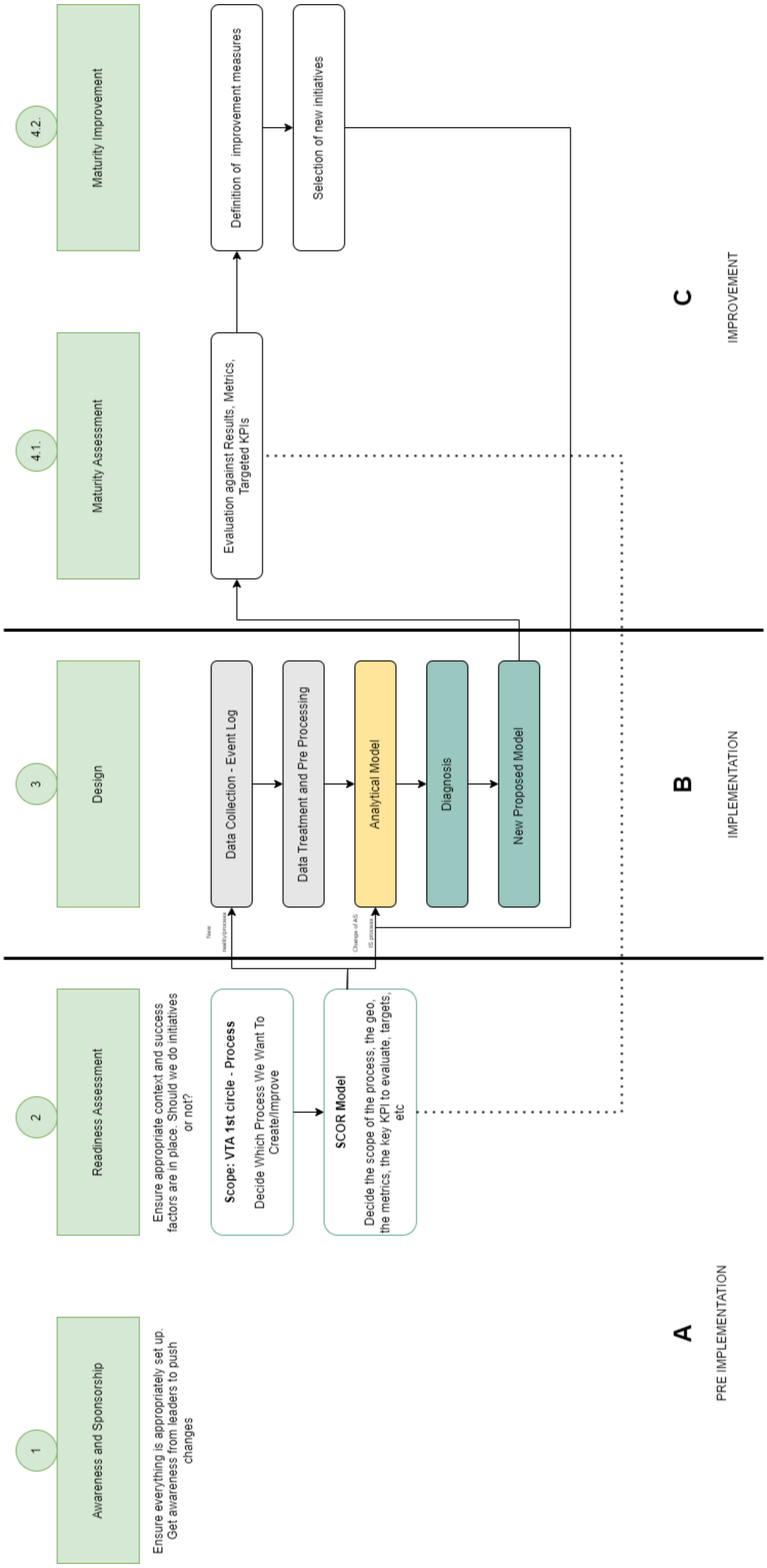

Figure 11 - Proposed Implementation Framework (own illustration)

The pre-implementation phase is when it is made sure that there is an appropriate environment for the successful adoption of initiatives. The goal is to satisfy all success factors and implementation requirements in a way that is ensured that the appropriate techniques and methods are chosen based on the reality that we are looking forward to improving, making it possible to achieve the maximum number of benefits in their implementation. The actual implementation phase, or phase B, takes place when the strategy is translated into practice – this translation is explained in more detail in the next subsection.

The last phase includes all the post-implementation steps, ensuring the continuous improvement and maturity of the use of data in supply chains. This is done by assessing the maturity level based on the reflection of the actual performance and measuring the results/the targeted process against defined goals to identify improvement opportunities.

These phases are summarized in Figure 11 which shows the first proposal for the framework for the improved model that illustrates how a company can use data to improve their processes within their supply chain.

The main goal of this framework is to outline a standard process that describes the phases and order of steps that need to be performed to answer this thesis problem and use process mining techniques in the process of collecting and analyzing even logs within a supply chain successfully and in a sustainable matter, which can be further translated into four main steps when looking into a high level, less detailed perspective. These four steps start with setting up the success factors for a good implementation of a new framework with an organization that moves the needle in several activities and end with the continuous assessment and improvement of all initiatives that target the scoped process.

From Chapter 3 of this paper, we understand a few challenges that are the reason for some drawbacks when auditing data in supply chains. For the framework, we can highlight the ones that we are targeting that fall into the SCM challenges – the coordination between activities and information, the management of the resources to drive a sustainable SC, and the lack of transparency and information sharing that delay data retrieval that comes with it affecting every stage of the logistics network. Audit data challenges include data privacy, access, information overload, and the need for the right expertise to be possible to process and analyze the data evidence which could mean further investment. On the process mining side, we are facing challenges such as the event data being non-process oriented, incomplete, noisy, and mismatched, resulting in data quality issues and making it harder to apply on the supply chains. Hand in hand with this, we are also facing the lack of the right expertise, skills, and tools that are user-friendly for nonexperts to be able to interpret the process mining results with suitable and easy-to-read representations/visualizations.

5.2.1 The Pre-Implementation Phase (A)

The implementation and investment of new techniques and tools integrated into a new framework can affect the entire organization and requires careful consideration of all stakeholders to reduce any resistance and increase awareness and involvement. Therefore, the first step is concerned with the needed awareness and sponsorship from the stakeholders, making sure that the leaders are ready to push forward for changes. This step, together with the second step, has crucial importance for the successful implementation of the framework, as before enacting any significant changes within a large-scale corporation, it is imperative to contemplate numerous factors. It becomes important to evaluate if the necessary competencies, proficiency, and momentum are readily available within the organization, as these elements lay the foundation for the successful implementation of the proposed initiatives. Thus, these initial deliberations serve as the cornerstone of our preparatory steps.

During these introspective steps, it is essential to thoroughly examine the current state of the organization, considering its capabilities, limitations, strengths, and vulnerabilities. It is crucial to involve stakeholders in this dialogue, soliciting their views and insights, as their buy-in can significantly influence the success of this initiative. Key questions that warrant consideration include: Is the organization well-positioned to embark on this transformative journey at present? Do we possess the necessary drive and momentum to propel change forward at this moment? Are our teams equipped with the required skill set, or should we consider expanding our talent pool? Do we have the essential tools, resources, and technologies to facilitate a smooth transition?

At this point, it becomes imperative to delineate the roles, accountabilities, responsibilities, and decision-making authorities necessary for the task at hand. Ensuring an effective governance structure rooted in senior management is crucial, and the organization must determine who will shoulder the responsibility for the execution and consistent monitoring of the entire process, engaging all parties involved.

This decision might be involved by factors such as the size and structure of the company. Alignments of all roles and governance structures are critical in fostering a conducive environment for cross-functional collaboration between business and IT units.

The final factor to consider in this stage is the necessity for cultural and organizational change that prioritizes the adoption of innovative techniques to optimize data usage within supply chains. Regular updates and continuous engagement with stakeholders are key to fostering awareness, securing buy-in, and promoting active participation. This approach boosts change initiatives, justifies necessary investments, and provides updates on progress, thereby enhancing motivation and participation.

The forthcoming table outlines a proposed communication strategy catering to the needs of various stakeholders, which can be adjusted as required. It should be augmented with specific communication timeframes and needs for all stakeholders. These awareness and communication plans should emphasize the importance and benefits of process mining techniques in supply chain optimization, maintaining a steady flow of communication throughout the journey, and celebrating achievements along the way.

Stakeholder(s)	Level of Detail	Suggested Communication
IT/Data department Supply chain parties affected by the change	High	Regular updates on initiatives with a specific timeframe (e.g., weekly) Updates on goals achievement Updates on challenges and opportunities Updates on new practices
Employees Management	Medium	Updates on goals achievement Updates on challenges and opportunities
Customers Shareholders	Low	Updates on high-level goal achievements

Table 4 - Proposed Communication

Before the actual implementation begins, all involved stakeholders must be thoroughly informed and educated about the planned initiatives and their purposes. In addition to setting up effective communication channels for ensuring awareness and transparency, fostering knowledge sharing, training, and collaboration are fundamental steps toward successful implementation. With enhanced knowledge, stakeholders are more likely to drive for innovation, exhibit a willingness to participate, and apply process mining solutions effectively.

Additionally, it is vital to acknowledge that although the envisioned changes offer potential benefits, they demand substantial investments, both in terms of finances and time. It is not merely about immediate expenditure but also about committing to long-term investment, entailing sustained financial, operational, technical, and human resources. This awareness, coupled with readiness for the investment, forms an integral part of the preparatory phase.

Once we have ensured that all stakeholders are adequately informed, engaged, educated, and equipped with the necessary skills, we can start on the actual implementation of the proposed design. However, it is crucial to approach these preliminary considerations with the diligence they deserve, as they ultimately shape the success to achieve the end goal.

We are now ready to deep dive into the second step of the framework. For the scope of this thesis, our focus will be primarily centered on supply chain processes. These processes, existing ones, or the creation of new ones, are derived from event logs that provide invaluable insights into operational patterns. Within this designated scope, we'll identify and select specific company processes that we aim to improve with the help of process mining techniques. Scoping is particularly important in the context of large corporations where it is impractical to overhaul processes across all geographical locations simultaneously.

We will employ the SCOR model. This model enables us to dissect the supply chain into ideal business processes and categories, facilitating a comprehensive cross-company analysis of all information, financial, and product flows within the value chain. Consequently, based on our data analysis, we can plan long-term, medium-term, and short-term while coordinating and comparing processes between suppliers, manufacturers, and customers, thereby augmenting efficiency and effectiveness.

Implementing the SCOR Model's ideology empowers us to precisely define the scope of the process we seek to improve, decide on the geographical location of the company (especially for global entities), and establish the metrics and KPIs that we will evaluate post-improvement. These pre-defined factors will serve as our basis of comparison moving forward.

The aim is to advance your process to a higher maturity level by leveraging new technologies. By reassessing your KPIs post-implementation, you can evaluate the improvement in the process's maturity level. This comprehensive approach ensures that the process enhancement aligns with the broader business goals while fostering efficiency and effectiveness.

5.2.2 The Implementation Phase (B)

Should the assessment yield a favorable outcome, we can then proceed to the design phase or the third step of the framework. This phase centers on the integration of process mining techniques within our supply chain, aiming to enhance process efficiency.

Step 3 unfolds in three distinctive parts, each corresponding to one of the three categories of process mining techniques, as identified in the literature review: process discovery, conformance checking, and process enhancement.

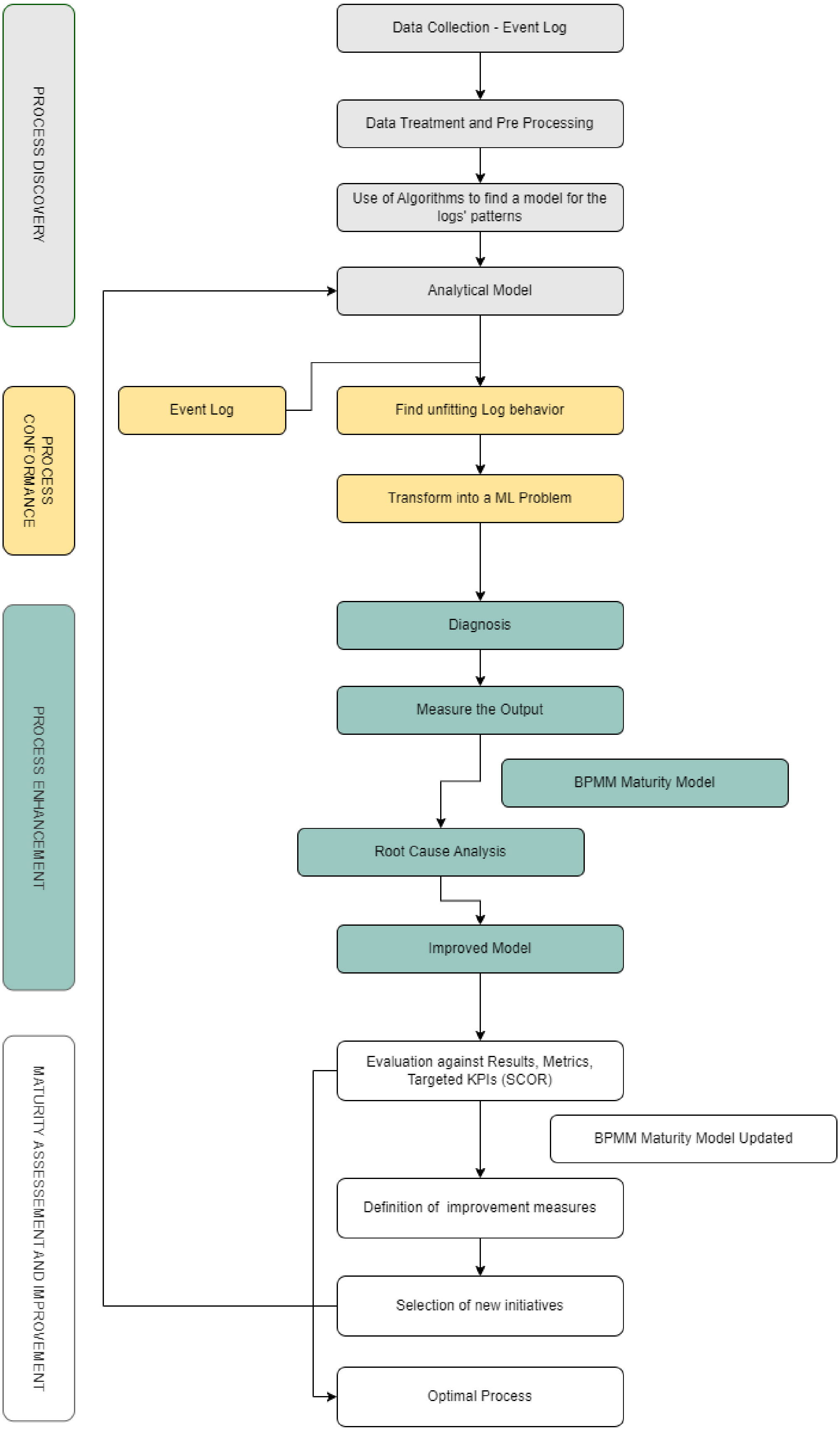

Figure 12 - Proposed Framework for The Implementation Phase

Process Discovery

The process discovery stage primarily targets new processes. Given that it is a new process, data collection begins with the event logs from the previously defined process in step 2. In today's supply chain environment, one of the pressing challenges is information overload. Additionally, event log data is often non-process-oriented, incomplete, noisy, and mismatched. Therefore, the subsequent stage in process discovery involves pre-processing and treating the data. The importance of this step cannot be overstated, as the quality of the event logs data significantly influences the outcome of the process.

The task of data treatment can be undertaken by auditors, stakeholders, analysts, and individuals responsible for data management. They may employ algorithms or machine learning techniques such as alpha, heuristic, and inductive mining algorithms. While this thesis does not have a scope for the details of these algorithms, several studies discuss their applications within big data event logs. The most recent study, explained in the literature review in the subsection, evaluated three algorithms and based on the results defined that the best suitable algorithm was the Inductive Miner as it did not exclude low-frequency activities.

The goal is to generate an output that offers a visual representation of the process operation based on historical data, or in other words, the AS-IS process within the defined scope. We denominate this as the Analytical Model.

Process Conformance

In the second part of the implementation phase, we will be focusing on the process conformance technique – the yellow part in Figure 12. As stated above in the literature review chapter, the process conformance technique the first step is to compare the previous process model, the analytical model that we have as an output of the process discovery technique, to an event log associated with the same process, through the use of conformance checking algorithms and performance analysis algorithms.

This makes it possible to detect any discrepancies between the real process and the modeled process, meaning an event log and an a priori constructed model as input. Through this step, we can find unfitted log behavior, meaning that we can identify any new event log that does not follow the same behavior/pattern as the analytical model. It is assumed that each event log observed and used for a comparison with the analytical model has been preprocessed ahead of comparison, meaning that the quality and consistency of the log are assured. This may include removing duplicates, handling missing data, or even transforming the logs into a suitable format for the analysis.

To measure how well a process model describes the observed behavior, different quality dimensions are used, which are shown by measuring the output using four metrics. These four metrics are explained in the subsection of the literature review chapter 4.2.4. By doing this, we make sure that the event logs are corresponding to the activities within the process model and that there is a connection between observed behavior and the expected behavior described by the model. It is also assured that all the activities in the process model are present in the event logs and vice versa. It shall also be identified any missing or unobserved activity that deviates from the expected and modeled behavior. This includes the analysis of deviations of any kind and the investigation of the reasons behind the detected deviations. This investigation forms our Machine Learning (ML) problem.

The conclusion of the process conformance technique is the collection of deviations seen between the observed behavior and the modeled behavior that will constitute the final diagnosis ahead of the process enhancement technique.

Process Enhancement

The last process mining technique is process enhancement, which takes the discrepancies found previously and uses them to improve or expand the existing process model based on the information got from the event logs. To make this possible we take the prior made diagnosis where it is identified which areas we can improve. This can be done through process mapping, data analysis, and even feedback from stakeholders.

Then, to know that after improvement our model increased its maturity level, we need to measure it. This evaluation can be made through maturity models explained in the literature chapter sub-section 3.2. We can also measure a process' maturity level, although less recommended as they only provide an idea of what the maturity level might be, by using process audits, benchmarking, performance metrics, and expert assessments. Although the overall framework also includes KPIs as a scope, it is more to scope the process and set targets, while identifying at the same time, which steps in the process are deviating from meeting the defined targets.

Based on the information previously gather, the analysts are now in a favorable position to identify the root cause of the problems or the deviations that justify the underlying causes of the issue found, the lack of efficiency and/or effectiveness of a process. By identifying and addressing the root causes, organizations can implement effective corrective and preventive measures and actions to avoid the issue to reoccur in the future.

Knowing that each company is different and there is no pre-defined guideline to implement corrective actions, we assume that each company will apply the corrective actions in the way they fit best while making sure that those bottlenecks that we identified are resolved.

By having those bottlenecks not exist anymore, we find ourselves with a new and improved model that not only does not contemplate the initial bottlenecks anymore but also mirrors the reality of the scoped process and the pattern and behavior of the event logs of that process better.

This model also can demonstrate two different types of improvement: one being the change of a preexisting model to better reflect the actual process or, two, expanding the model to include a new reality that is correlated to event logs.

5.2.3 The Continuous Improvement Phase (C)

The final stage of the framework involves the application of the maturity model that was established during the implementation phase. It facilitates the evaluation of performance in a meaningful way by using the new model generated from the implementation phase.

This new model serves as a standard for evaluating the process's effectiveness and efficiency and is also used for benchmarking against predefined goals. The comparison involves measuring the new model's behavior against previously defined and agreed-upon KPIs, metrics, and goals.

If necessary, data points can be collected to facilitate this comparison. An advantage of evaluating performance based on these predefined measures is the ability to communicate clear expectations and reduce ambiguity in assessment. Common KPIs and measures may vary based on the chosen activity within the supply chain where the process is taking place. These measures can also be employed for benchmarking against competitors, thereby enhancing the company's standing.

Evaluation against goals

Following the application of the improved model, it must be assessed against the goals established in the pre-implementation phase. The use of a maturity model with various dimensions assists in identifying bottlenecks, pinpointing specific dimensions that did not meet the targets, and investigating the reasons behind missed improvements. Additionally, it's crucial to communicate these results to stakeholders in a manner that's comprehensible to the target audience.

Selection of new KPIs, targets, and hypothesis of improvement

If all the goals have not been met, improvement measures should be defined for the ongoing processes in the next implementation cycle. An in-depth investigation into the reasons behind unmet goals is essential to ensure that the improvements yield the desired outcomes. If the causes aren't readily identifiable, it is advisable to reassess the readiness factors, particularly focusing on awareness and attitude dimensions.

Once the implemented processes have been evaluated and enhanced according to the targets and KPIs chosen in the pre-implementation phase (step 2), new targets aligned with the maturity model should be selected. This enables further optimization and improvements in the supply chain processes, incorporating new areas of interest within process mining and auditing that might be studied in the future and have applicability in the scope of this thesis.

Alternatively, a new process for improvement can be selected. In such cases, the framework is considered complete for the current process, and a new cycle begins with the newly targeted process. An appropriately structured maturity model guides the entire process along a path of continuous improvement with well-defined requirements and conditions.

Ideally, the improved model's maturity level should exhibit progress when compared to its state before the framework's implementation. We anticipate an increase in the maturity level of a process after going through the framework, with each identified bottleneck resolved, thereby enhancing the overall maturity, and adding value to the company, the environment, and society.

If the evolution does not meet the company's needs, the evaluation of new KPIs and targets represents an iterative process of selection, deployment, and assessment of the already improved process. This process continues until the maturity level of the process improves, and the process meets the expected KPIs and targets. Therefore, the choice of KPIs or targets forms a new base for comparison when repeating the implementation phase.

5.3 Use Case

This subchapter will illustrate the use of the framework presented in the previous section. To demonstrate the use of the framework, a fictional use case is described. Since the use case is merely fictional and only for demonstration purposes, the content of the tools, algorithms, event logs, and models is only showing a few examples and it is not intended to represent complete and valid inputs. Furthermore, not every step of the proposed framework is documented in the use case with a great level of detail, as that would only be possible in a real application of the framework in a company.

In line with what was researched concerning the design science research methodology, this fictional case should be considered the first iteration to test the applicability of the proposed framework. Furthermore, it presented the next three expert interviews that were conducted that also confirm the validity of the framework.

The picture below shows the group of steps present in a normal, theoretical, supply chain: Procurement, Production, Warehousing, Distribution, Retail, and Customer, each with several activities. The aim is to present several processes where you can get event logs from and the type of event logs that one may encounter. Take note that the group of processes, or the event logs demonstrated below are not complete and do not intend to demonstrate the entire reality of the processes and event logs that can be found in a supply chain nowadays. Its purpose is merely to give a first idea of where an analyst, or the team responsible for the implementation of the framework can get the data collection of the event logs from.

For the fictional use case, let us consider an e-commerce retail company that experiences inefficiencies in the inventory management process. The company has been experiencing frequent stockouts of a popular pair of jeans and they want to understand and identify the root causes and patterns behind these stockouts using the event logs available.

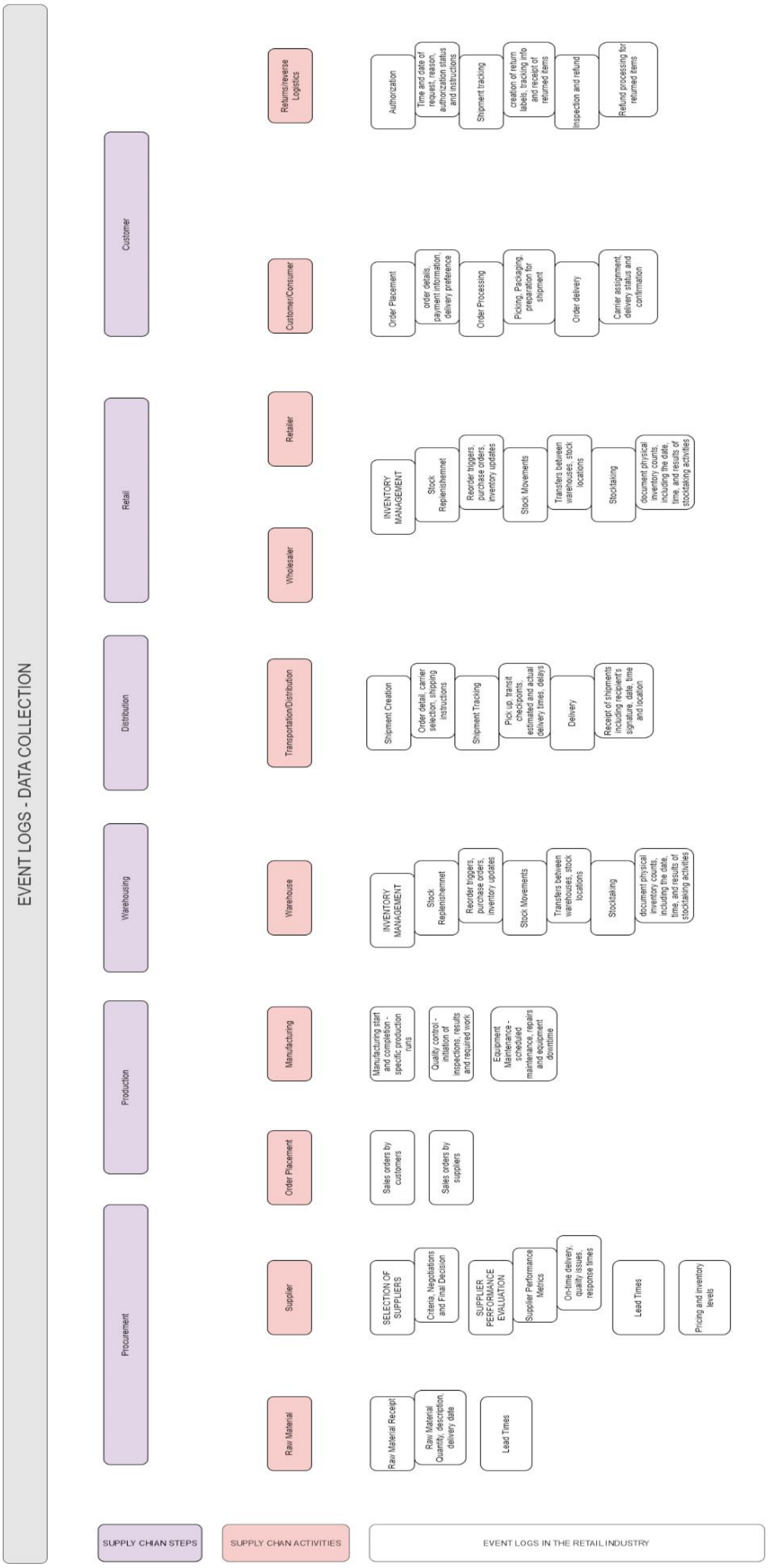

Figure 13 - Sample group of activities, processes, and event logs in a supply chain

The first step the company will follow is to proceed with the data collection of event logs that capture relevant data about inventory management that include information regarding stock updates, deliveries, stockout events, replenishments, reorder triggers, safety stock quantity, et cetera.

Taking these event logs, which are all related to the inventory management processes, the analysts of the company shall proceed to clean, treat, and preprocess the data logs to ensure data quality and consistency before moving into the framework. This may include removing duplicate information, handling missing data, and formatting the data for process mining analysis. The analysts of the company may even understand which process discovery algorithm works best for their data type to get to an analytical model that describes the behavior of the event logs within that process. This is possible as the company is now looking at historical data and is mapping the process AS IS.

Then, we enter the process conformance checking where the aim is to compare the process model with event logs and identify deviations or mismatches. For example, we know from the process discovery model that the replenishment time should be around 5 business days, and from the event logs we are seeing that it is taking 7 business days instead. Theoretically, this can be due to long lead times for order processing, slow supplier response, or inadequate reorder mechanisms. Another discrepancy that we may encounter in the event logs is the insufficient safety stock levels.

When evaluating the maturity level of this process, the company concluded that it had been previously planned, documented, monitored, and controlled, but these efforts were primarily focused on individual process levels without a holistic perspective.

When comparing the observed event logs with the expected behavior, the analysts notice that the root cause problem was set in the safety stocks being too low for a popular with stable demand product in their portfolio. That would lead to a delay in the reorder trigger by one day of business day. Analysts also noticed that from the date in the event logs that the supplier was also having one business day delay that also would compromise the level of available stock in the company warehouse.

With these new insights available, the company was facing a decision-making point: what to do to improve the process and eliminate the bottlenecks found? The stakeholders of the company met and decided with each other that the best approach was to increase the safety stock number to get the reorder trigger sooner, meaning having a considerable amount of stock available at the moment of the reorder trigger. This number was decided based on data analysis of historical data and the demand trend of the product, considering that the demand has been stable for the past few months. This would give more time to reorder the correct amount of stock without compromising the sales.

The second decision was around the supplier. Initially, the supplier was chosen to take into consideration a certain amount of weekly product that, in the meantime with the product becoming a best seller, grew. The supplier no longer could provide the same service to the company with such an increase in inventory asked. The tradeoff to be able to deliver the new increased amount of product was the service provided being delayed by 1 business day. As the supplier could not deliver the demand asked by the company, it was decided that it would exist a change in the supplier for that specific product. The company put together several scenarios with different capacity/cost analyses of different suppliers and took the decision to trade off part of the revenue – meaning choosing a more expensive supplier – to be able to get the inventory necessary to satisfy the total demand on time.

After these two decisions, the analysts of the company were able to improve their previous model to 5 business days, while answering the weekly demand without any stockouts. The overall trade-off for the company was a hit on the overall margin as the cost for the supplier increased.

When measuring against the KPIs, the company saw an improvement in customer satisfaction, conversion, and fill rate while seeing a decrease in the stockout rate and the backorder rate. They also met the targets for the weekly sales and revenue for that product while trading off with the hit on the margin target. The maturity level of the process also improved as we saw an improvement in several KPIs, and the process is now well characterized and understood.

The company also documented the key learning from this exercise and decided to not continue improving this specific process. However, they were able to learn and understand possible future behavior for other best-selling products, making it possible to target other processes' improvements in the future. In this documentation, the company also stated procedures, tools, and standards available to use for future reference, showing an improvement in the maturity level of the process.

5.4 Evaluation

To substantiate the efficacy of the proposed framework, we conducted semi-structured interviews with three distinguished experts, each possessing unique occupational backgrounds but sharing a deep understanding of supply chain processes.

The first expert is a director at the world's largest sports retail company, overseeing consumer inventory. With an educational foundation in industrial engineering and a doctoral degree in logistics management systems, this individual brings substantial practical and theoretical insights to the table. The second expert has 15 years of experience in operations and digital supply chain, having worked in multiple facets of these areas. Currently, leading the digital supply chain for the EMEA region. Lastly, the third expert, serving as the Senior Director in charge of supply and inventory planning for the EMEA region, brings invaluable insights from the world's leading sports retail brand.

The selection of these senior role-holders, each with different backgrounds but a common focus on supply chain management, aids in validating the proposed implementation framework from multiple perspectives within the same area of expertise. This qualitative approach, featuring expert insights, serves to enrich the existing scientific research and address the identified research gaps.

Each interview, conducted individually in June 2023, lasted between 30 to 45 minutes. All participants consented to be recorded, facilitating the transcription of conversations integral to this master's thesis (refer to Appendix A - Appendix C for details). An interview guide, encompassing research goals, the implementation framework, and four guiding questions, was shared with each participant. These questions sought to elicit the experts' viewpoints on the utility of the proposed framework, their critique, their willingness to implement the framework, and their recommendations for further improvements. The purpose of these questions was not just to guide the conversation but also to encourage a critical evaluation of the framework.

Expert 1 drew parallels between our proposed framework and his early professional role, which involved process mapping and event log analysis. His opinion centered on the framework's utility, which is especially evident in scenarios where data is abundant, or event logs are readily accessible. Because the quality of data or the event logs is so widely considered crucial, expert 1 added that a positive readiness assessment in the pre-implementation phase implies that the company possesses data and a certain level of process maturity. If these conditions are met, the framework can be incredibly beneficial.

Expert 2 concurred with Expert 1's positive rating of the framework. She specifically highlighted its relevance in the current digital era and appreciated the inclusion of the sponsorship component, emphasizing that leadership alignment is a significant asset for successful process improvement. Expert 3 agreed with Expert 2's sentiment and highlighted the importance of assessing the impact of process improvements. He regarded the framework as highly beneficial, primarily because it facilitates the use of data to understand the real-time behavior of a process integral to the business but also because it gives the space for learnings ahead of future implementations, retrieving the emotional component by dealing with facts of behavior in supply chain and allowing to use those facts to identify where we should focus our attention as leaders.

Expert 1's reflection on a past project that aimed to expedite supply chain processes for same-day, or next-day delivery further underlined the framework's usefulness. Despite the technological constraints of the time, the project's success highlighted the value of evaluating AS-IS processes, understanding each step, and calculating optimal scenarios.

Despite the many advantages of the proposed framework, the experts also identified potential challenges, risks, and opportunities

Experts 2 and 3 emphasized the importance of incorporating an analytical perspective to understand the potential impact of process improvements – comparing the use of resources, time, and cost with the impact of the improvement. Expert 2 also acknowledged that the implementation required comprehensive knowledge about process mining and data mining techniques within the implementation team. While to Expert 3, one of the potential challenges is data collection, as supply chains often exhibit varying levels of maturity within a single end-to-end process. This may lead to data variants, different tracking measures, diverse metrics, and distinct levels of process maturity, making the inherently fragmented nature of a supply chain a considerable challenge.

Through this discussion, the experts' insights and constructive feedback not only strengthened the proposed framework but also provided essential learnings for its future applications.

To summarize the conducted expert interviews, all experts agreed that the framework, with a special focus on the pre-implementation phase and the data collection step, was very useful, especially if we want to continue driving growth and improvement in supply chains that are each day more complex.

Further, they all agreed that they use the framework, or a big part of it if they were to start a business at this precise moment, demonstrating that even the senior people see and understand the value that auditing data and understanding process can make it easier for companies to drive effectiveness and efficiency on the way they set up their processes in their supply chains.

The key consideration to take for the revised model is the need to include the impact analysis when comparing with the resources, time, and budget used for a certain improvement to make sure that it is justifiable to improve that scoped process. Finally, the last key learning of the interviews was that it took a lot of time to explain the process, presumably because of its holistic nature, which was intended and appreciated. Yet, to understand and apply the framework, a complete case study could be useful, which was recommended by all experts as a further improvement to the framework.

5.5 Revised Model

The feedback from the conducted expert interviews and the proposed framework model present in section 5.2 of this thesis was instrumental in refining the pre-implementation phase and continuous improvement phase. Specifically, it is integrated into it the impact analysis of the decision to improve a process considering by including the cost-benefit analysis and the utilization of the resources available, budget, and time before initiating the process improvement.

The overall framework remained unaltered, with the revised model incorporating this additional analysis at two specific points within the framework. These points occur once during the pre-implementation phase and again during the continuous improvement phase. The goal of having two different moments across the framework is to ensure a rational basis for proceeding with the implementation of process mining techniques or further improving a process, considering various factors that drive change within an organization.

The factors are the cost of change when compared to the effort necessary to drive those same changes, both in terms of resources and budget. If a process or process improvement is not justified by those factors, it may result in ineffective utilization of resources that could otherwise drive more impactful changes with less investment.

The adapted model is shown below in Figure 12. After the two steps explained previously that already took part in the initial proposed framework, the idea is to add a third step to the second point in the readiness assessment in the pre-implementation phase. Meaning that after scoping and choosing the process in which we want to integrate process mining techniques in, and after deciding on the evaluating measures (KPIs and targets), we will perform a cost-benefit analysis, evaluating if the benefit that would come as an output of that process improvement is justifiable according to the investment necessary to be made of the resources available (human resources available, budget available for improvement of the processes, time needed to improve the process).

If the analysis supports the investment of resources because the benefits and impact that it brings are substantial, then we shall proceed to phase 3 of the framework, the design and development phase. If not, the organization should take note because implementing process mining techniques in that specific process is not justifiable at that time. his could be due to various reasons – excessive investment required, lack of necessary expertise for successful implementation, or sustainability concerns in terms of budget or company vision for a long-term process improvement. If a process does not pass this step, it should no longer be a target for the framework, and a more suitable process, with leadership support and full readiness assessment, should be chosen.

The same approach is applied to the continuous improvement phase. As an output of the implementation phase, we have the new improved model, which the company needs to decide whether they want to improve further or not. If they decide to do so, new targets and KPIs are set in place to set the new ceiling of objectives for that process. n the revised framework, we add a step before repeating the implementation phase, involving a cost-benefit analysis to justify the repetition of the design phase and continuous improvement of this process.

If the analysis does not justify the investment of any more resources and the positive impact is expected to be minimal, then we should consider the process complete. However, if the opposite is true, then the implementation phase is repeated. It is important to keep in mind, and as stated in the literature review and the guiding principles of process mining, that process mining should be viewed as an ongoing and continuous process as the objective is to encourage users and analysts to engage with it daily. We should also note that all the models described above are abstractions reflecting reality as they capture the behavior of event logs. However, when considering a single event log, multiple views may prove to be valuable.

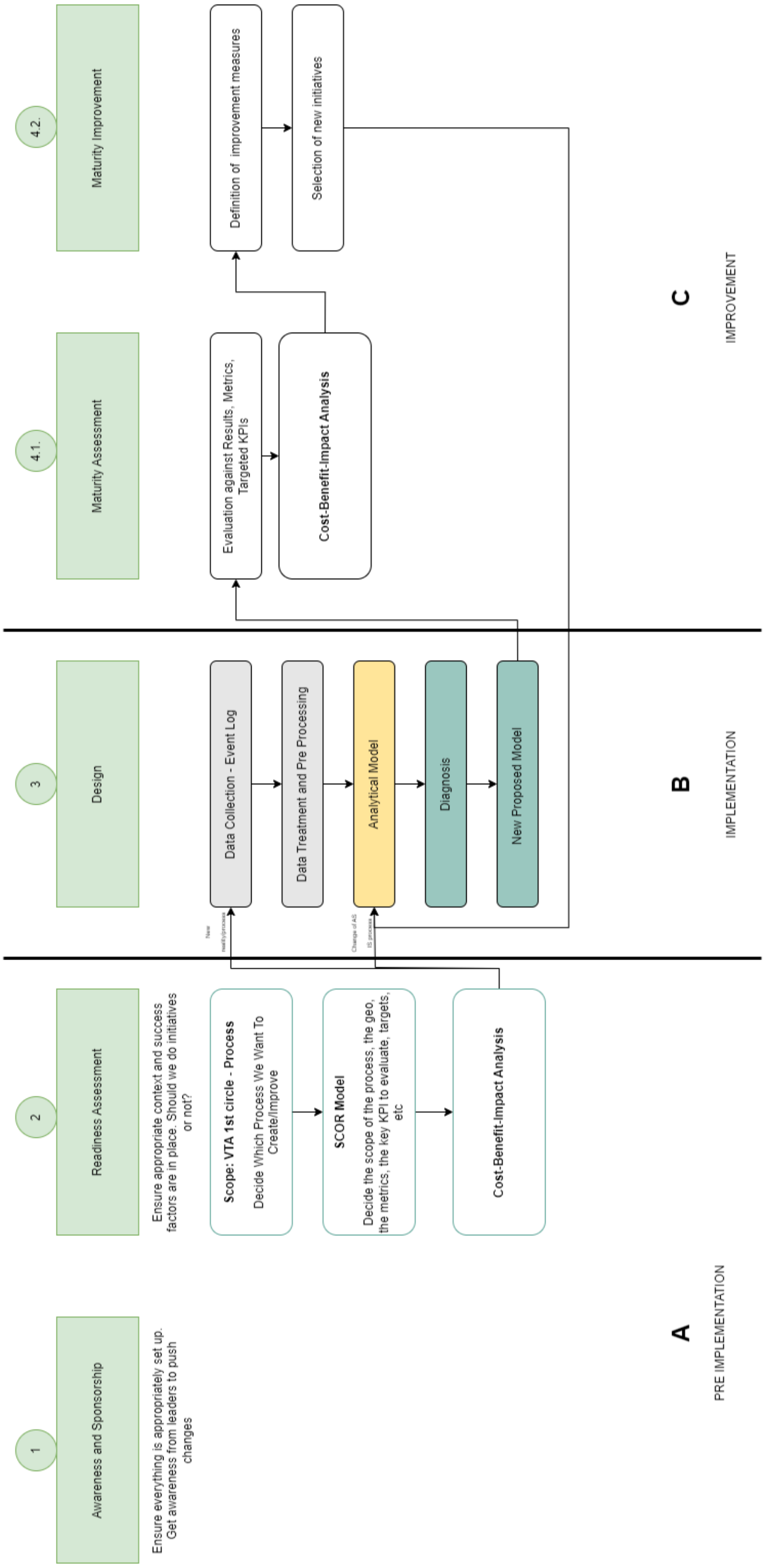

Figure 14 - Revised Model

6 Conclusion

After conducting an extensive analysis of existing literature available it was established that there was not available any comprehensive framework that answered this thesis' problem which was how could we develop a proper model, which embraced a new set of tools and techniques that look forward to auditing the available data in the most efficient way possible and find the optimal supply chain based on the data provided and the maturity level of the business.

This aligned with the need found when evaluating the supply chain environment nowadays and was further supported by the feedback from three experts who validated the initial version of the framework, all of whom deemed it to be highly useful and insightful. Considering the insights gathered during these interviews, the framework underwent a revision to incorporate suggested improvements. The revised framework will now serve its purpose by contributing to the ongoing research in the field.

6.1 Synthesis of the Develop Work

This paper and the proposed framework developed followed a structured design science research process, meaning that several qualitative methods including a literature review, an environment study, a fictional use case study, and expert interviews were chosen to validate the models' designs thoroughly. First, the literature review was conducted to define the knowledge basis that includes the latest research, challenges, and opportunities of Audit Data Analysis and Process Mining.

Based on initial findings, the literature review was further extended by a review of the current environment of the supply chain's processes to establish the current point of implementation of the literature review in a real-world setting. The results of the literature review and environment were used to develop and propose a framework for process mining techniques implementation with a focus on auditing event logs available in supply chains to improve the overall processes. The proposed framework was then validated qualitatively by gathering feedback from three experts within the supply chain industry background. The criticism of these interviews was used to improve the framework and highlighted important points for further improvements in subsequent research.

6.2 Limitations and Recommendations for Future Work

The first limitation of this paper is the lack of previous similar frameworks within the same area of expertise that, although it shows the need for the development of the proposed framework, does not allow any comparison with the work of other researchers, which would be valuable to validate the assumptions and the process established.

Another limitation worth highlighting is that the proposed framework is intended to capture a high-level, holistic view of the implementation of process techniques while auditing data in supply chains. This means that there is less focus on details and practical application of the framework. Because of that, the framework was only revised in a qualitative matter once by the three experts. Due to lack of time, there was no second revision. The quantitate revision was also out of scope as it would require implementation in a real-life scenario.

Nonetheless, several important contributions have been made. For once, the audit data analysis and the process mining subject have been analyzed in the literature review together including challenges and opportunities, providing a comprehensive overview of the current research on the areas that most influence how we improve processes. Furthermore, it has been developed a holistic framework for the implementation of process mining techniques when auditing data in the supply chain's processes to improve them, including a step-by-step approach that can be followed and adaptable to the company resources. This can help guide organizations, increasing the likelihood of successfully improving processes by using new techniques, without losing track of actions or using models in a meaningless order. Researchers should consider this work as a first step which should be tested quantitatively and qualitatively in real use cases, using those findings to further improve and develop this framework, including more detailed recommendations for each step.

Most importantly, and to conclude, this research had the goal to be the bridge between recently discovered areas of expertise that have been growing exponentially over the past few years and one of the areas that drive so much value within an organization.

Bibliography

(CSCMP), C. o. (2023). *What is Supply Chain Management?* Retrieved from https://cscmp.org/CSCMP/Educate/SCM_Definition_and_Solutions/CSCMP/Educate/SCM_Definition_and_Solutions/SCM_Definition.aspx

(IAASB), I. A. (2016). Exploring the growing use of technology in the audit, with a focus on data analytics.

(PwC), P. (2017). Presentation at the PCAOB standing advisory group meeting. Retrieved from https:// pcaobus.org/News/Events/Pages/SAG-meeting-May-2017.aspx

Aalst, W. v. (2016). *Process Mining: Data Science in Action.* Berlin: Springer.

ACCA. (2021). *Data Analytics and the Auditor*. Retrieved from https://www.accaglobal.com/in/en/student/exam-support-resources/professional/exams-study-resources/p7/technical-articles/data-analytics.html

Aelker, J. (2013). Managing complexity in supply chains: A discussion of current approaches on the example of the semiconductor industry. *Forthy Sixth CIRP in SciVerse ScienceDirect*, 79-84.

Aelker, J., Bauernhansl, T., & Ehm, H. (2013). Managing complexity in supply chains: A discussion of current approaches on the example of the semicondutor industry. *Forty Sixth CIRP Conference on Manufacturing Systems* (pp. 79-84). Procedia CIRP 7.

AICPA. (2020). *Audit Evidence AU-C Section 500*.

Akçetin, E., Celik, U., & Yaldir, A. (2016). Analysis of Volvo IT's Closed Problem Management Processes By Using Process Mining Software ProM and Disco. *Alphanumeric Journal, 4(2)*, 171-782.

Akhil, K., & Sharman, G. (1992). We love your product, but where is it? *Sloan Management Review, 33(2)*, 93-99.

Alaoui, I. E., & Gahi, Y. (2019). The Impact of Big Data Quality on Sentiment Analysis Approaches. *Procedia Computer Science, 160*, 803-810.

Alexander, R. M., & Gentry, J. K. (2014). Using social media to report financial results. *Business Horizons, 57(2)*, 161—167.

Alles, M. G. (2015). Drivers of the use and facilitators and obstacles of the evolution of big data by the audit profession. *Accounting Horizon, 29(2)*, 439-449. Retrieved from doi:10.2308/acch-51067

Alles, M., & Gray, G. (2014). *A framework for analyzing the potential role of big data in auditing: A synthesis of the literature (Working Paper).* Rutgers University.

Alles, M., & Gray, G. L. (n.d.). Incorporating big data in audits: Indentifying inhibitors and a research agenda to address those inhibitors. *International Journal of Accounting Information Systems*.

Appelbaum, D. (2016). Securing big data provenance for auditors: The big data provenance black box as reliable evidence. *Journal of Emerging Technologies in Accounting, 13(1)*, 17-36. Retrieved from Journal of Emerging Technologies in Accounting

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. *Auditing, 36(4)*, 1-27. Retrieved from doi:10.2308/ajpt-51684

Banathy, B. H. (1996). *Designing social systems in a changing world.*

Beinhocker, E. (2007). *The Origin of Wealth: The Radical Remaking of Economics and What it Means for Business and Society.* Harvard Business Review Press.

Benabdellah, A. C., Benghabrit, A., Bouhaddou, I., & Zemmouri, E. (2016). Big Data for Supply Chain Management: . *IEEE*.

Beske, P., & Seuring, S. (2014). Putting sustainability into supply chain management. *Suppl. Chain Manag. Int. J.*, 322-331.

Brown-Liburd, H., & Vasarhelyi, M. A. (2015). Big Data and Audit Evidence. *Journal of Emerging Technologies in Accounting, 12(1)*, 1-16. Retrieved from doi:10.2308/jeta-10468

Buijs, J., Van Dongen, B., & Van Der Aalst, W. (2013). Quality Dimensions In Process Discovery: The Importance of Fitness, Precision, Gneralization and Simplicity. *nternational Journal of Cooperative Information Systems*.

Cao, M., Stewart, T., & Chychyla, R. (2015). Big data analytics in financial statement audits. *Accounting Horizons,, 29(2)*, 423-429. Retrieved from doi:10.2308/acch-51068

Capello, A., Lösch, M., & Schmitz, C. (2008). Achieving Top Performance in Supply Chain Management. In F. Budde, *Value Creation: Strategies for the Chemical Industry* (pp. 281-296).

Çetinkaya, S., & Lee, C.-Y. (2000). Stock Replenishment and Shipment Scheduling for Vendor-Managed Inventory Systems. *Management Science, 46(2)*, 217-232.

Chaudhry, S. S., Forst, F. G., & Zydiak, J. L. (1993). Vendor selection with price breaks. *European Journal of Operational Research, 70(1)*, 52-66.

Chen, F. (2001). Market segmentation, advanced demand information, and supply chain performance. *Manufacturing and Service Operations Management, 3*, 53-67.

Chopra, S. (2003). Designing the distribution network in a supply chain. *Transportation Research Part E Logistics and Transportation Review, 39(2)*, 123-140.

Cleven, A., Gubler, P., & Hüner, K. M. (2009). Design alternatives for the evaluation of design science research artifacts. *Proceedings of the 4th International Conference on Design Science Research.* - DESRIST '09, 1.

Collatto, D. C., Dresh, A., Lacerda, D. P., & Bentz, I. G. (2018). Is Action Design Research Indeed Necessary? Analysis and Synergies Between Action Research and Design Science Research. *Systemic Practice and Action Research 31*, 239–267.

Cook, J. E., & Wolf, A. L. (1998). Discovering Models of Software Processes from Event-Based Data. *ACM Transactions on Software Engineering and Methodology*, 215-249.

Coosmans, D., Smyth, C., Lee, I., Hancock, T., & Yang, J. (2009). Unsupervised Data Mining: Introduction. In *Comprehensive Chemometrics* (pp. 559-576). Elsevier.

Correia, I., Carvalho, H., Azevedo, S., & Govindan, K. (2017). Maturity Models in Supply Chain Sustainability. *sustainability*.

Costa, C. J. (2022). Comparing Process Mining Tools and Algorithms. *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*.

Dadik, D., Stefanović, D., Sladojevic, S., & Vučković, T. (2019). Process Mining Possibilities and Challenges: A Case Study. *17th IEEE International Symposium on Intelligent Systems and Informatics.*

Dasgupta, S. (2003). Dasgupta, S. (2003). Multidisciplinary creativity: The case of Herbert A. Simon. Cognitive Science, 687–707. *Cognitive Science*, 683–707.

Delipinar, G. E., & Kocaoglu, B. (2016). Using SCOR Model to Gain Competitive Advantage: A Literature Review. *Procedia - Social and Behavioral Sciences, 226*, 398-406.

Denning, P. J. (1997). A New Social Contractt for Research. *Communications of the ACM (40:2)*, 132-134.

Dickersbach, J. T. (2009). *Supply Chain Management with SAP APO™.* Springer.

Disney, S. M., & Towill, D. R. (2003). Vendor-Managed Inventory and Bullwhip Reduction in a Two-Level Supply Chain. *International Journal of Operations & Production Management, 23(6)*, 625-651.

Divyaranjani, R. (2008). Supply Chain Performance of Customer and Supplier Relationship on Indian Retail Sector. *Int. J Sup. Chain. Mgt (Vol. 7, Issue 2)., Vol. 7, Issue 2*. Retrieved from http://excelingtech.co.uk/

Divyaranjani, R. (2018). Supply Chain Performance of Customer and Supplier Relationship on Indian Retail Sector. *International Journal of Supply Chain Management*, 168-175.

Doty, J. R. (2013). *The Role of the Audit in the Global Economy*. Retrieved from PCAOB: https://pcaobus.org/news-events/speeches/speech-detail/the-role-of-the-audit-in-the-global-economy_455

Dowling, C., & Leech, S. (2007). Audit support systems and decision aids: Current practice and opportunities for future research. *International Journal of Accounting Information Systems, 34(4)*, 75-103. Retrieved from doi:10.2308/HORIZONS-19-121

Dunzer, S. S. (2019). Conformance checking: a state-of-the-art literature review. *Proceedings of the 11th international conference on subject-oriented business process management*, (pp. 1-10).

Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, 493-500.

Education, I. C. (2021, July 8). *IBM Cloud*. Retrieved from Why use Process Mining?: https://www.ibm.com/cloud/blog/why-use-process-mining

Eilifsen, A., Kinserdal, F., & Messier, W. (2020). An exploratory Study into the use of audit data analytics on Audit engagements.

Gartner. (2013). IT glossary: Big data. Retrieved December 1, 2014, from http://www.gartner.com/it-glossary/big-data

Georgiadis, M. C., Tsialis, P., Longinidis, P., & Sofioglou, M. K. (2011). Optimal design of supply chain networks under uncertain transient demand variations. *Omega, 39(3)*, 254-272.

Greco, G., Guzzo, A., Pontieri, L., & Saccà, D. (2006). Discovering Expressive Process Models by Clustering Log Traces. *IEEE Transactions on Knowledge and Data Engineering, 18(8)*, 1010-1027.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly, 3(2)*, 337-355.

Gunasekaran, A. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, 333-347.

Gupta, A., Maana, C. D., & McDonald, C. (2000). Mid-Term Supply Chain Planning under Demand Uncertainty: Customer Demand Satisfaction and Inventory Management. *Computers & Chemical Engineering, 24(12)*, 2613-2621.

Halldorsson, A., & Arlbjorn, J. S. (n.d.). *Research Methodologies in Supply Chain Management - What Do We Know?*

Hameed, W. U. (2017). Enterprise Risk Management (ERM) system: Implementation problem and role of audit effectiveness in Malaysian firms. *Asian Journal of Multidisciplinary Studies*.

Hameed, W. U. (2017). The Role of Distribution Channels and Educational level towards Insurance Awareness among the General Public. *International Journal of Supply Chain Management*, 308-318.

Hazen, B., Boone, C., Ezell, J., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications. *International Journal Production Economics*, 72-80.

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice (Integrated Series in Information Systems, 22, Band 22).*

Hevner, A., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 75–105.

Hidayatno, A., Suzianti, A., & Hartono, M. (2021). A corporate sustainability maturity model for readiness assessment: a three-step development strategy. *International Journal of Productivity and Performance Management, 70(5)*, 1162-1186.

Higgins, A., Beashel, G., & Harrison, A. (2006). Scheduling of brand production and shipping within a sugar supply chain. *Journal of the Operational Research Society, 57(5)*, 490-498.

Hopkins, M. S., LaValle, S., & Balbone, F. (2010, October 01). *10 Insights: A First Look at The New Intelligent Enterprise Survey*. Retrieved from MIT Sloan Management Review: https://sloanreview.mit.edu/article/10-insights-a-first-look-at-the-new-intelligent-enterprise-survey/

Huser, V. (2012). Process Mining: Discovery, Conformance and Enhancement of Business Processes. *Journal of Biomedical Informatics, 42*.

Ishwarappa, & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science, 48*, 319-324.

Jardim, L., Pranto, S., Ruivo, P., & Oliveira, T. (2021). What are the main drivers of Blockchain Adoption within Supply Chain - an exploratory research. *CENTERIS* (pp. 495-502). Procedia Computer Science.

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital tansformation at the intersection of big data analytics and supply chain management. *International Journal of Operations & Production Management, 37*, 10-36.

Kaur, N., & Sood, S. K. (2017). Dynamic resource allocation for big data streams based on data characteristics (5Vs). *International Journal of Network Management, 27(2)*.

Kend, M., & Nguyen, L. A. (2020). Big Data Analytics and Other Emerging Technologies: The Impact on the Australian Audit and Assurance Profession. *Australian Accounting Review, 30(4)*, 269-282. Retrieved from doi:10.1111/auar.12305

Klimko, G. (2001). Knowledge management and maturity models: Building common understanding. *In Proceedings*, (pp. 269-278). Slovenia.

Knoll, D., Waldmann, J., & Reinhart, G. (2019). Developing an internal logistics ontology for process mining. *12th CIRP Conference on Intelligent Computation in Manufacturing Engineering* (pp. 427-432). Munich: Procedia CIRP .

Kohlegger, M., Maier, R., & Thalmann, S. (2009). Understanding maturity models: Results of a structured content. *In Proceedings of the I-KNOW ’09 and ISEMANTICS ’09, Graz*, (pp. pp. 51–60). Austria.

KPMG. (2014). *Going beyond the data: Achieving actionable insights with data and analytics.* Amstelveen, Netherlands: KPMG International Cooperative.

Lamghari, Z. (2022). Process Mining: Definitions and Concepts. *International Journal of Systematic Innovation*.

Landers, T., Cole, M. H., Walker, B., & Kirk, R. W. (2000). The virtual warehousing concept. *Transportation Research Part E Logistics and Transportation Review, 36(2)*, 115-126.

Larson, P. (2001). Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies, David Simchi-Levi Philip Kaminsky Edith Simchi-Levi. *Journal of Business Logistics* , 22(1).

Lee, H. L., & Billington, C. (1992, April 15). *Managing Supply Chain Inventory: Pitfalls and Opportunities*. Retrieved from MITSloan Management Review: https://sloanreview.mit.edu/article/managing-supply-chain-inventory-pitfalls-and-opportunities/

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60(3*, 293-303. Retrieved from https://doi.org/10.1016/j.bushor.2017.01.004

Lee, Y. H., & Kim, S. H. (2002). Production–Distribution Planning in Supply Chain Considering Capacity Constraints. *Computers & Industrial Engineering, 43(1-2)*, 169-190.

Leek, J. T., & Peng, R. D. (2015). What is the question? *Science, 347(6228)*, 1314. Retrieved from doi:10.1126/science.aaa6146

Liang, W.-Y., & Huang, C. (2006). Agent-based demand forecast in multi-echelon supply chain. *Decision Support Systems, 42(1)*, 390-407.

Lnaders, T., Cole, M. H., Walker, B., & Kirk, R. (2000). The virtual warehousing concept. *Transportation Research Part E Logistics and Transportation Review, 36(2)*, 115-126.

Mahendrawathi, E., Arsad, N., Astuti, H. M., Kusumawardani, R. P., & Utami, R. A. (2018). Analysis of production planning in a global manufacturing company with process mining. *Journal of Enterprise Information Management*, 317-337.

Mantelaers, E. &. (2018). *Data-analyse nader geanalyseerd II*.

Mantelaers, E. &. (2021). *Data-analyse (nader) geanalyseerd (III)*. Retrieved from https://www.accountant.nl/vaktechniek/2021/1/data-analyse-nader-geanalyseerd-iii/

Mantelaers, E., & Zoet, M. (2018). *Data-analyse nader geanalyseerd.* Retrieved from https://www.accountant.nl/vaktechniek/2018/11/data-analyse-nader-geanalyseerd/#

March, S. T., & Smith, G. (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems (15:4)*, 251-266.

Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A Design Theory for Systems that Support Emergent Knowledge Processes. *MIS Quarterly*, 179-212.

McCormack, K., Ladeira, M., & Oliveira, M. (2008). Supply chain maturity and performance in Brazi. *Suppl. Chain Manag. Int. J.* , 272-282.

Mentzer, J. T. (2004). *Fundamnetals of Supply Chain Management: Twelve Drivers of Competitive Advantage.* SAGE Publications, Inc.

Mettler, T. (2010). Thinking in Terms of Design Decisions When Developing Maturity Models. . *Int. J. Strateg. Decis. Sci.*, 76-87.

Morash, E., & Clinton, S. (197). THE ROLE OF TRANSPORTATION CAPABILITIES IN INTERNATIONAL SUPPLY CHAIN MANAGEMENT. *Transportation Journal*.

Mula, J., Poler, R., Garcia-Sabater, J. P., & Lario, F. C. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics, 103(1)*, 271-285.

Munoz-Gama, J., & Carmona, J. (2011). Enhancing precision in process conformance: Stability, confidence and severity. *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 184-191). IEEE.

Neuhauser, C. (2004). A maturity model: Does it provide a path for online course design? *J. Interact. Online Learn*, 1-17.

Niiniluoto, I. (1999). Defending Abduction. *Philosophy of Science*.

No, W. G. (2019). Multidimensional Audit Data Selection (MADS): A Framework for Using Data Analytics in the Audit Data Selection Process. *Accounting Horizons, 33(3)*, 127-140. Retrieved from doi:10.2308/acch-52453

Owen, S. H., & Daskin, M. S. (1998). tegic facility location: A review. *European Journal of Operational Research, 111(3)*, 423-447.

Papageorgiou, L. G. (2009). Supply chain optimisation for the process industries: Advances and opportunities. *Computers & Chemical Engineering, 33(12)*, 1931-1938.

Park, Y. B. (2005). An integrated approach for production and distribution planning in supply chain management. *International Journal of Production Research, 43(6)*, 1205-1224.

Paulk, M. C., Weber, C. V., Garcia, S. M., Chrissis, M. B., & Bush, M. (1993). *Key Practices of the Capability Maturity Model, Version 1.1.* Carnegie Mellon University.

PCAOB. (2010). Auditing Standard No. 15 Audit Evidence.

Peffers, K., Rothenberger, M., & Tuunanen, T. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*.

Peffers, 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. Retrieved from https://doi.org/10.2753/MIS0742-1222240302

Pigneur, Y. (2013). *Design Science Research in information systems.*

Power, D. (2005). Supply Chain Management Integration and Implementation. *Supply Chain Management, 10(4)*, 252-263.

Purao, S. (2002). *Design Research in the Technology of Information Systems: Truth or Dare.* Working Paper, Department of Computer Information Systems. Georgia State University.

Pyke, D. F., & Cohen, M. A. (1994). Multiproduct integrated production—distribution systems. *European Journal of Operational Research, 74(1)*, 18-49.

Real, E. M., Pimentel, E. P., Oliveira, L. V., Braga, J. C., & Stiubiener, I. (2020). Educational Process Mining for Verifying Student Learning Paths in an Introductory Programming Course. *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Reefke, H., Ahmed, M., & David Sundaram, D. (2014). Sustainable Supply Chain Management—Decision Making and support: the SSCM Maturity model and system. *Glob. Bus. Rev.*, 1-12.

Reinkemeyer, L. (2020). *Process Mining in Action Principles, Use Cases and Outlook: Principles, Use Cases and Outlook.* Springer.

Röglinger, M., Pöppelbuß, J., & Becker, J. (2012). Maturity models in business process management. *Bus. Process*, 328–346.

Romme, A. G. (2003). Making a difference: Organization as design. *Organization Science*, 558–573.

Santoso, T., Ahmed, S., Goetschalckx, M., & Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research, 167(1)*, 96-115.

Sari, Y., Hidayatno, A., Suzianti, A., Hartono, M., & Susanto, H. (2021). A corporate sustainability maturity model for readiness assessment: A three-step development strategy. *International Journal of Productivity and Performance Management, 70(5)*, 1162-1186.

Scheiter, K., & Gerjets, P. (2007). Learner Control in Hypermedia Environments. *Educational Psychology Review, 19(3)*, 285-307.

Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics, 36(1)*.

Seuring, S., Muller, M., Reiner, G., & Kotzab, H. (n.d.). *Is There a Right Research Design for Your Supply Chain Study?*

Shapiro, J. F. (2004). Challenges of Strategic Supply Chain Planning and Modeling. *Computers & Chemical Engineering*, 855-861.

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2018). *Designing and Managing the Supply Chain.* McGraw-Hill Education.

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science*, 471–480.

Simon, H. A. (1996). *The Sciences of the Artificial (3rd edition).* MIT Press.

Simon, H. A., & Laird, J. E. (2019). *The Sciences of the Artificial: Reissue of the third edition with a new introduction by John Laird.*

Snyder, L. (2004). Facility Location Under Uncertainty: A Review. *IIE Transactions, 38(7)*.

Tax, N., Verenich, I., La Rosa, M., & Dumas, M. (2016). Predictive Business Process Monitoring with LSTM Neural Networks. International Conference on Advanced Information Systems Engineering.

Tesch, J. F. (2016). Discovering the Role of Scenarion Planning as an Evaluation Methodology for Business Models in the Era of the Internet of Things (IOT). *Twenty-Fourth European Conference on Information Systems (ECIS).* İstanbul: ECIS 2016 Proceedings.

Timpe, C. H., & Kallrath, J. (2000). Optimal Planning in Large Multi-Site Production Networks. *uropean Journal of Operational Research, 126(2)*, 422-435.

Tsiakis, P., & Papageorgiou, L. G. (2008). Optimal production allocation and distribution supply chain networks. *International Journal of Production Economics, 111(2)*, 468-483.

Tsiakis, P., Shah, N., & Pantelides, C. (2001). Design of Multi-echelon Supply Chain Networks under Demand Uncertainty. *Industrial & Engineering Chemistry Research, 40(16)*.

Tsichritzis, D. (1998). The Dynamics of Innovation. In *Beyond Calculation: The Next Fifty Years of Computing* (pp. 259-265). New York: Copernicus Books.

Van Der Aalst, M. (2022). Foundations of Process Discovery. In M. Van Der Aalst, *Process Mining Handbook* (pp. 37-75). Springer.

van der Aalst, W. (2011). *Process mining : discovery, conformance and enhancement of business processes. Heidelberg.* Springer.

Van der Aalst, W. (2011). *Process Mining: Discovery, conformance and enhancement of business processes.* Springer.

Van Der Aalst, W. (2014). Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity. *International Journal of Cooperative Information Systems, 23(1)*.

Van der Aalst, W. (2020). On the pareto principle in process mining, task mining, and robotic process automation. (pp. 5-12). Proceedings of the 9th International Conference on Data Science, Technology and Applications.

Van Der Aalst, W. A. (2021). Process Mining Manifesto. *Business Process Management Workshops: BPM 2011 International Workshops* (pp. 169-194). Clermont-Ferrand, France: Springer Berlin Heidelberg.

van der Aalst, W. M. (2011). On the representational bias in process mining. *2011 IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises* (pp. 2-7). IEEE.

Van Der Aalst, W., & Weijters, A. (2006). *Process Mining with the Heuristics Miner-algorithm.*

Venable, J. R. (2006). *The Role of Theory and Theorising in Design Science Research.* CA: DESRIST.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, 36-59.

Wang, Y. (2019). *Designing a Blockchain Enabled Supply Chain.* IFAC PapersOnLine.

Xu, H. (2020). *Minimizing the ripple effect caused by operational risks in a make-to-order supply chain.* Brisbane: UQ Business School.

Yoo, J. S., Hong, S., & Kim, C. O. (2009). Service level management of nonstationary supply chain using direct neural network controller. *Expert Systems with Applications, 36(2)*, 3574-3586.

Yoon, K. H. (2015). Big data as complementary audit evidence. *Accounting Horizons, 29(2)*, 431-438. Retrieved from doi:10.2308/acch-51076

Zacharia, Z. G., Sanders, N. R., & Nix, N. W. (2011). The Emerging Role of the Third-Party Logistics Provider (3PL) as an Orchestrator. *Journal of Business Logistics, 32(1)*, 40-54.

Zald, M. N. (1993). Organization studies as scientific and humanistic enterprise: Toward a reconceptualization of the foundations of the field. *Organization Science*, 513–528.

Zmud, R. (1997, June). "Editor's Comments,". *MIS Quarterly (21:2)*, xxi-xxii.

Appendix

Appendix A- Expert Interview 1

Interview 1 Transcript, Interview Date: May 27, 2023

Beatriz: To start by giving you a little bit of the introduction. So, I'm doing a master thesis on how we can use audit data to improve supply chain processes, especially when we take into consideration process mind techniques. So that's the main focus of this thesis. And for this part, we to go to a specialist on either supply chain or performing some kind of supply chain position in a professional setting, to just look at our framework or model and, say if it is okay, or if this can add any value, or if there are any constraint or challenges here and there. So more in that sense, that's why I asked you here. Maybe to get started, you introduce yourself and then just say what kind of experience you have with supply chain and what you do right now.

Julio Yon Rabbe: Sure, where do I start? I mean, before the company even?

Beatriz: Yes, I mean if you want to, for example, say what you studied and then just step into the company, maybe that's easier.

Julio Yon Rabbe: My background is in industrial engineering. I mean, already there you study some business and supply chain. I have an MBA also in Cardiff University. And I did a doctorate in logistics management systems. So, the supply chain is my thing, I can tell you that!

Beatriz: Perfect.

Julio Yon Rabbe: I think workwise, I've also always been in the supply chain in a way. My first job ever was in retail, and Sears is an American company, and I was there doing…well, it was in the auditing department, which is interesting. But, interestingly enough, it wasn't auditing of finances, but it was our out auditing processes, which is probably very, very interesting here. And I oversaw doing process mapping and understanding the different steps on the processes to make sure that we had an improvement of the inventory health, which is actually kind of what I do now in a way. Making sure that - because it was retailed, there were no processes that would hinder inventory shortages – that things would tally all the time, and therefore inventory was always healthy, and it was also traceable. Then I worked in the telecommunications company and logistics – Huawei Telecommunications the Chinese company. And then, for my doc doctor's degree in logistics, I did some consulting projects as part of my thesis. I was in the steel industry doing some optimization of steel industry processes. And now on the company I've been in the supply chain, mostly inventory management since the beginning, what is it now? Almost eight years. Starting with forecasting of inventory, inventory management, and supply management.

Beatriz: I know. It's perfect for my thesis then. Okay, so let me get started just to introduce how, basically I'm thinking about doing this. So just give you the identification of the problems that I identified and the objectives of my thesis overall. Then I'll walk you a little bit through my design and the framework, I'll keep it a little bit high level just because of time management, but just to give you the overall and then ask you questions for the evaluation part. Starting with a problem, so basically what I saw was with globalization and information systems growing we have this overload of information and data that sometimes you don't know what to do with it at some point. And sometimes it's really difficult to identify the complexity of a process and if that complexity is good or harmful for the company as a whole.

So, people start having a hard time just finding the optimal solutions in some processes, especially when you go into AI and new technologies things, maybe your expertise or the expertise of the organization is not that defined. So that was mainly my problem. When facing all of this and its consequences in terms of increase of risk, decrease of profitability, etc. Another thing is that you have a lot of studies around, you know, data auditing, data mining, supply chain, but you don't have many that have those combined, also because it's a new subject, meaning less than 20 years of studies and interest for the community. A solution would be to do some internal auditing, meaning adding these new techniques into a process that can go into either new event logs or just old ones and say: okay, we have these bottlenecks, this is what we should do based on new models that decrease the overload and the risk associated. So that's the main idea. This is more a theoretical one, especially during the part of the use case, it's just purely theoretical. But the idea is that you can manipulate the process for your own company.

Julio Yon Rabbe: But then what you're defining, and maybe you'll get to it, it's like a process that can apply to different companies?

Beatriz: Yes. More like a framework from the beginning to the end. With that clear then, let’s get to the process. I divided this into three parts as well. The first one is the very overall end-to-end situation. This has its base on the thought that before you're trying to apply anything to one organization, you need to make sure that you have the right conditions to do so. Also, during my literature review, I found that one big challenge is to have the awareness and have stakeholders come and say, okay, this is worth using AI or technology because we'll have a benefit associated. So that was one of the main challenges. Another one was like, you know, information spread across the whole company, how can we get reliable data, how we get the right expertise. All those challenges are included in the first step just to make sure that the company is okay with them, although this is a challenge going forward, giving it a try to implement new technologies within the process.

And if the company is ready and ok taking the step towards the challenge, we go to the second point, which is the decision on the scope. A company has a lot of processes, a lot of objects, and decisions - for my thesis, I am going to scope down to just include the processes, either new ones or already implemented ones. And then within that process, the idea is to apply the SCOR model and decide which process we want to improve/create which geo…Especially if we are talking about a global company you want to start with a smaller scope and then expand. Also, the metrics that we want to evaluate, as well as the KPIs. And then obviously, the targets, because this will be our basis of comparison.

This is all pre-implementation phase. Then we go to the implementation, which is the design. And I'll get to this in more detail in the next slide. But it just goes through the idea that you collect data, you treat it, you got a new model diagnosis, and the new improved model. Having that said, and having the new proposed model, basically suppose on new improvements as well, how you're going to continuously improve that process or decide that this is the optimal one, we are okay with this for the next five years, for example. And you do that by against the results that you have from the new proposal model to compare with the metrics and the target KPIs that decided on point number two.

So if you see that results are more beneficial in terms of the KPIs and metrics that decided on point number two, the question is do you want to continue improving those or are those enough for now? Or do you want to focus on a new process? So, if those are not enough and you want to continue improving that specific process, you decide on new measures or new targets. Followed by a new hypothesis or initiatives that you want to take into those new improvements. So, for example, imagine that you see a new bottleneck - if you want to take care of it then you go back to the design part. On the design part, double clicking on point 3, I sub-divided these into three parts, incorporating a technique for process mining, which is the division between process discovery, conformance, and enhancement.

And it just targets which kind of decisions you make within each part of the process. So, imagine targeting a new process. You get the data collection or the event logs from the process, that you decided on your scope, and then you're going to treat the data. And this part maybe nowadays companies do Excel analysis, or you take three weeks to get that treatment and process of the data. And here I'm proposing machine learning applications such as clustering and classification, to reduce your workload and risk of biases as well. Having that said, and having your data treated - because obviously when you get into very big volumes of data, you, for example, have data that is not there, or duplicate data and you want to treat those as well - then you are going to use algorithms as well to find patterns within that data. And here there are several Python PromM tools in this part of the process, having that said and taking into consideration that we are in a new process that we want to implement, we get our first analytical model of how we should data behave within that process. If that's not the case, this part is already done and set. So, you just want to improve the process that we already have, which is based on this, but prior algorithms. And that brings us to our second part of the process, which is already having the model that you have or the behavior that you want your logs to have.

You can input a new event log in case it's a new process or you want to improve on it. So you just take new event logs and you try to see if they behave the same way as your model. And if they're not, they probably know unfitted log behavior and you want to see what is happening there. So what brings to this part is you transform this into your machine learning problem. This is an outlier. Why is this an outlier? Why is this happening? Is this something that we want to see or not?

Julio Yon Rabbe: So is it correct that, so I have understood you first said sort of a baseline. And then comes a new event, and then you have that baseline benchmark to compare against. Yes. And then that gives you information.

Beatriz: Yes, exactly. Imagine that we are talking about a process in a factory and you're making shoes and this shoe just took three hours longer than the other ones. Is it something because it was like a machine problem? Was it a supplier problem? In which activity of the supply chain is the bottleneck. And that gives us the diagnosis of bottlenecks and the challenges that we're seeing within that process. We going to measure the output of your model and understand that: we have this model; we have these outliers coming or these bottlenecks. How is the process in terms of maturity?

This is important because you have processes that are not mature enough in the sense, they don't use technology whatsoever. You don't treat the data; they are reactive and unpredictable. So, its initial ones and other processes are mature, controlled, and well-implemented. So maybe you don't want to touch so much on those. So, you want to measure - and there're four points of measure based on quantitative and qualitative measures. And this is our first basis of maturity comparison. So when we improve the model, we want to compare it to this step in the process maturity level, as we want to increase the maturity level.

With the diagnosis or the bottlenecks, we can understand what the root cause of the bottleneck is - so the point in the process that we want to take care of and improve it as well. And do that in our organizations today, but we also know that each company is different. So, you cannot put everything in one box, especially when you change industries. For example, the pharmacy industry doesn't treat the same bottlenecks the same way as retail. But the idea here is that you treat your bottlenecks and challenges, and the output is your improved model. Yes. That circles back to point number four. So, we are out of the design part and are now in the improvement part where we going to compare the results of your improved model to the ones that we decided on the point to with the score model.

So, metrics, targets, KPIs, all of that. We again, evaluate the maturity model of the process. So, imagine that we are in the second level of maturity. With your improved model, you want to go from two to three or two to four, depending on of bottlenecks you treat. And then, as explained before, you decide if this is good enough for you or not. So, imagine that you now have a process that now is a maturity level four, while you still have the rest of the company at maturity level two, you may want to take a look at the other process first before going to improve again the process level 4. So, if that is the case that's finished, it's done. But if that's not the case and you still want to improve it to a five then you going to do it and restart the same process again. Does this make sense?

Julio Yon Rabbe: I mean, so far so good. What I'm trying to imagine is an example in real Life and At the current company, which is my closest thing, or maybe not at my current company.

Beatriz: Because in my use case, what I'm going to do is go deep into which area/activities from the supply chain you can get event logs from.

Julio Yon Rabbe: And this framework is what you've designed.

Beatriz: Yes, so the new stuff here is really about algorithms and the steps toward improvement that you might want to use in your process. Not so people-based, but more machine based. So here (slide of event logs) I designed an overall perspective where you can get event logs within the supply chain. So you have the procurement, production, warehousing, distribution, retail, and customer. You have several activities throughout all of that. And for example, if you go to a supplier level, you can get the lead times event logs, you can get supply performance metrics, event logs on time, delivery, quality issues, and response times. So, all these event logs can be a scope for this framework. That's what I'm thinking. So for example, we decide that we are not good with the transfer between warehouses and stock locations. We want to improve this process; we're going to get these event logs and do all of that process with technology.

Julio Yon Rabbe: And then to your point, it could be maturity level zero and then you run it through the process, and you iterate until you elevate the maturity level to what you desire it to be.

Beatriz: Yes. So, in a use case, imagine that you now have a lot of cases in which you think the stock is in location A and it is in location B, it takes a lot of time, the consumer is not happy, you're not happy, the organization is not happy, and you want to implement all of this. And the outcome that you want to have been maybe we have one or two times there is the failure to identify where the stock is, but the majority, where the stock is and what to do with it as well as that the company knows the procedure if something fails because the process is understood for the company and controlled to a certain extend.

Julio Yon Rabbe: But you said that also this is very data-driven and how do you link process and data?

Beatriz: So, in the sense of using the event logs. Everything nowadays is made through machines and standardization. So, if you're making a shoe you know where the shoe starts with the supplier through event logs of the supplier.

Julio Yon Rabbe: Okay, now I'm tracking my industrial engineering background Is helping me here.

Beatriz: You can, for example, go to the supply chain and you have all the dates, the time, where it is, who has it. So, you get those event logs and then you try to improve them or the process that includes them.

Julio Yon Rabbe: Now I am tracking.

Beatriz: We're back on track then. Let me go back to the other one then. So, my use case would be that can go to any kind of these event logs collection, and you just decide which one fits your company the best. And you go from there. We want to see the improvement in the long run of the risk, the biases, the lack of information, all of that. The idea is a company that stands by this framework or something similar can improve its processes, maturity, and efficiency that it has as a company by a lot. So the more processes you apply this, the more efficient your company might be. In an ideal world. Then just for the last part, which is the evaluation, you may foresee these questions already. The first one is if you find it useful? And if not, why not? Or if yes, why?

Julio Yon Rabbe: I think, Yes, and maybe I'm old and outdated because I recognized all your event logs there. And actually, it takes me back to my first job on a team where I remember I associate all of this with process mapping and studies of movement and where you're logging things, and I think I remember something related to this. So, where you're tracking every single event log, and then you trace it back, and then you check in and you find the faults, right? Well, this is an added efficiency because you can see that there are some mismatches here or something happened. And I think it's probably very useful in companies where you have the accessibility of data logs. And if I even tried to apply anything of that at my current company, I mean we have tons of data, but I'm no more thinking with my process mindset. So yes, probably very useful where data is accessible. However, the reality is that you have a lot of companies where data is not accessible already like that or where it's processed or the service industry, for example. And I think that's where probably this sort of framework is a little bit more complicated because I mean, you can still create data logs even manually in the process by documenting things, but I think that's where something like this can have a fall a little bit short.

Julio Yon Rabbe: On the flip side, I mean you have tons of companies that are tons of data fully available. And then, but I can imagine that it's the complicated part of it, especially in the first when you're designing if you have too much data, so what do you take into account? What is successful, and what is not successful? So that's how I would answer the first one.

Beatriz: Yes, for sure. I mean, it's one of the challenges right now within the subject in the areas: you either have a lack of good information or you have an overload of just information, right?

Julio Yon Rabbe: And how do you skim through it? It was in your problem statement.

Beatriz: Yes, exactly. So, one of the points I want to make within the first two steps is that company is ready to incorporate the right expertise, tools, or skills to get that. So for example, you don't want to go to a company that is not data-driven to implement this kind of algorithm because they don't know how to do it, right? And that's why it's so important to take into consideration the stakeholders first and the big guys within the company to get the yes to, okay, technology is important, information is important, let's get the most out of it.

Julio Yon Rabbe: So, there's also a pre, and maybe that's why you said the pre-readiness assessment of whether this can be useful and successful or not.

Beatriz: Yes. Because that's one of the main challenges for sure. And also, for example, when you talk about the share of data within, when you have a global company, you go through the supplier to the carrier, the share of that information within the parties of the supply chain. So that's another challenge as well. And I think obviously that you cannot put everything into a basket. You cannot put startups in the same basket as a global company. So, it's more just providing the framework of if you are ready to take that step forward, this is how you should go with it mostly.

Julio Yon Rabbe: Well then, I think that that's one, and I, in a way, also, I'm addressing two the answer, the question two, because that would be my criticism, but you already captured it.

Beatriz: And I think he's also why has so few studies or so few studies around supply chain and this because if you go to financial auditing or if you go to pharmacy or healthcare industry, the number of studies is completely different. There are a bunch of studies there, but when you go into the supply chain is a completely different story. And I think it's because this part, so information and if the companies are ready to take that step on,

Julio Yon Rabbe: And I mean finance is also just very numbers and precisely driven, I guess, but I see what you mean.

Beatriz: Yes. And number three, would you consider implementing, so maybe not at my current company because it is huge and difficult to implement one of those things, but if you would start up your own company, for example, would you take into consideration steps of this and the use of new technologies such as algorithms or even really the process part?

Julio Yon Rabbe: I think absolutely. Where I'm struggling is, how do I implement algorithms. This is my mind not being very in a, well, I guess you can mean you can optimize everything. You can optimize services, you can optimize flows, you can optimize. And that's what you mean by algorithm, then? Yes, yes. I think maybe my answer is going to be yes, also in support of your studies.

I mean, it's almost like a simple case. I mean, in cases where companies are ready to, they meet the criteria to implement something like that, they're assuming that they have data, there's a certain level of maturity, there has to be a certain level of process. If it's all chaos, then it's also probably not ready. But if you have all of those and you have the data available and you can run it through a process where you can set a baseline of what is acceptable, and then use the data through algorithms to improve it, to make sure that you elevate your maturity in a certain level, in whatever process you, I mean, why not? I think, yes. The only thing is, and maybe I'm answering now, Yes, number four is how do you simplify it?

Beatriz: What do you mean by simplifying?

Julio Yon Rabbe: Well, I mean I've been sitting here listening to you and I can absorb a certain level of information, but in simple terms, how do I guess the implementation of it in practice is maybe the piece that is a hundred percent clear to me. I can picture it, but I can't see it. So I mean, again, to implement it in practice, I mean anything that is academic or that is theoretical, there needs to be a practicality to it. And I'm where I need to think hard about how to Make sense to apply this.

Beatriz: I mean the bigger the company, the harder it is. Even for me, my key, or the thing that made me think about this more clearly was, for example, when you think about the activities in the supply chain individually. So for example, if you go to the carrier part, get the product to the people's house, and imagine that you're taking three days to get to that, and it's, imagine it's because of the location where your product is or the amount of time that the carrier takes it. How can you improve that? So you just take the logs from the carrier, imagine the DHL, that's the carrier here. So imagine when they receive it, where the location event log is, and how much time does it take to process within their warehouse? How much time does it take to get to the client's house?

Was the client there, or was it not? They'd have to come back the next day. So you get all of that data and then you try to improve. So, imagine that maybe that carrier is not good enough for you. You need to buy a new premium service with another carrier to improve your pro process, right? Maybe you need to change the location within the euro parts of your supply chain to match the carrier location as well. Yes. So more in that sense. So, what helped me was to think about the activities individually, as one process. And that's why it's important to scope it before.

Julio Yon Rabbe: Not you're going to solve the world and you just made me think of a project maybe that could have been relevant to this. I remember a few years ago, there was a thing here in one company where it was a maybe project where, I think it was the VP of operation at the time, challenged the organization to have, I'm not sure if it was same day deliveries or next day deliveries for “company”.com, which at the time was possible. I think we were delivering in three days or something. It was probably, I don't know, maybe six years ago. And that was it. That was like, this is the challenge, go do it.

And then the teams then of course started to go through a process where they had to do that. You said, to be able to ship out and deliver something within a range of 24 hours directly to a consumer, you need to start tracing your stamp, right? And then you go to the data and the logs in the steps. Yes. It's like how much does it take to pick a product from the warehouse to pack it, to pick it, to put it in the transportation? Does what the carrier take to deliver, and what does it carrier network look like? What does it cost associated? I mean, all those types through data. And they did something in the end somehow, and they did manage to deliver within 24 hours, but at that time it only lasted for a very small period. Then I think it was probably too expensive. I think nowadays we're delivering until within the next day delivery. So, Yes, I mean, maybe that's an example of how you implement.

Beatriz: Yes, the idea is to reduce the people work in the sense of, for example, if we run an algorithm in Python, for example, you put all the event logs there, and then you see, okay, there's an outlier during this time, in this location for the carrier with this product. And then you try to understand if that product, it was something, onetime thing, or it was a big product, so you know, need a take more time for it,

Julio Yon Rabbe: And you use historical data for that?

Beatriz: You use historical data for that, and you can just also just continue to improve that model based on new data that you receive or new event logs that you receive. So, it's more than all of that tracking and all the model work of that that you do, it's almost automatic, for new and old processes. Yes. And then of course you need, as you were saying, you can get the outcome that it's better than the previous process and increase the maturity level of it. Does that make sense?

Julio Yon Rabbe: Yes.

Beatriz: Nice. Well, that was it for me

Appendix B- Expert Interview 2

Interview 2 Transcript, Interview Date: May 28, 2023

Beatriz: I prepared just a couple of slides just to give you the overall context and also show you the thing that I need from you as well. Thank you for checking the time I will just start off introducing myself and then ask you to introduce yourself. So Yes, IBP intern this year and I'm doing my master thesis at the same time I'm studying for my master's degree, which is in information management but specialize in business intelligence. And I came from a bachelor's in supply chain management. So, I try to combine the two areas for my master thesis and that's why I'm studying the use of Audit data for the improvement of supply chain processes, especially when you look at process mining techniques. And we'll get into that later, but that's a little bit of context, and if you, like to introduce yourself and maybe give a little bit of context on the supply chain part that you have within your career.

Eugenia: So I am Eugenia Martelli. I've been working in company XXX for 15 years and within those 15 years, I have worked in almost every operation role that exists. So, it starts in Laakdal in our logistics center, I did profitability analysis and cost to serve and a lot of warehousing and profitability. So working closely with accounts. I was in supply chain finance for a while. I then moved to "always available" so I was in the responsive team, just really creating automated ways of replenishing our accounts. So, we are saying, hey, always available product should be always available on our shelves, but we as a company do not manage that for our customers. We can do it in the direct channel, but we cannot do it in our accounts. So how do we make sure that we work with them to make sure that the product is on the shelf and not just in the back room or in the warehouse.

Then I moved to customer operations. I did marketplace operations for what back then was Central Europe, so Czech, Slovakia, Slovenia, Hungary, and Croatia. So, a couple of countries there. They're all very different and within the EU, some outside the EU, a lot of complexity in managing that marketplace and those accounts. I then come back and joined the DSM team. So, coming really from downstream and shipping and the day-to-day when to the upstream in the supply chain, really planning and planning three years ahead and just getting perspective on how to learn the strategy through the way were planning. I did the planning and inventory management for our key categories. I've done running, I've done kids, I've done training. I did a global role for about a year and a half. I was in a team that was building the foundations for it, really developing a process, a global process for marriage, and financial planning.

And then after that, I wanted to get back to the business. It just always feels a little bit detached when you're doing this global project and process work, you kind of feel far away from the business. So I move into retail planning. So, I did retail planning for stores and then we had CDA. So then with CDA, I was moved to the central MFP team working on that longer-term planning as well from a certain central perspective and with the lens of consumer planning. And after that, I decided to move back again. You're feeling a little bit far away from the business when you're planning it is 18-month rolling. I felt like I needed to get into the day-to-day and the digital landscape, so that's why I moved to MPO for digital. So I'm in the digital supply chain now and that's what I've been doing for the last year and it's been awesome.

Beatriz: Wow. I mean I think you've been everywhere at this point. No, that's awesome. It's also the scope that I want for my thesis, so Perfect. So maybe we can go into why I am doing the thesis, just in terms of what's the problem here and then we'll go to just the design to give you the overall framework and then I'll ask you 3, 2 questions just on your opinion as a specialist in the supply chain on

Eugenia: The framework. And if we don't have time, Yes, let's just go through these, and then I can also just write some of the questions, and the responses and give them to you through email or something like that.

Beatriz: Sounds good

Eugenia: Yes, because I want you to take the time to give the context as well so that I can give you the proper answer

Beatriz: Okay, perfect. Starting with the problems. So I found that with globalization, like supply chains, especially in global companies are so complex nowadays, and sometimes you don't have the expertise, the skills, or even the tools to deal with that amount of data, especially when you to big, big volumes or sometimes you have the information, you have the expertise, but you don't have the right information and you don't know how to do with how to deal with that. So, it's starting to be complex to identify what is the right complexity for the company or a certain process. It comes with several risks, being less profitable, et cetera. And at the same time, I was searching on my literature review, and I found that you have a lot of studies on auditing data, for example, in financial companies or for the pharmacy industry or healthcare, but you don't have that development for supply chain, especially when you go into more AI stuff, some more new technologies.

So that was my problem. So, we don't have the studies, this is highly valuable because companies are losing money and development because of that. What can we do? Can we design a framework that incubates these techniques into an overall framework that can be applied to a company if they're ready for it? So that's mostly the idea here. So, the problem and then the context of it goes to the design. Then just rather quickly, I divide this into three parts. So, this is the first one, and it is more like a high-level view. You have the pre-implemented phase, then the implementation, which is the framework in depth, and then the continuous improvement, which comes after. And if we double click on the first part, the pre-implementation, we know that before you apply anything to a company, they need to be ready for it.

Even if it goes, it goes into the human resources, they're okay with it and they're ready to take that step forward or even the technical issues that you might have within your process. So that's all included in the A part. So, ensure that everything is aligned, and the stakeholders are okay with it, this may be a big investment right now, but in the long term, this is beneficial for us. And because this goes into a lot of changes within the whole reorganization, especially when you look at companies that include from the supplier to the end customer, so the E2E supply chain. And because it crosses so many changes in so many activities, it's important to have the stakeholders with you at all times on the same page. So that goes onto the first number one part. So have them give your, okay, and Yes, let's go forward with this.

And then it comes to the second point, which is the readiness of it. So, they can be okay with it, but not ready for it in terms of how your process is already designed, for example. And here for the scope of my thesis, I'll just go for the process. I will not go to objects or decisions of any kind. So, we'll just be focusing on the process, either new ones or already implementing the ones that you want to improve. And then basically my first part is do we have the right resources, skills, and expertise? And if yes, which ones do we want to target. So, for example, you want to go into a certain process in your factory. So, we are going to target that one process and then we are going to decide on the geo, because especially I looked at global companies, you cannot do everything at once.

You need to be specific, about the metrics and the KPIs that you want to evaluate. And that will be the base comparison that we'll have when you improve then your process, and obviously where you want to reach. So, the idea here, and I'll talk more about it on point number three, but the idea is you have your processing at a certain maturity level, and by evaluating again your KPIs after improving your model, you can increase your maturity level of the process in terms of new technologies. So that's the idea for the first part, just really make sure that everything is set in place before going to any big investments. And when you do decide to go in big investments is the design three, which I'll talk about in the next slide, so let's keep just over it. But skipping that and you have the new proposed model, everything is okay, then you need to decide, well, did we improve?

So, is this an improved model? So, you're going to compare with the scope that you did on number two. And then based on that, do you want to improve more or you're fine with that by now? So, I imagine you have a company that has processing level two and you improve one to level three. Do you want to continue to improve to number four or you want to continue with the number two ones, if that makes sense. So Yes, based on that decision, and if that's okay, that's the end of the process. But if you want to continue improving, you need to decide on new measures, new targets and, new KPIs that you want to take into consideration, and then you new hypothesis of what you want to do, what is the bottleneck, all of that. So that's overall. And now going to the point number three. And on the design part, which is mainly the specific framework. So basically, that is number three, I'm incorporating different process mining techniques, tools, algorithms, machine learning into this as well. So, all of that in order to automate your process for big volumes of data with less risk. So obviously you can do all of these in Excel with a lot of people, but the idea is to optimize the risk here.

So the first part on the process discovery piece is to collect the event logs first. And this is mainly when you have new processes as well. You go into historical data and understand what you have and what you can do with that. So those are the event logs. So, for example, if you go to a machine, the time that it start to produce something, how long will it take? Then you're going to treat your data. And that's the difficult part. Sometimes when you have big amounts, it's like sometimes it's duplicated, sometimes it's misinformation. So, all of that is going to be treated in this part, and we're going to use machine learning clustering and classification. You can also use other ways to go around this which I go into more depth in my literature review but just to give you an idea.

And with this you can find the patterns and you can create an analytical model, which be your model for that specific process. Now imagine that you either already have the process in mind or already set in place, so it's only just wanted to improve one, or you already went through these four steps, then you are going to take new event logs. You took historical data before, now you want to take a new one and see if it fits your analytical model. So, your pattern until now, if it works perfect, all going well, but that's not the reality. So basically, something will go wrong, and it'll be a outlier for your pattern. So that's when you want to find why is it an outlier? So, what is causing this unfitted log behavior. Again, machine learning techniques, but Yes, you'll get a diagnosis out of this is just an outlier.

So, for example, when you, you're carrying, when you're talking to a carrier and he took longer to deliver a package, was it because it didn't leave the warehouse in time or was it because he had a flat tire along the way? Those kinds of situations. You'll define the bottleneck and challenges in this part, and then you going to measure your process here. Then you're going to take several measurements that there are, and you will classify them. And there's a five-maturity level scheme to define your process. And here this will be what you want to improve. Let's consider the process processes on the level three. You have the data and the process in place, but you do not use statistical and other quantitative techniques. So that's where you have that part. And because you define a bottlenecks and challenges, you can go through the root cause of them and try to improve them.

So, if it was not a flat tire, what they go wrong and how can we improve that? And that you cannot put everything into one box because each company is different. And the idea here is just to provide a general framework, but idea here is to the company then go to evaluate those bottlenecks. Having that improved model because now we have new logs that came in, they were fitted, but now they're part of a new model, and supposedly it translates better the reality. Then you, you're going to evaluate your results, your metrics, and your KPIs. And now it's already the part four of the previous slides. So Yes, this part four. Yes. So, you're going to evaluate this results or the metrics against what you defined before implementing anything, and you're going to measure everything the same way again. And what was the number three before now it's hopefully a number four. So that's the idea. Obviously, it's not so black and white situation, but each company is different. So, each company then will be able to contextualize this into their own environment, and then it's the same conversation. They decide if they want to improve further or not, and how they're going to do that. This is the general framework, the general idea. Now I do have a slide on a couple of processes within the supply chain that I thought you could get data collection or event logs from. So that first one, bucket.

Eugenia: One point that I would like to bring on your slide on the selection criteria of new initiatives. You can also use analytics as well to understand what the impact is. Again, you can choose to if we want to continue improving further, but just the impact that you bring, you move this much the needle. So maybe something to consider there on election criteria, if you want to add something there on what are the tools, the modelling tools that you can use to help you select if, what are those initiatives that move the needle? Because sometimes we just go like, no, we want to keep improving, and then you keep doing this. And then you might be better off just going to another completely different process that is going to help you 10 times more than if you just keep obsessing on this one.

Beatriz: Yes, that's really, good. I didn't think of that. That's

very good. I need to go that I'm thinking about.

Eugenia: Many times, and this is the reality of a company, you decide to improve something and then you're doing everything to improve this. And it gets a point as you say that there is so much juice you can squeeze from one you want, right? So maybe squeezing that extra bit it is not worth your resources vs the impact.

Beatriz: Got you. Great feedback, honestly. And that brings me to the supply chain that I designed. So obviously we have the activities of the supply chain, so procurement, production, warehousing, distribution, all of that. And then what I did was to think about process it that you could find within each one and what kind of event logs you could take out of each activity. So, for example, let's say if we go to the supplier, we can get the lead times of a supplier or we can get the pricing and their own levels of inventory that are going to influence your own eventually levels. If you go to inventory management, we can take the transfers between warehouses and the stock locations in which one.

Within the framework, I thought about all of these processes you can take big data from, and probably these are the principal ones that you want to start improving in your company because it will have the most impact, if that makes sense.

Eugenia: Yes. The one thing that I'm getting hung up, so maybe it's the wrong place to get hung up, but even looking at this high level, and when you're looking at the entire supply chain steps and so on, you're talking about what are the supply chain activities? And you have a noun there instead of if it is an activity, I would expect a verb as you have there, manufacturing warehousing, if it is suppliers, is it fine in the supplier? Is it, what is the activity that you're trying to collect here? So, in procurement, you're selecting the raw material or what is it? Because then that also aligns with if it is about selecting the raw material, then the event log, is that receipt or that build off materials or whatever. If it is selecting a supplier as you have there because you have it as one of your first thing, well, what is my supplier ranking? That would be the first thing that you could look at in data order placement. That's the login of all the orders that you have done in the past. So Yes, it feels a little bit more of a process and not so much.

Beatriz: I understand what you're saying. So, if I want to look, for example, if you go into sap, the tool, right? What I was thinking is the kind of logs that you have in there is what I want to put in here, reflect here in the sense of if you want to go to a carrier, for example, have, when they get the product in their warehouse, when did they ship it? Yes, how long will it take. Yes. So that was more kind of the events log that I Yes, understand. So, I get what you're saying with the verbs.

Eugenia: Again, you go from what are the big steps in supply chain management. You start from manufacturing, you go to warehousing, distribution, and then how does it go back? Right?

Beatriz: Oh, okay. But wait, this kind is like you choose a process you want to, so you don't do everything.

Eugenia: I get it. But just for the sake of how you are visualizing this. Yes. Just think if you're visualizing a process or visualizing activities within the process.

Beatriz: That even if I just take these arrows out, then it's already an improvement in the sense of if I don't have these arrows, it's already just activities.

Eugenia: Exactly, Yes. Because that's not a process. You also have one team doing procurement and really looking at the materials, another completely different team. Well hopefully not completely different looking at vendors and suppliers. And then there's another one that is looking at orders and sales orders. In one company you have, Yes, 15 different teams even within one geo.

Beatriz: Well, the last slide for you was really just a question actually. Well, the last one was just a question, but I think you already answered a few of them. So, the challenges and recommendations, I think we went through a little bit of that. And for me it was just, you can add the one and the two maybe together. If you were to start a company by your own right now, if you would think about using process mining techniques within your own company from the start, because you just know that data is valuable or not. And if you would consider implementing something like this for you?

Eugenia: I think the framework makes sense. Yes, I mean you're taking the right steps. This is something that is very relevant, especially in the digital world when you have tons of data and then you still have to find value in that data, and you don't have a lot of resources. So, I think it's just having a clear step by step, this is what we're doing. And I love that you, that is sponsorship component. Because sometimes half of the battle is really making sure that the organization and the leadership alignment with really this is what we should be doing.

Beatriz: Even with budget

Eugenia: Well, exactly. Because investing on some of the things are little longer term too. And the model is not going to be perfect. It must learn. You must invest on making it better. And there is this black box mentality that when you have a model and you get an output, why should I trust? So, analytics has a lot of that reputation too. So, I think good to have a framework. I think the challenge is to implement the model. I think not everyone truly understands what process mining or data mining looks like and what are the techniques. So, I think there is a whole level of education on some of that that needs to happen because some people think that they're doing it or you're trying to approach, everyone's just trying to approach it in a different way. And sometimes different teams are not also sharing best practices either. The skill sets are not there in the organization.

Beatriz: So that first part is important when you define expertise, skills, budget, resources.

Eugenia: Exactly. How do you make sure that you have the right resources to help through this? Because if you tell me, would you consider implementing, well, sure, if you have that, I'll do it. I'm not sure if I have a time or the skill sets to create the, well for sure not the model, but even to hire the right people either, right? I don't even know whom I need or what I need. Makes sense. So, I feel like that is a big component that in which hiring or in developing your teams, you know, need to start embedding this as the way forward period. There is no not applying a framework. This framework we need to get, if we want to continue squeezing that juice as much as we can, we are going to have to do a little bit more of that. And the question is not if when we get and how fast can we get? Because this is a competitive advantage in any organization. So, I think it's more, as a leader, how do you get ready? How do you get your teams ready? How do you learn who you have to contact, where do you get the expertise to make sure that you're able to start looking at processes like that? And I think younger generations might feel more comfortable with managing data, but I think we all need to learn those skills .

Beatriz: For sure. No, it's incredible that you talk about that, because I interviewed Julio from S&IP yesterday and he told me the same. So, it's really that first bucket to be really, really set in stone to we have this, have that, then we can move forward because it's important, but we can move forward without that.

Eugenia: Yes. So, it is becoming more important, but then I think it's maybe not organizations, it's not that complicated. If you have the kind of skillsets top down, Yes, just go on, follow the steps. But if it's not there, you do need that education

Beatriz: First. Yes, for sure. Well, thank you. That was it from me.

Appendix C- Expert Interview 3

Interview 3 Transcript, Interview Date: May 28, 2023

Ben: I am a Supply Chain leader with experience in transforming, building, and developing high-performing teams, cultures, and capabilities. I have over 20 years' of experience at a leading sports retail company leading Global and Regional teams across Demand Planning, Supply and Inventory Planning, Sales and Operations Planning, Marketplace Operations, and Integrated Business Planning. For the last 6 years, I have been in the role of EMEA Senior Director for Supply and Inventory Planning where I have been a leader, coach, and mentor to a team of more than 50 employees who are accountable for managing supply and inventory in EMEA business. In that time, we have had to address at short notice the supply chain challenges resulting from the COVID pandemic but we have also purposefully and proactively implemented several growth initiatives that are designed to develop and prioritize both the S&I planning processes and tools, as well as the team's skills and well being. Our intention in doing this from a business POV is to enable both the company Direct channel and Marketplace Partners to achieve their retail sales and revenue goals while creating a positive and energizing workplace and culture from our employees' POV.

Beatriz: So, my bachelor was in supply chain management and my master's is in information management specialized in BI. So, I wanted to combine the two for the thesis, and that's why I came up with using audit data to internal audit data to improve our processes, especially when you look into process mining techniques. But we'll get into that. And just rather quickly, I just wanted to give you an overview of how I'm going to subdivide this interview. So, I'm going to take a look at you first on the problem or what was the reason to come up with this thesis, and then I'm going to walk you through a high-level perspective of the framework that I'm proposing, and then just ask you two or three questions.

Ben: That sounds fine. What would you like from me? Just to answer the questions.

Beatriz: Yes. It's more towards the questions which require feedback, constraints that you might see on the application of this if it is valuable, and the framework if you see this happening. If not, why not? Those kinds of things.

Ben: Okay. I will wait then for you to ask me questions and then we'll take it from there.

Beatriz: Yes, Yes. Sounds good.

Ben: And if I'm unclear on something when you're explaining, please, then I will ask for clarification.

Beatriz: Yes, stop me please. Okay, perfect. So going to the problem, basically what I saw is with globalization, our supply chain processes are super complex. You either have a bunch of data that you don't know what to do with it, or you don't have sufficient data or misinformation that you also don't know what to do with it. And obviously there's good complexity in supply chains that can add value through your company, but there is harmful complexity as well, and we want to take a look at that to improve our processes. Obviously when you have a lot of complexity, there's a several of consequences in there such as increase of costs while seeing the decrease of an efficient and effective process.

What I also saw was the lack of studies. You have a lot of studies on audit data in financial companies. You have a lot of audit data in healthcare situations, but not in supply chain. And mostly that's because of tradeoffs that we're going to talk about, but it's still important to apply it with the globalization that we see nowadays. So, lack of studies, high complexity - that was a problem, that was opportunity that I saw. Moving on and to just give a little bit of insight down to the framework. I have a couple of slides. So firstly, we have a three-step framework, which is a very high level one, and we are going to double click on the number three, which is where I double click on algorithms and all of that. But this is the overall one. What is important about this is the three A,B and C parts. So the A one is the pre-implementation of anything you can think of before moving into the implementation, and then you have the continuous improvement or the improvement of maturity levels of the scoped processes. Going into the A phase, before implementing anything on a big company, you probably want to take into consideration if you have the right skills, expertise, momentum for that, so that all of that happens on the first and the second points. So, you're going to look at your company and you're going to take your stakeholders and ask them, do we want to do this now?

Do we have the momentum? Do we have the skills, or do we need to hire the skills? Do we have the tools? Are we ready to make that investment? Because obviously this is a very good thing, but it comes with a very big investment as well, and a long-term one as well. So that's for on the first part. So just make sure that there is this awareness within your leaders that they also probably need to educate people if they go forward. So, if they're ready then we go to the next step. Keep in mind that the scope for my thesis is just to concentrate on processes. So, we are not going for decisions, we're not going for objects. It's just really supply chain processes you can take event logs from. And then within that scope, we are going to then look at the processes of the company and decide which one we want to improve.

Because especially when you go to big companies, you cannot go for all the geos at the same time. So that is the second square, or the second point, which is really defined the scope. So you define the process that you want to take part in the go, and then you also decide the metrics and the KPIs that you want to measure after. And this will be our base of comparison afterwards. So that's all decided before and before you need to tell people, we are going to do this. We're going invest billions of in data and machine learning and all of that for this, and we are going to educate ourselves on this as well.

And if we're ready, then we go to step three, and I'm going to double click after on this. It is basically just the framework on the models and how we improve the model that we have. And then taking consideration that after the step three, we have the new proposed model, we want to make sure if we are in the right stage right now to be okay with this as it is, or do we want to even improve further? So that's 4.2. If we have that decision of like this is not good enough still, we decide on new measures, new KPIs, new targets, and do the whole process again. So now that we click on the number three, there's a technique ongoing, which is a process mining technique, and basically just divides the whole thing in three parts. And that's why I'm trying to do this in here. So, the first one is the grey one, which is a process discovery. And this is mostly for process that are not incubated yet. Because processes that are already incubated and you just want to improve them you probably go already for the yellow part, which we'll talk about later. Considering that it is a new process you start by collecting the data from event logs from your process that you defined prior. Then you're going to treat them: pre-process the data - because when you have big volumes of data, it's a problem to analyze it without treating it first. So, you really want to make sure, and here we use algorithms or machine learning clustering and classification, which is just the BI part of it to make sure that the data that you're working with is actually good enough to drive great results and conclusions. And done and all working very well we have an analytic model of your process. So, what do you expect to happen when you introduce a new event? Log into the expectancy time, for example, a carrier to take from the warehouse to the consumer. So, let's say two working days, the whole process through all those little steps to working days if that makes sense.

Ben: So, collected all of the event logs, and then I'm just repeating it so that if I can say it, then that I've understood it. We take the event logs, and the event logs effectively portray the picture of how the process works because it says event one, there are in the distance between object one going through event one and object one going through event two is two days. So, you know that the process between event one and event two is two days.

Beatriz: Exactly. Basically, designing a pattern of behavior of your logs that you can get from that process.

Ben: And the essence is that you are following multiple objects through the same process, which is your event log,

Beatriz: But it's historical data mostly here. That's why it's still data collection and not introduction of new event logs that might come up in the future.

Ben: Okay. You're just basically portraying the AS IS through data.

Beatriz: Yes, exactly. This is an AS IS. And then when you go to the yellow part is what do you want to be. So, imagine that now based on historical data, you have a process or a model that portrays any new event log behavior, and then you have a new event log that is an outlier of all of this. You need to adjust your model because of that event log. But obviously you need to understand if that event log is an outlier, so it's a one-time thing or sufficient enough to change your whole model. So, for example, just making more actionable, imagine a carrier if it was a flat tire on the drive to the consumer house and because of that the delivery took longer or it was something that was a huge package that we need to actually look into because huge package takes longer to get to the consumer. Does that make sense?

Ben: It does make sense. Yes. Yes. Okay. So, you're looking for to take away any of the natural outliers, which would be unique circumstances that have impacted the event log. Yes. Behavior. What you're looking for is whether there is a pattern of outliers that you can dig into to identify what it is that is causing those outliers so that you can address the root cause of the problem

Beatriz: Yes! Because obviously if you have a new pattern, you probably, like you mentioned, I have a constant pattern in that specific activity within the process, step within the process that just makes your lead time much higher. You probably, it's a bottleneck there. And you want to go into bottlenecks in the first blue box, which is already the final part of number three, which is, okay, we have this diagnosis model, we have these outliers in this bottleneck. So, this is our bottleneck or our challenge within the process. So, we have the new model where we can say, in the activity or the step three of the process, we have a bottleneck or a challenge, but does this overall process is considered mature or not? So, in that conversation of the five maturity levels, where does it stand?

Ben: Help me clarify one thing, because by the end of stage three, you were talking about identifying the behavioral pattern of the process, but in stage two, you're looking for consistent outliers versus individual occurrences.

Beatriz: Yes and no. It's more in the sense of you have a first model and if you get enough unfitted log behaviors or outliers, you start wondering, is it a bottleneck or a challenge that I want to improve?

Ben: Right. Okay. So, you're not actually, what you're looking at is here is my standard process, and for the purpose of your recording, I'm visualizing with my hands the end-to-end process. And I'm saying that the data that you've captured will say we have a standard model, but we are seeing point 3 consistently has data variances that suggest it's a bottleneck. Yes. You're not changing the model per se. What you're saying is the data suggests this is an issue.

Beatriz: And now you're going to target that issue. But obviously you need to define the maturity level of the whole process because you cannot define one activity. So, you define that, and then you're going to look to understand the root cause of that bottleneck that you've identified. And because every company is different, you cannot put a framework on how they're going to take care of that bottleneck. A retail company probably doesn't take care of that the same way as a pharmacy company does or whatever it is. So just for the sake of this, we assume that the company does take care of that bottleneck on its own way, in the sense to the output being the improved model without that bottleneck as it was.

Ben: Okay, got you. So, you're, you're also rating, you're rating the process based on the maturity of the process. And when you say the maturity of the process, my understanding there is that five is a very well established, very mature process. One would be a very new process. Does the model include a reflection of the importance of the process to the business?

Beatriz: That's a very good point because I just interviewed a Eugenia before and she was actually one of her points of including how beneficial it is to go for that process over and over again. So, I need to include that in here, but it's a very good point

Ben: It sounds like Eugenia and I are asking the same thing. Yes. That says how do you identify the impact of solving the problem?

Beatriz: Exactly. She did make that point further on the process, but same point. Yes, okay. Which is

Ben: Perfect. That's fine. Good.

Beatriz: But Yes, if you have a new improved model that already took care of that bottleneck and supposedly, and what it will do is, I'm already going back to the fourth point in this one. So already going to the number four

Ben: Maturity assessment...

Beatriz: You're going to evaluate again on the scope that you've done in point number two. So before implementing all of this. And then what is hoped is when you, again, double cross with the maturity level, your maturity level went up because you already don't have that bottleneck and you used algorithms or machine learning to actual make your process a lot quicker or a lot more automated.

Ben: Okay. So then if I understand what you're saying there, the very fact that you can run this analysis against a process indicates that the process is reasonably mature?

Beatriz: Yes and no. I mean, you have to have some sort of data collection, right? You must have already some kind of way to have event logs recorded. That's why it's so important. The part two, it's one of the most important parts, which is making sure that you're all right to take that step and that investment and making sure that you have the expertise to take care of all of it.

Ben: So then in essence, by definition, if the process passes the readiness assessment, then it is inherently something that is reasonably mature, reasonably well invested and reasonably important because it has the data and the collection of data that would be associated with something that is important.

Beatriz: We don't expect this to be a level one but the number two or three. But then again it goes back to the number four that we were already talked about. I did one thing, which is based on the event blocks, I went to a supply chain. And this is not a process, just like activities, example of activities you can get event logs, because idea here is to be you general, so you can apply this to whatever company you might want. But I put together the main activities of a supply chain. And then for example, if you want to go into inventory management, imagine that you want to take care of the event logs of your inventory and you want to improve that specific process. I give a few examples of process that you might want to get data collection or event logs from to improve that one. And from my use case in the future, I'll just choose one of these and apply to a formal use case. Obviously, I will not go into this, but it's just like this can be applied to all of this, if it makes sense?

Ben: Yes, it does.

Beatriz: Yes, perfect then. So this was my framework. Now let me go to the questions. I mean, it's four questions, but I think we can put them into just two. So, these are the four ones. So, the first two we just considering if you were to start your own startup right now, not startup, but your own company right now, if you would consider implementing the framework or part of the framework. And if not, why not? And if yes, why, and which constraints would you see there?

Ben: Okay, so if I was to start up a company and the company was involved in supply chain, one of the first things that I would want to do would be to measure the process and collect sufficient data that I can understand the behavior of goods, materials, products, or whatever you want to say as they move through the process. Because the framework that you've outlined here, I would find incredibly beneficial. And the reason I would find it beneficial is that it offers us the opportunity using data to understand real time behavior of a process that is integral to any part of the business. It allows us to make recommendations for change, or at least to identify whether there is an opportunity for improvement based on real data. And thirdly, it allows us to improve our understanding of typical issues that a similar process might find. So if I'm starting up a business, for example, I started up in Hilversum some and I expand, I want to go into Amsterdam or I want to go to Antwerp, then I would take the learnings from this framework and say, I'm not going to design the process as I did in Hilversum.

I'm going to upgrade to level two or level three or level four based on what I'm seeing in the framework. Do I consider the proposed framework as useful? Yes, I do. Why do I see it as useful? Because it allows us to take away the emotion and deal with the facts of behavior in supply chain, and it allows us to use those facts to identify where we should focus our attention, which for any business is the most efficient question that you can ask.

Beatriz: Taking consideration that the first and second bullet or parts are really, really set in stone. And what are some challenges that you foresee in this?

Ben: So, as it relates specifically to my current company, we are, our maxims and our behaviors and our approach towards business are driven through speed, agility, disruption, innovation, and an entrepreneurial spirit. So, the strap line for the company is just do it, which means that the rigor that would be required to set up a process with event logs and data collection at the relevant points in order to implement this framework is typically something that we would not do because we would just say, Hey, there is a problem, can we get around it? Yes, we can. We'll go around it. We'll put a band aid over whatever the problem might be, and we'll reroute the process to accommodate. So, the challenges within supply chain are internal, and it's our discipline around process engineering, around process development and data collection. We don't enjoy that sort of stuff. We want to do the more innovative, sexy stuff.

So, I can see that being a challenge within my currnet company. In a more broader and general supply chain. I can see the data collection being a challenge because typically a supply chain will have multiple vendors and multiple locations, therefore, you will have different levels of maturity within your supply chain for an end-to-end process. If your process of anticipating or delivering product from factory to a logistics center, for example, if you want to take something out of a footwear manufacturing facility in China and deliver it to our logistics facility in Antwerp, that will involve the factory. It will involve the consolidator, it will involve the freight forwarder, it will involve the shipping company, it will involve the consolidator, the carrier from the decon to our logistics company, the logistics facility, and then our inbound receiving process. So that's seven or eight individual instances where you can have data variants, you can have different tracking, you can have different metrics, a different level of maturity. And so, the variant nature, the fragmented nature of our supply chain, I can see as being a challenge to putting this type of framework in place.

Beatriz: Makes sense. Especially when you go to bigger companies probably as well. Right? The bigger, the more difficult.

Ben: And, we have a tendency the bigger the company, the more difficult it will be.

Beatriz: Yes, okay. So makes sense. I mean, it makes total sense, and I think that's why I need to highlight, underline everything on that really second part of, put your scope really, really sharp, put your expertise really, really sharp because one point Eugenia made is if she were to start this one, she wouldn't even know who to hire to have the expertise necessary for the modelling part or the algorithms part. So, it's really that part of knowing what to do, the equational part, if that makes sense. Yes. Because the rest just follows through.

Ben: And it's fascinating. The concept of a digital supply chain, which is essentially what we're talking about here, is taking hold, and the concept is accelerating in terms of adoption. There areisy number of pharma or life sciences companies, petrochemicals, who are looking at putting a digital supply chain in place, not because they're seeing pain points, but because they're wanting to use machine learning as an opportunity for improvement where they didn't even know improvement was possible. So rather than looking for the outliers that are late, they're looking for the outliers that are faster and they're starting to say, how can you drive everything from being the lagging majority into being the faster the new majority? So I think you need a high degree of expertise in understanding and interpreting, manipulating and, analyzing big data. There is a huge amount of mathematical knowledge required to understand the different variances. So, if I was hiring for this, I would go straight to one of the technical universities and start looking for anyone who is Engineering, mathematics, data, et cetera, in industrial engineers with a focus on programming or that type of stuff. And that I would love to do it, the idea of having a virtual supply chain with a set of parameters that you can adjust and say, Hey, what if we changed the carrier in our low end facility and we dropped a day off the processing time there, you might suddenly realize that you've created a bottleneck somewhere else because you've suddenly changed the flow of goods and that type of end-to-end understanding and interdependency. I find it fascinating. I would love it.

What I will say now is that too often as a company, we focus on where there are problems, not where there is opportunity. So having this type of model, even if you just get to stage three where you've understood what the typical behavioral pattern might be and where there are pain points, the ability to understand it, the ability to then do a comparison with a best in class digital supply chain that really unlocks opportunities. So, for example, how does the company's supply chain compare with geese who do fresh fruit manufacture in South America and they bring it across into Europe. It's like, hang on a moment. You've got a very time-specific supply chain. We are not running that. We're running kind of a two-to-three-month delivery. Hey, can we get better?

You start to understand the opportunity that onshore manufacturing might be. Typically, our footwear manufacturing facilities are based in Asia, in Thailand, in Vietnam, in locations like that, there is an opportunity cost of moving production to be more market relevant. So, what if we were to work with our manufacturing partners to put a facility in Prague, right? How does that change? What do you gain in terms of speed to market? And then if you can shorten the lead time, you can also go in and start to say, if I can delay my decision by on placing a PO and I can increase my forecast accuracy by 10 to 15%, you look elsewhere in the business, then you can start to say that the benefit in terms of revenue and margin is X. So, the return on investment, I must invest 500 million in building a manufacturing facility and training up the labor inside Prague or close to Prague, but the return on investment is less than eight months because I'm able to capture unfulfilled demand and I'm able to minimize the margin and the revenue loss that I would get from having products sitting in a two to three months project. That's where it gets fascinating for me.

Beatriz: I mean it's incredible. I think what I've tried to do with this one, because there's so little or so few studies around all of this, I tried to do the first step one, but I mean, while I was doing this, I was thinking about the opportunities after this. If people start applying investments into supply, digital supply chains, and my brain was just pinpointing around.

Ben: Do I have any recommendations or suggestions...

Beatriz: Besides the impact situation? Because I noted that one.

Ben: I think the only one that I would get to be, and it's kind of linked to impact, I think the importance of the process that you are undertaking the analysis for in terms of revenue or margin, but also in terms of resource and in terms of the complexity and also the size of the change that would be required. So, if the data shows that there is an opportunity to address the way, using your example, a big parcel flows through our process as opposed to just a simple shoebox, then I would want to understand which process is based on the event logs I would need to look at. And then I would need to understand the cost of running those individual processes. Because if it is as simple as saying we can get an extra four deliveries out of a spur if we live load rather than palletize a product in a warehouse, then I would say that's a no-brainer. But if we say that we can get another four deliveries out of a spur in 24 hours, if we reconfigure the way in which the cartons are coming out of the apparel and the footwear facility so that they are scheduled again, I will say that's an investment that's not worth making. So, the one bit that I'm not getting into yet, and I know that it's not part of your framework, but the one thing I would want to get into would be to understand the cost benefit analysis of any sort of improvement.

Beatriz: No, but that's super okay. Because what I do in my thesis I have my first framework and then I have the evaluation with the specialist, and then I have the improved framework or model. So, I'm definitely including the impact cost slash benefit analysis into this because I think it's a major one for sure.

Ben: Other than that, I think it's great.
Beatriz: Well, that was it as well. Thank you

.

7 Annex

Table 1. Design and Design Science Process Elements from IS and Other Disciplines and Synthesis Elements for a DS Research Methodology in IS

Common design process elements	**Archer [2]**	**Takeda, Veerkamp, Tomiyama, and Yoshikawam [46]**	**Eekels and Roozenburg [14]**	**Nunamaker, Chen, and Purdin [33]**	**Walls, Widmeyer, and El Sawy [55]**	**Rossi and Sein; Cole, Purao, Rossi, and Sein [10, 40]**	**Hevner, March, and Park [20]**
1. Problem identification and motivation	Programming Data collection	Problem enumeration	Analysis	Construct a conceptual framework	Meta-requirements Kernel theories	Identify a need	Important and relevant problems
2. Objectives of a solution			Requirements				Implicit in "relevance"
3. Design and development	Analysis Synthesis Development	Suggestion Development	Synthesis, Tentative design proposals	Develop a system architecture Analyze and design the system. Build the system	Design method Meta design	Build	Iterative search process Artifact
4. Demonstration			Simulation, Conditional prediction	Experiment, observe, and evaluate the system			
5. Evaluation		Confirmatory evaluation	Evaluation, Decision, Definite design		Testable design process/product hypotheses	Evaluate	Evaluate
6. Communication	Communication						Communication

Annex 1 - Design Science Process Elements from IS