NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Exploring the Applications of Process Mining Techniques in the Industry

Emanuele Aldera

Internship Report

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

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

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

EXPLORING THE APPLICATIONS OF PROCESS MINING TECHNIQUES IN THE INDUSTRY

by

Emanuele Aldera

Internship Report presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics.

Supervised by

Pedro Maia Malta

November, 2023

STATEMENT OF INTEGRITY

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

[Lisbon, 21st November 2023]

ACKNOWLEDGEMENTS

I feel the need of acknowledging all the people that made this work possible, starting with the Professor Pedro Maia Malta, whose feedback and constant presence made the work much easier and valuable, and helped me find the concentration and focus in times where I needed it.

Many thanks also to Miriam, Vasco, Mario, Breno, and all the colleagues in Solvay I had the opportunity and luck of encountering during the internship in the fantastic Process Mining team.

My family also played a very important role in achieving this very important goal in my life, so a big thanks to all of you: Mamma, Papá, Ele, Gio, Nonni Carlo e Maria Piera e Nonna Cin Cin. Thanks for being so close to me and supporting me from day one, with the love only a family can give.

Thank you Tere for bringing color and creativity in the world, for being there for me every day, for all the nice memories and moments we shared together, and for always supporting my ideas and choices; you have been a very important part in my life during the last years.

To all my long-time friends which were always there a big thank you! You were always important over the years, and it is an honor having you in my life. In particular I'd like to acknowledge Lonzo, Ette, Vindro, Maik, Saggio, Riccardinho, Umbi, Gionni, Lapaduros&Co, and many more.

A big thanks also to all my university colleagues which shared with me this experience; I'll always remember with a smile all the memories we created.

ABSTRACT

Process Mining is a discipline that has been growing very fast in the last years, and it has been adopted by many companies to have greater visibility on the ongoing processes within the company itself. Process Mining leverages the increasingly intensive adoption and use of ERP systems by companies, using the *digital footprints* created by those systems to implement the necessary data needed to achieve its main three objectives: Discovery, Monitoring and Improving.

As a point of contact between Data Mining and Business Process Management (BPM), Process Mining analyzes a large volume of data to extract insights, patterns, and other useful information within the data, as well as analyzing and improving existing business processes within the organization.

The objective of this internship report is to dissect and describe the usage of Process Mining encountered during this internship within Solvay, a global leading company in the chemistry sector, using a qualitative methodology approach to describe the usage of the Celonis tool, the leading Process Mining software, supplemented by a literature review to provide a comprehensive analysis of the topic.

KEYWORDS

Process Mining; Business Process Management; Data Mining; Solvay; Celonis; Business Process Intelligence

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

PM	Process Mining
BPD	Business Process Diagram
BPM	Business Process Management
BPMN	Business Process Model and Notation
BPR	Business Process Reengineering
DES	Discrete-Event Simulation
DFG	Directly-Follows Graph
GBU	Global Business Unit
SBS	Solvay Business Services
ERP	Enterprise Resource Planning
EMS	Execution Management System
SL	Service Line
CI	Continuous Improvement
IEEE	Institute of Electrical and Electronics Engineers
P2P	Purchase-To-Pay
O2C	Order-To-Cash
AP	Accounts Payable
R2R	Record-To-Report
AR	Accounts Receivable
MMD	Material Master Data
COE	Centre of Excellence
KPI	Key Performance Indicator
IT	Information Technology
DMAIC	Define, Measure, Analyze, Improve, Control
RCA	Root Cause Analysis
RPA	Robotic Process Improvement

PQL	Process Query Language
PO	Purchase Order
PREQ	Purchase Requisition
OM	Order Management
AR	Accounts Receivable

1. INTRODUCTION

In this document, the objective will be conducting an analysis regarding the real cases of Process Mining techniques encountered during the internship carried out in Solvay, starting from October 2022. The analysis will be performed using a qualitative methodology approach, together with the support of a detailed literature review covering the discussed fields.

Process Mining is an approach that has been growing a lot in the last years; the reason behind is the opportunity to have better transparency on the processes within the company. By having this visibility, it is possible to improve those processes to reduce waste, optimize resources utilization, shorten cycle times, increase automation, check processes conformance, and identify the root causes impacting the processes the most.

The contents of this internship report will show how the implementation of Process Mining techniques in Solvay helped the company to improve some of their processes and to create value; in particular, the document will present the solution frameworks for the challenges the company faced. Moreover, those solution frameworks might be scaled to different markets and sectors to overcome the common challenges present in organizations.

1.1. COMPANY OVERVIEW

Solvay Group is public global company, specialized in the chemical and plastics markets, having its headquarters in Brussels, Belgium. It was founded in 1863 by the Belgian chemist Ernest Solvay. It provides solutions for the Aerospace, Automotive, Agriculture & Feed, Batteries, Building, Consumer Goods, Electronics, Food, Green Hydrogen, Healthcare, Industrial Applications, Resources, Environment and Energy markets.

Figure 1 - Solvay Logo

The purpose of the company is “We bond people, ideas and elements to reinvent progress”, while the vision is to “create sustainability shared for all”. Both the vision and the mission reflect one of the key challenges that the chemical market faces: reinventing progress to address industrial, social, and environmental challenges.

Solvay is structured by dividing its businesses into the sectors, presented below; each segment is then divided into Global Business Units (GBUs), that will also be presented within their parent segment.

1. Materials: this segment specializes on providing solutions for sustainable mobility, light weighting, and energy efficiency, mostly the automotive and aerospace industries. The Materials segment comprises the following GBUs:
    a. Specialty Polymers
    b. Composite Polymers
2. Chemicals: this segment excels globally in producing key chemical intermediates for daily use with top-notch assets, technology, and cost-efficient industrial innovation. The Chemicals segment comprises the following GBUs:
    a. Soda Ash & Derivatives
    b. Peroxides
    c. Coatis
    d. Silica
3. Solutions: The Solutions segment offers custom formulations for surface chemistry and liquid behaviour to increase efficiency and reduce environmental impact; this allows Solvay to deliver innovative and competitive solutions that create value for customers and support a healthier, more sustainable world. The Solution segment comprises the following GBUs:
    a. Novecare
    b. Special Chem
    c. Technology Solutions
    d. Aroma Performance
    e. Oil & Gas

To provide global shared services, to support with Solvay's major administrative processes as well as information services, in 2013 Solvay launched Solvay Business Services (SBS). The key functions SBS provides are split into service lines (SLs), and they are the following ones:

1. Finance: takes care of Treasury, as well as Management, Financial and Statutory & Country Accounting.
2. Procurement: takes care of Provisioning (energy, raw materials, etc), Accounts Payable, Travel & Expenses, and Data & Analytics related to their processes.
3. Credit: takes care of Cash Collection, Account Receivable and Customer Credit Analysis.
4. Human Resources: takes care of Payroll & Compensation, Global Mobility, Employee support and the entire Hire to Retire process.

In addition to these four SLs, there are two additional ones, considered transversal due to their shared interests within the previous four:

5. Transformation Office: takes care of driving SBS ISO certifications for Quality & Information Security Management, training and supporting SBS teams through their Continuous Improvement journeys and to aim to service excellence.

6. Merger & Acquisition: takes care of advising the business in their acquisitions, as well as managing the transitions to ensure business continuity.

1.1.1. Company History

Solvay was founded in Belgium in 1863 by Ernest Solvay, and his brother Alfred, out of a technological breakthrough; the two brothers, together with a small circle of relatives, developed the ammonia-soda process, securing this manufacturing advantage with a patent. Following a period of technical challenges, the enterprise quickly expanded to become one of the biggest multinational corporations, thereby encompassing various national cultures. At the turn of the twentieth century, Solvay was a single-product company leading a fast-growing industry. With its associated companies in Great-Britain, Germany, Austria-Hungary, the US and Russia, the Group owned 32 plants, employed 25,000 people and produced nearly two million tons of alkalis. The importance of the company can also be testimonied by the 1911 Solvay Council of Physics held in Brussels, where within the participants we have physicist of the calibre of Einstein, Marie Curie, Planck, Lorentz, Rutherford, and many others, and the 1927 Solvay Council of Physics where leading figures Albert Einstein and Niels Bohr famously debated quantum mechanics at the conference, and of all the attendees, 17 of the 29 were or became Nobel Prize winners.

Thanks to its closely held manufacturing secrets and family shareholder base, the Group successfully endured both World Wars. In the early 1950s, Solvay resumed its global expansion and diversified its operations. Diversification of operations meant also investing a lot into the creation of research centres. While until the 60s Solvay only served only industrial markets, the company began to expand into consumer products too. The main consumer products Solvay put in commerce were in field of plastics (PVC) and peroxides. Being the largest producer of PVC in Europe, Solvay was strongly tied to the raw material needed to produce PVC, ethylene coming from oil. Ethylene was not sourced directly, and therefore Solvay's financial results have been partially tied to oil prices.

In 1967, after more than a century of private property of the company by the founding families, Solvay transitioned to be a high-profile public company; the decision was taken in order to finance new activities.

The company in the following years kept refocusing its activity portfolio and by increasing its presence into emerging markets; after the fall of the Berlin wall Solvay regained importance in Eastern Europe, and the objective of gaining presence in Asia had a successful outcome after Solvay established business in Thailand, South Korea, Japan, India and China. Post-oil crisis, Solvay acknowledged the necessity of transitioning into more stable fields. Life sciences, including human and animal health and crop protection, were promising areas. Solvay achieved this through acquiring firms such as Kali-Chemie and Salsbury, with a later focus on human health, resulting in significant investments in the sector. Another sector Solvay decided to invest in specialty polymers, with the focus on engineering plastics that yielded higher added value.

The last 15 years witnessed an intense transformation of the Group's profile. The divestment of the pharma business and the acquisition of Rhodia, a French-based chemicals company, in 2010-2011 kicked-off a radical process of metamorphosis.

With the launch of Solvay One Planet in 2020 and Solvay One Dignity in 2021, Solvay raised the bar on its Environmental, Social, and Governance commitments.

1.1.2. The future of the company: POWER OF 2 Project

In March 2022, Solvay announced the plans to separate the company into two independent public companies. The companies will be:

1. EssentialCo: will contain mono-technology business including Soda Ash, Peroxides, Silica, Coatis and Special Chem, all reported as the Company's Chemicals segment. In 2021, these businesses generated around €4.1 bn in net sales.
1. SpecialtyCo: will consist of the Materials segment, which includes the high-margin Specialty Polymers and high-performance Composites, and most of the Solutions segment, such as Novecare, Technology Solutions, Aroma Performance, and Oil & Gas. These businesses collectively generated roughly €6.0 bn in net sales in 2021.

The reason behind this decision is to sharpen the focus and competitiveness of the different businesses, as well as providing value to the company's shareholders. Having followed successful financial and operational strategies in the last years, the Materials and Solutions segment have gained more resilient, self-sustaining, and profitable characteristic; the Chemicals segment kept its record of resilient cash generation.

The transition is expected to happen in the second quarter of 2023, as publicly announced by the Company, and to be completed by end of the same year.

1.1.3. Relevance of the Company in today's market

As mentioned in the previous subchapters, Solvay is a well-established leading company in the chemical and materials industry.

Based on the 2022 Solvay annual report, the company generated €13.4bn in net sales across 61 countries around the world: majority of the sales occurred in Asia, Middle East, and Africa (33%), followed by Europe (27%), North America (26%), and Latin America (14%). The sales are distributed between the Solutions (€4.84bn), the Chemical (€4.50bn) and the Materials (€4.07bn) sectors.

In 2022 the Company counted 22,000 employees, half of which are employed in European countries. To support the production processes and manufacturing, the company counts 99 different industrial sites around the world. The Research and Innovation department in 2022 counted 2,030 employees and €349M allocated to finance the projects.

1.2. INTERNSHIP OVERVIEW

During my 1-year long internship in Solvay, I had the opportunity to cover the role of Process Mining Trainee inside the Solvay Process Mining team. As a Process Mining Trainee, most of my responsibilities were in the Data Engineering field, especially Extract, Load, Transform (ETL) and Data Modelling. As a matter of fact, the team is composed by the Process Mining Manager, two Data Engineers and two

Process Mining Trainees. During my internship, I had the opportunity of following closely many different projects and use cases: some were already implemented while I joined and we had to work on maintenance and support, while some new projects were an opportunity to follow closely on the development. In particular, the main processes that the team manages are Purchase-To-Pay (P2P), Order-To-Cash (O2C), Accounts Payable (AP), Record-To-Report (R2R), Account Receivable (AR) and Material Master Data (MMD).

Due to the increase of projects the team has been involved into, starting from 2023 it launched the Process Mining Extended Centre of Excellence (PM CoE). The reason behind the decision was to share knowledge and responsibilities around PM, Celonis, project visibility and such between as many users as possible. The roles defined for users belonging to the PM CoE are PM Analyst, Process Owner, Champion, and Value Realization Owner. The PM Analyst has knowledge about the business and the data and works to validate and prioritize potential improvements. The Process Owner is owner of KPIs, Alerts and Automations and provides the PM team with specific definitions and requirements. The Champion is a key business user and has the responsibility on training new users and to manage maintenance topics with the PM Team. The Value Realization Owner, finally, is responsible of the improvement projects' realization, as well as making sure that the value is realized during the year and reporting to the company about value realization.

The team was created in early 2020, and since then it grew during the years, both considering the number of members working within, and for the number of projects carried out. The team is part of the Transformation Office of the SBS organization, specifically inside the Quality, Continuous Improvement and Process Mining department.

The Process Mining software used by the team, since 2020, is Celonis. Celonis is a company based in Munich, Germany, created in 2011, same year in which the Process Mining Manifesto was published by Wil van der Aalst and the other founding members of the IEEE Task Force on Process Mining. Celonis is nowadays considered the leader of the commercial Process Mining tools, as testimonied by the Gartner report published in 2023.

1.3. INTERNSHIP OBJECTIVES

The Solvay's Process Mining Team, where I carried my internship, has the objective of working together with GBUs and/or Service Lines within SBS to help them having visibility on the real state of the processes running within their department, to find improvement or automation opportunities; the final goal is to reduce costs and inefficiencies, and to increase automation and optimization of resources.

On an academic standpoint, my personal objective while working within the Company was to increase and deepen my knowledge of the subject; it was a great opportunity to learn how organizations can leverage Process Mining projects to improve their ongoing processes and to bring value and support to the stakeholders, as well as exploring further the existing literature in the field.

1.4. STUDY RELEVANCE

This document's objective is to carry out a detailed analysis on Process Mining techniques, applied in real world scenarios to solve real world challenges. The results will contribute to the literature in two ways. First contribution consists in providing information about Process Mining techniques applied in

the industry, their frameworks, and the evaluation on the impact that those techniques had within the organization. Secondly, it will contribute to understanding the value that Process Mining can leverage within organizations.

2. THEORETICAL BACKGROUND

Processes are present everywhere in the context of organizations; by using process management, an organization is able to maintain high-performance processes, which "operate with much lower costs, faster speeds, greater accuracy, reduced assets, and enhanced flexibility" (Hammer, 2015). It can be stated that the majority of businesses implement a methodology centred on processes for overseeing their activities, and that the notion of Business Process Management (BPM) is a widely recognized and acknowledged concept (Zairi, 1997). Within the domain of BPM, Process Mining grew to be an important branch for process management; it aims to "Discover, Monitor, and Improve" processes (W. Van Der Aalst et al., 2012).

In this chapter it will be presented an extensive literature review regarding the main topics of the domains that this document will encompass, to better understand the analysis that will be discussed in the next chapters.

2.1. BUSINESS PROCESS MANAGEMENT

BPM holds a strong significance to deeply understand the contents of this document: in fact, Process Mining sinks its roots in BPM. Business Process Management (BPM) is a "comprehensive system for managing and transforming organizational operations" (Hammer, 2015). BPM approach sets the objective to analyse and continually improve core business activities, such as manufacturing, marketing, communications, and other key sectors of a company's operation (Zairi, 1997).

The global goal for BPM is therefore to improve the process performances; the main variables to be optimized, in the businesses, usually are throughput time (complete the process faster), cost (reduce the cost in performing the processes), quality (achieve better products/services), and resource consumption (reduce as much as possible the resources involved in performing the process).

Another key field in which BPM is used is Process Governance and Compliance Checking. Business Process Compliance aims to make sure that the operations within organizations are aligned with the laws of regulatory entities, and with the internal statute of the organization itself (Hashmi et al., 2018). Two different techniques are identified: forward or backward compliance. The backward compliance analyses post-execution processes, while the forward compliance takes place at design- or run-time (Kharbili et al., 2008).

BPM has its origins in various concepts and management strategies. Business Process Reengineering (BPR) is certainly one of them: in the early 1990s, it was observed that several businesses had significant increases in performance after drastically altering one or more of their processes (Hammer & Champy, 1993). BPM however evolved to take a different approach on process redesign, preferring a more continuous and gradual approach, in contrast to the significant altering of business processes praised by BPR. Another important intellectual antecedent to BPM is the statistical process control, which "led to the modern quality movement and its contemporary avatar, Six Sigma" (Hammer, 2015). The first works published regarding this topic appear in the early 1930s, and during the 1950s the statistical controls started to spread across industries; they were a way to the growth of companies, which strongly relied on the continuous improvement of the efficiency and optimization of production and distribution (Derning, 1953).

The way in which BPM is implemented within organizations and manages business processes is usually defined BPM Lifecycle; it shows the way in which the different steps of business process managing process are defined in the organization (Szelągowski, 2018). Dumas described the BPM Lifecycle as a continuous cycle, containing the following phases:

- Process Identification: in this phase, once identified a business problem, the processes relevant to the problem are identified, delimited, and related to each other. The output to this phase is a process architecture that provides visibility on the processes and their relationships.
- Process Discovery: in this phase, the current state of the processes is mapped and documented, either in one or multiple as-is process models.
- Process Analysis: in this phase, there's a quantitative evaluation of the processes, by using performance measures. The output is then a prioritized list of gaps within the process, and their impact.
- Process Redesign: in this phase, the necessary changes to the act on the gaps identified in the previous phases are identified, to achieve the desired improvements in the performance measures. The output usually is a to be process model.
- Process Implementation: in this phase, the changes to move from the as is to the to be process are tackled. These changes can be either regarding organizational change management or process automation.
- Process Monitoring and Controlling: once the to be process is implemented, the performance measures are calculated to check whether they are conformant to the performance objectives. Once this phase is complete, the cycle will continue to identify new gaps, to carry on the Continuous Improvement method.

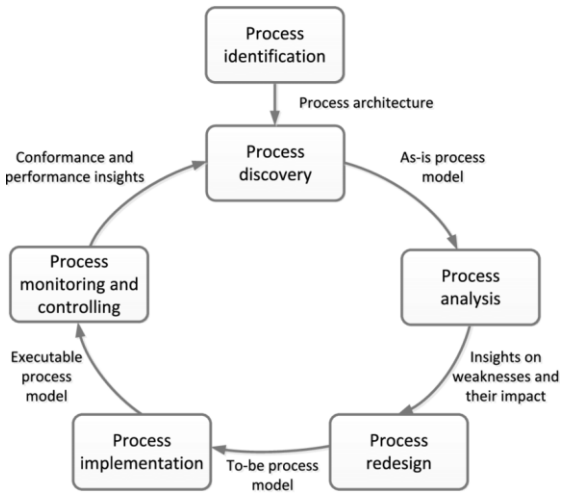

Figure 2 - BPM Lifecylcle

Adapted from source: (Dumas et al., 2013)

Six Sigma is another example of Process Lifecycle, similar to the DMEMO cycle (acronym for Design, Model, Execute, Monitor, and Optimize) (Szelągowski, 2018).

Six Sigma, one of the most advanced methodologies originated from statistical process control, was first introduced in 1988 by Motorola University Design as a manufacturing training program, and evolved later being used in many different sectors, such as services, sustainability, and supply chain

management, as methodology for quality management (Tjahjono et al., 2010). The most common method used in Six Sigma is DMAIC, which acronym stands for Define, Measure, Analyze, Improve, Control, the different phases composing the model. The Define phase aims to understand and define the scope of the project, as well as the gap that is aimed to improve. The Measure phase has the objective of collecting data about the process performances, and to build a benchmark for the desired improvement. Analyze phase focuses on exploring the data to identify the most impacting root causes for process performance, or variation. Improve phase aims to implementing changes within the process, by tackling the identified root causes. Finally, the Control phase is to make sure that the identified changes and improvements are established, and to monitor performance to make sure that the performance is maintained.

2.2. PROCESS MINING

Over the last two decades, Process Mining gained increased interest as a field, due to its significant potential to be implemented by organizations to monitor and improve their internal processes' performances.

Process Mining is born as a hybrid approach between BPM and Data Mining. Data Mining techniques are used to extrapolate from historic data information and insights; however, most Data Mining techniques are not process centered. Classic BPM approaches instead use process models as static descriptions or to drive a BPM system; when process models are simply qualitative, they tend to not describe well the reality behind the processes. When, however, the models are used to configure BPM systems, they tend to force people to work in a particular manner (W. M. P. van der Aalst, 2011). Process Mining can be then seen as the point of contact between both disciplines; similarly to BPM techniques, it is process centered, and it is driven by factual and historic data rather than an approach of modelling processes on predefined goals, requirements, and assumption.

The Process Mining Manifesto, published by van der Aalst et al. in 2012, states the Process Mining techniques can extract knowledge from event logs commonly available in today's information systems. These techniques provide new tools to discover, monitor, and improve processes (W. Van Der Aalst et al., 2012).

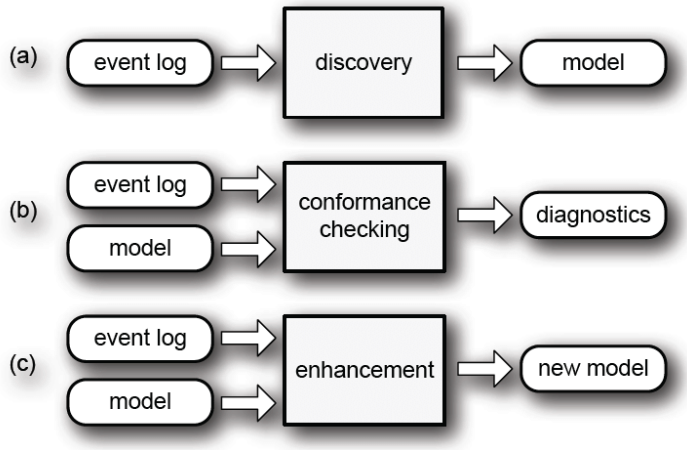

Figure 3 - Three types of Process Mining

Adapted from source: (Rudnitckaia & Humby, 2014)

Originally there were three different types of Process Mining (Rudnitckaia & Humby, 2014):

- Discovery: a discovery technique takes the event-log and creates a process model; there is no need of a-priori information regarding the process in scope. This usually is the first approach that is tackled by organizations when implementing PM tools. One example is the Alpha-algorithm that takes an event log and creates a process model explaining the behaviour recorded in the logs.
- Conformance: here, a pre-existing process model is compared to the event logs existing for the same process. These techniques are used to see if the *de-facto* state of the process conforms to the model, and vice versa.
- Enhancement: the main goal of these techniques is to modify and improve an existing process model using the information provided by the *de-facto* process created from the event logs (e.g., eliminating bottlenecks, avoid rework, etc.).

2.2.1. Process Mining Terminology: Event-Logs, Cases and Activities

Process Mining's backbone is the event logs (often referred as digital footprints); without them, it is not possible to talk about Process Mining. In the era of digitalization, organizations see IT systems as crucial due to the central significance that data assumes for business competitiveness in all sectors. Because of that, organizations adopting IT systems are generating volumes of data unprecedented in the history. In complex application scenarios in large companies, the data required to extract all information might be in different databases and IT systems used by the organization. Most of the organizations and IT systems do not store the data in a process centric way, and therefore the initial challenge can be locating the right data and applying the necessary transformations to create the data structure needed by Process Mining techniques (Diba et al., 2020).

There is a list of assumptions to be made about event logs (W. van der Aalst, 2016):

- A *process* consists of *cases.*
- A case consists of *events* such that each event relates to precisely one case.
- Events within a case are *ordered*, usually chronologically.
- Events can have *attributes*.

It is important to deeply understand some formal terminology that will be exhaustively used in this document.

- *Event:* an event refers to a specific occurrence or action that takes place within a process. It represents a unit of work or activity performed by an entity or system and is typically recorded in an event log.
- *Event Log:* An event log is a collection of recorded events that capture the activities performed within a system or process. It typically includes information such as timestamps, case identifiers, activity names, and other relevant attributes; some attributes could be the user

who performed the activity, the system information where the event took place or the cost it took to perform the activity.

- *Case*: A case represents a process instance (e.g., a Sales Order in the O2C process). It is identified by a unique case identifier, often called the Case ID. Each case encapsulates a sequence of activities, that constitutes the activity history.
- *Activity:* An activity refers to a specific and well-defined step or task within a process. It represents an action performed by an actor or system during the execution of a case. Activities are often associated with descriptive names or codes.

Case ID	Activity Name	Event Time	UserName	Cost ($)	…
…	…	…	…	…	…
1	Receive Sales Order	30/05/2023 12:00:00	USER01	0	…
1	Confirm Stock Availability: Successful	30/05/2023 13:45:00	USER02	0	…
1	Send Sales Order Confirmation	30/05/2023 14:00:00	USER01	0	…
2	Receive Sales Order	30/05/2023 14:15:00	USER03	0	…
1	Create Delivery	30/05/2023 14:30:00	USER01	25.00	…
2	Confirm Stock Availability: Not Available	30/05/2023 14:30:30	USER03	0	…
1	Sales Order Picked Up for Delivery	30/05/2023 16:00:00	DELIV01	0	…
2	Reject Sales Order	30/05/2023 16:15:00	USER03	0	…
…	…	…	…	…	…

Table 1 - Example of an Event Log's fraction

Event Logs need to be structured in a standardized way. Table 1 shows an example of what a small fraction of an event log (simplified as much as possible) built for an Order-To-Cash process could look like. Some attributes, such as Case ID, Event Time, Activity Name are necessary to create the most basic event log structure; it is key to refer a specific event to a specific case, and know when the event happened. Other attributes, such as UserName and Cost ($), are only optional; their presence, however, adds further information regarding the event characteristics.

2.2.2. Abstraction and Correlation in Event Mapping

Abstraction and correlation are concepts with a deep meaning in the Process Mining field.

Abstraction is used to link event data to events which represent the execution of activities; as such, it enables the understanding and analysis of the data in terms of specific activity execution (Diba et al., 2020). The abstraction relations can be either 1:1 or n:m. One high-level activity might result in multiple low-level events (e.g., cashing an invoice typically requires multiple steps), and one low-level event might relate to multiple high-level activities (e.g., contacting a customer via e-mail). Higher level events usually are easier to understand for the stakeholders, but low-level ones tend to contain more information and insight about the process they belong to. Event abstraction is then the method of grouping low-level events to higher-level events, and it is essential to help guiding process discovery methods towards discovering a process model that can be understood by stakeholders and is more useful for answering process questions (Mannhardt et al., 2018).

Correlation, on the other hand, involves the identification and establishment of relations between different activities within a process. Using correlation techniques in event mapping enhances the effectiveness of process mining by uncovering hidden insights and facilitating data-driven decision-making. Many Event Correlation techniques have been proposed depending on the available information, ranging from event attributes, available process models, or relations between data (Diba et al., 2020). The main techniques identified are finite state machine based, rule-based event correlation, codebook based, genetic algorithms, graph based, model-based reasoning, neural network based, and probabilistic approaches (Grimaila et al., 2012).

2.2.3. Data Quality and Event Logs Pre-Processing

Data quality is another key aspect to monitor when implementing any Process Mining; as all Information Technology fields, the common concept "*Garbage In, Garbage Out*" is applicable in Process Mining as well. Process Mining's inputs are event logs; therefore, the reliability of the output from process mining will be directly proportional to the quality of the event logs data that is used in the input.

In the Process Mining Manifesto (W. Van Der Aalst et al., 2012), it is stated that event logs need to be treated as first-class citizens; events should be trustworthy (should be safe to assume that the events really happened, and their attributes are correct), event logs should complete (no event in the process scope should be missing in the logs), any recorded event should have a well-defined semantics, and that the data should be safe (privacy and security are taken in account when recording the events). It is also defined an event log maturity score, ranging from one star to five stars, with the characteristics for each maturity level. Table 2 shows those maturity levels, making consideration on the different possible data sources.

Empirical and experimental evidence shows how most real-life event logs tend to be incomplete, noisy, and imprecise; moreover, contemporary processes tend to be complex and subject to a wide range of variations (Bose et al., 2013). A single event log can have multiple different issues; however it is been identified how many of those issues are systematically appear, and can be tackled by using well known solutions. In particular, the research made by Suriadi et al. in 2016, identifies the different event log imperfection patterns, and serves as a contribution to develop a systematic methodology for improving event log data quality.

Level	Characterization	Examples
★★★★★	Highest level: the event log is of excellent quality (i.e., trustworthy and complete) and events are well-defined. Events are recorded in an automatic, systematic, reliable, and safe manner. Privacy and security considerations are addressed adequately. Moreover, the events recorded (and all of their attributes) have clear semantics. This implies the existence of one or more ontologies. Events and their attributes point to this ontology.	Semantically annotated logs of BPM systems.
★★★★	Events are recorded automatically and in a systematic and reliable manner, i.e., logs are trustworthy and complete. Unlike the systems operating at level ★★★, notions such as process instance (case) and activity are supported in an explicit manner.	Events logs of traditional BPM/ workflow systems.
★★★	Events are recorded automatically, but no systematic approach is followed to record events. However, unlike logs at level ★★, there is some level of guarantee that the events recorded match reality (i.e., the event log is trustworthy but not necessarily complete). Consider, for example, the events recorded by an ERP system. Although events need to be extracted from a variety of tables, the information can be assumed to be correct (e.g., it is safe to assume that a payment recorded by the ERP actually exists and vice versa).	Tables in ERP systems, event logs of CRM systems, transaction logs of messaging systems, event logs of high-tech systems, etc.
★★	Events are recorded automatically, i.e., as a by-product of some information system. Coverage varies, i.e., no systematic approach is followed to decide which events are recorded. Moreover, it is possible to bypass the information system. Hence, events may be missing or not recorded properly.	Event logs of document and product management systems, error logs of embedded systems, worksheets of service engineers, etc.
★	Lowest level: event logs are of poor quality. Recorded events may not correspond to reality and events may be missing. Event logs for which events are recorded by hand typically have such characteristics.	Trails left in paper documents routed through the organization ("yellow notes"), paper-based medical records, etc.

Table 2 - Maturity levels for event logs

Adapted from source: (W. Van Der Aalst et al., 2012)

Therefore, data pre-processing assumes a key role to have a state-of-the-art Process Mining implementation.

In Table 2 we some different pre-processing approaches typically used in Process Mining frameworks identified and described in Sani's work published in 2020; to simplify the concept we'll assume the data as tabular, where rows correspond to cases and columns show the activities. Real-life pre-processing uses will likely use combinations of these approaches.

- Trace selection: this approach consists in selecting only a subset of rows (or cases) from the original event log and insert them as they are in the pre-processed event log. Some techniques, while performing this approach, perform removal of cases with outlier behaviour.
- Activity selection: this approach usually is applied to improve the performance of the later steps to reach the results; many PM algorithms have an exponential complexity on the number of activities (Hompes et al., 2015). However, by removing activities, there is the risk of adding behaviour not existing in the original data log; for example, removing activity *b* from sub-sequence *{a, b, c}* we implicitly assume a direct relation between *a* and *c* that does not exist in the original data log.

- Generalization: this approach consists of reducing the complexity of event logs by merging unique process instances or activities to general ones. This approach can be done at case level, or activity level, or both levels.

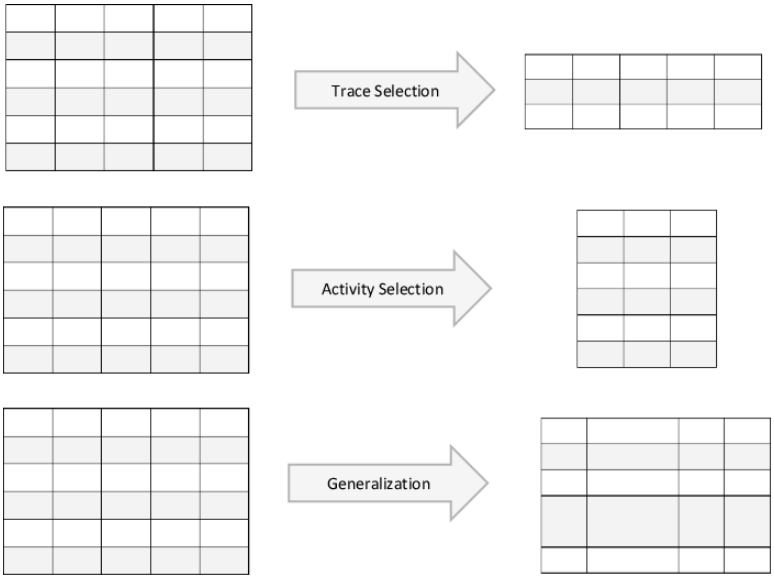

Figure 4 - Different event data pre-processing approaches

Adapted from source: (Sani, 2020).

2.2.4. Business Process Modelling

Business Process Modelling involves the creation of graphical representations or models of the different activities, tasks, resources, and interactions occurring within an organization; it is a valuable technique to better understanding and analysing business processes with the aid of the visual representation. Conceptual business process modelling is deployed on a large scale to facilitate the development of software that supports the business processes, and to permit the analysis and re-engineering or improvement of them (Aguilar-Savén, 2004). It can be considered the output of the Process Discovery phase.

Process Modelling is a tool for coping with the complexity of process planning and control; especially in enterprise-wide process management projects the design of process models can be a big challenge, due to the complexity and high-number of different process models to be created. This could raise issues if the understanding of these models is limited only to “modelling specialist”, and not to all people involved (Becker et al., 2000).

Many different Business Process Modelling techniques have been identified and researched over the years; in the next subchapters we’ll go through the ones considered the most influential and significant for the scope of the document. The understanding of these techniques, and the significant importance

that Business Process Modelling carries for Process Mining, is important to deeply understand the content of this document.

2.2.4.1. Business Process Model and Notation

Business Process Model and Notation (BPMN) is a standard for Business Process Modelling, created by the Object Management Group. The standard's primary goal is to provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the business resources who will manage and monitor those processes. Thus, BPMN creates a standardized bridge for the gap between the business process design and process implementation (Object Management Group, 2013).

BPMN defines a Business Process Diagram (BPD); a BPD is made up of a set of graphical elements. One of the main drivers for the development of BPMN is to create a simple technique to create Business Process Models, while not impacting the inherent complexity of the processes (White, 2004).

In BPMN, elements are grouped into four different core sets, which can be seen in Figure 5, and described below (Kluza et al., 2017; Von Rosing et al., 2014; White, 2004):

- Flow Objects: the main elements in this category are gateways, activities, and events. Gateways are used to control the convergence and divergence of process paths and can be of many different forms. Activity is a generic term for work that gets done and can be either atomic (task) or non-atomic (sub-process). An event is something that happens during a business process; they affect the flow of the process, usually have a cause (trigger) or an impact (result), and can be either *Start, Intermediate* or *End*.
- Connecting Objects: these elements connect flow objects together in a diagram to create the flowchart-like structure. A sequence flow shows the order in which the activities are performed in the process. A message flow is used to show the flow of messages between process partecipants. Association flow is used to associate data, text, or other artifact elements to flow objects.
- Swimlanes: the goal of using *swimlanes* is to organize activities into separate categories to illustrate different capabilities or responsibilities. A pool represents a participant in the process (e.g., separate business entities, participants) and can contain multiple lanes. A lane is a sub-partition of a pool, and represents different specific company responsibilities (e.g., different departments within the same company).
- Artifacts: these elements allow developers to bring additional information within the model, and making it more readable and understandable (e.g., text annotations, data object to show the necessity of a database to perform a task, etc.)

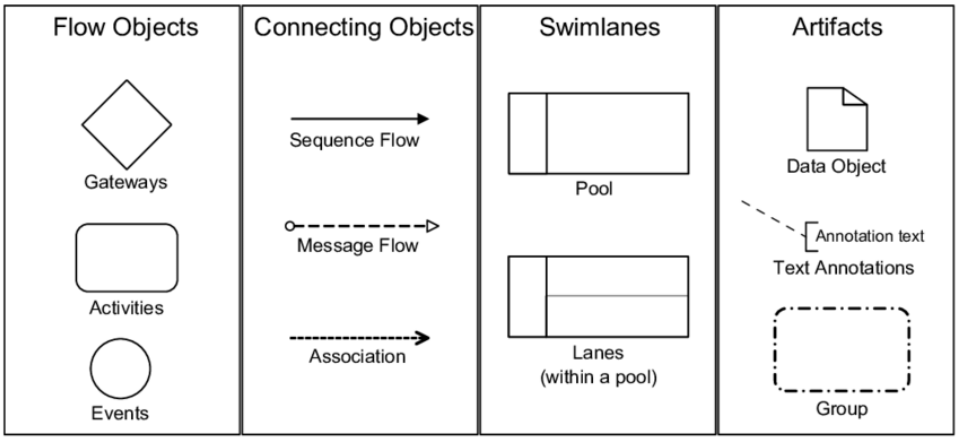

Figure 5 - BPMN core elements of BPD

Adapted from source: (Kluza et al., 2017)

2.2.4.2. Directly-Follows Graph

The Directly-Follows Graphs (DFGs) is a simple notation that is commonly used, and it is considered the *de-facto* standard for commercial PM tools (W. M. P. Van Der Aalst, 2019). Celonis, the software that will be used for the analysis in scope for the document in the next chapters, implements as well the DFGs.

As the name suggests, this modelling techniques aims to create a visual graph, showing the frequency of the direct relations between the activities that are captured in the event logs. The graph can be considered a square matrix, where activity names are indexes for both rows and columns; the values in the matrix cells are the number of times in which the activity on the row index happened before the activity on the column index (Jalali, 2020). Figure 6 provides a simple example on the steps mentioned above. In terms of graph theory, we have the nodes (activities) and the edges (connections between the activity).

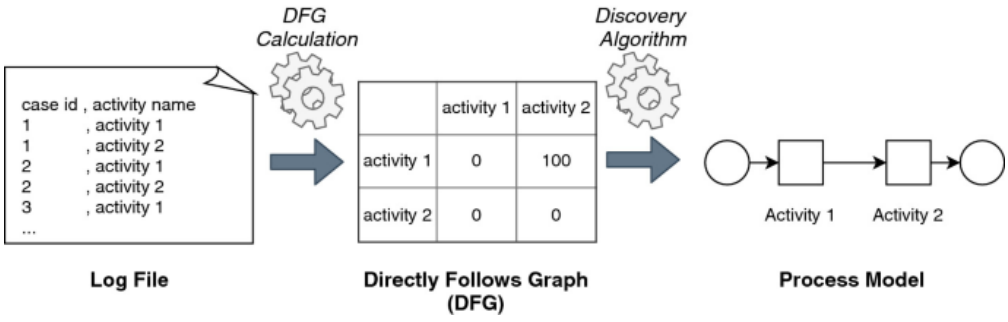

Figure 6 - Steps in a process discovery algorithm

Adapted from source: (Jalali, 2020)

Despite its apparent simplicity, the use of DFGs has some significant limitations that need to be addressed when the graph is used for process discovery. As the other modelling techniques, DFGs show a representation of the process reality; some misalignments with the *de-facto* process might exist if the preliminary steps occurred with errors (e.g., not considering an activity when creating the event log will mean that the activity won't show in the DFG).

Another limitation is dealing with the concurrency problem, that raises when activities have flexible ordering, and can appear in different order in different cases. This issue can lead to *spaghetti-like DFGs*, due to the loops between activities, even when activities are executed only once (W. M. P. Van Der Aalst, 2019). The name of this typology of process models comes from the visual similarity of a dish full of spaghetti, and the amount and complexity of connections between activities in the process model. Other reasons for why the spaghetti process models are discovered from event logs are: very low abstraction level on activity definition, high inner complexity of the process, poor event logs' data quality (Batista & Solanas, 2019).

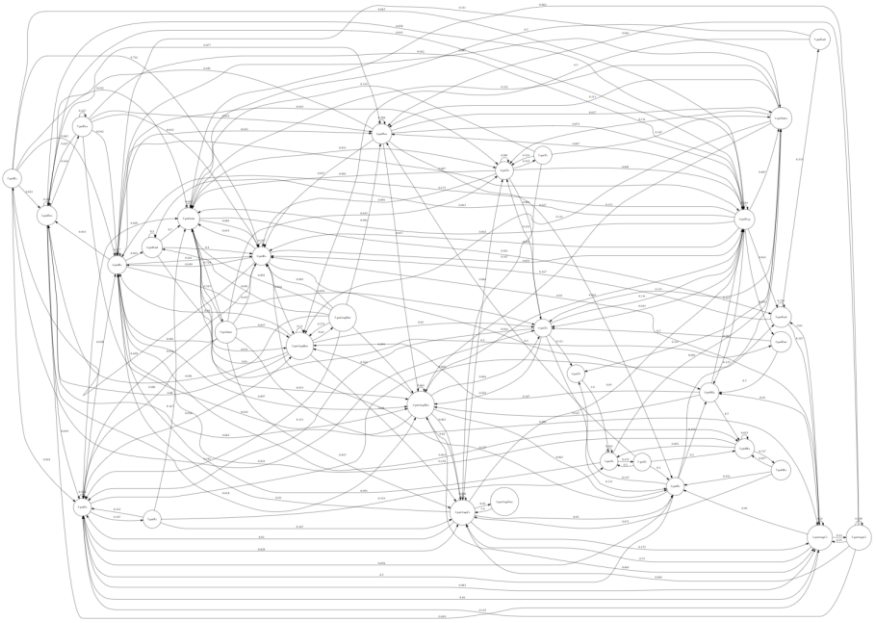

Figure 7 - Example of a spaghetti process model

Adapted from source: (Batista & Solanas, 2019)

One way in which this complex process model can be simplified is by setting a frequency threshold; by doing so, only activities and/or connections appearing more than the threshold chosen will be shown in the model. This allows the usage of a much-simplified version of the process model, focusing on its most common characteristics; however, the downside of the simplification is the loss of visibility of the activities and/or connections appearing less times than the threshold.

Furthermore, three approaches can be identified when applying filters to DFGs: arc-based, activity-based, and variant-based. Arc-based approach consists in filtering the DFG on the most common connections between activities, and not considering the least common ones. Activity-based approach is similar to the previous one, but in this case the DFG is filtered to keep only the most common activities and leaving out the activities with less occurrences. The last approach, variant-based, needs a definition to understand what a variant means. A process variant is "a subset of executions of a business process that can be distinguished from others based on some characteristics" (Taymouri et al., 2021); in this document, the characteristic that will be taken in account is the activity history, for

each case. Therefore, variant-based approach filters the DFG on the most common variants and leaving out the variants less occurrent. Based on the analysis that needs to be done, either of these approaches, or a combination of the three, can be used for filtering the spaghetti process model, as commonly done.

2.2.4.3. Petri Nets

Petri Nets are one of the most common modelling techniques; a Petri Net is an abstract, formal model of information flow (Peterson, 1977). In fact, BPMN inherits some of the elements and characteristics from Petri Nets.

The original structure of Petri net is a directed bipartite graph with two different types of nodes, called places (represented by circles) and transitions (represented by rectangles). Places represent states or conditions of the system; a place can hold different tokens indicating the presence of objects, resources, or statuses. Transitions instead represent events or actions that could happen in the system. Arcs are the connections that link different nodes; an arc cannot connect two nodes of the same types. Some extensions have been identified for Petri nets, such as colour, time, and hierarchy extensions (W. M. P. van der Aalst, 1998).

2.2.5. Modern Approaches to Process Mining

In addition to the three Process Mining approaches that were identified in the chapters before (Discover, Conformance, Enhancement), new approaches have been identified, developed, and integrated, both in organizations and in the academic field. The Process Mining Manifesto states that one of the key challenges faced for the field is to combine Process Mining with other types of analysis; some of the approaches described below are successful solutions for the challenge.

Process Simulation provides techniques to create a digital twin of the process, and test process re-engineering solutions before the actual implementation (Aguirre et al., 2013). This can be achieved by combining Process Mining with simulation techniques, such as Discrete Event Simulation (DES); Process Mining tends to be backward looking, but with Process Simulation we can explore possible re-engineering solutions and anticipate future performance gaps (W. M. P. Van Der Aalst, 2018). Figure 8 shows a possible scenario in which this combination between simulation and Process Mining can be used together: the process model's *de-facto* performances identified with the Process Discovery and Performance analysis phases is compared to the performance on the same process model calculated with the simulated event logs.

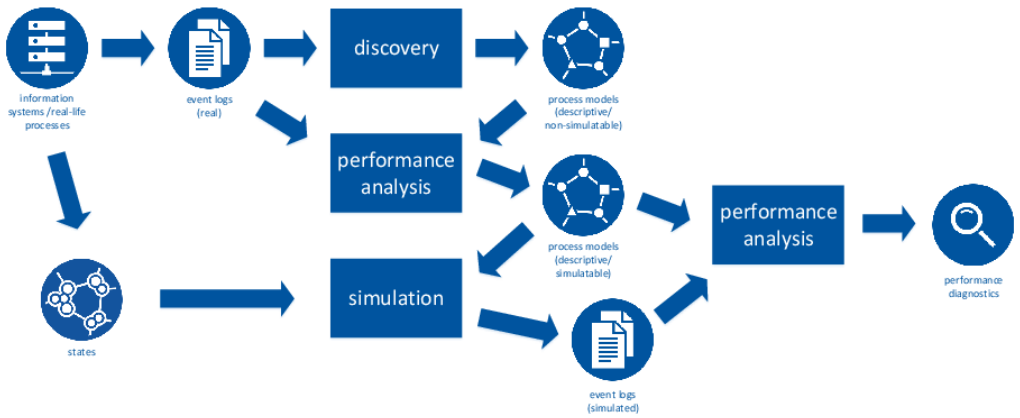

Figure 8 - An advanced scenario using a combination of Simulation and Process Mining

Adapted from source: (W. M. P. Van Der Aalst, 2018)

One of the main goals of Process Mining, as mentioned before, is Process Enhancement. This one phase can be divided into three different steps: first one is identifying the gaps within the process, then finding the root causes of each of the gaps identified before and estimating the possible effect if the root causes are fixed, and finally re-engineering the process to enhance performance (Qafari & van der Aalst, 2020). Focusing on the second step, the combination of Process Mining with Root Cause Analysis (RCA), can lead to continuous improvements to the process.

Robotic Process Automation (RPA) provides the opportunity to automate manual, repetitive and well-defined tasks, traditionally operated by humans, with no risk of manual errors. Process Mining can be used in parallel with RPA by organizations to achieve cost savings. The three steps approach, proposed by Geyer-Klingeberg et al., shows how PM can support RPA implementation. First step is assessing RPA potential; all tasks that potentially could be automatized are identified and prioritized. Second step consists in developing RPA application; some PM software, such as Celonis, have RPA suites embedded. Last step is sustaining RPA applications, by monitoring their usage and tracking the benefits, such as the return on investment (Geyer-Klingeberg et al., 2018).

2.3. Implementation Methodology of Process Mining in the Industry

As mentioned in the previous chapters, since Process Mining emerged as a new field, organizations implemented more and more PM as monitoring and decision-making tools. The variety of the markets and companies (such as banking, healthcare, chemical, etc.) adopting PM is proof of the scalability of the technique itself, and the opportunities it offers. To front the growing request of usage of PM within different markets, many companies emerged and developed commercial PM software, such as Celonis, IBM, SAP, Microsoft, Software AG, ProM (currently, the best open-source available), and many more.

Process Mining implementation within organizations faces multiple challenges, both on a managerial and technical prospective; not only the event logs and necessary resources need to be present, governance, strategy, and people play an important role as well for the implementation to be successful.

In the work by Grisold et al., a focus group consisting of process managers from various industries are interviewed on four key questions: (1) how one can initiate and measure a business case for PM, (2) which processes should be selected for PM, (3) how to align different PM initiatives and (4) what to consider regarding data availability. The key challenges identified are: (1) the difficulty in assessing and predicting outcomes when implementing PM, (2) following a continuous and on-going usage of PM and the difficulty of identification of the properties a process needs to have to be eligible, (3) creating a successful governance model ensuring PM is used to its full potential, (4) data availability and segregation between different systems, data privacy.

In the Process Mining Manifesto, the following list of challenges when implementing PM is presented: (C1) Finding, Merging, and Cleaning Event Data, (C2) Dealing with complex event logs having diverse characteristics, (C3) Creating representative benchmarks, (C4) Dealing with concept drift, (C5) Improving the representational bias used for Process Discovery, (C6) Balancing between quality criteria, (C7) Cross-organizational mining, (C8) Providing operational support, (C9) Combining PM with other types of analysis, (C10) Improving usability for non-experts, (C11) Improving understandability for non-experts. (Bigui & Cho, 2017) analyses each of these challenges, providing a comprehensive and critical literature review in the context of each of them, as well as some solutions found to overcome the challenges themselves.

3. FIELD WORK

To discuss the real-world use cases, central for this document, this chapter is structured in the following way: first, the method used to build this analysis will be discussed further. Afterwards, an overview of the main technologies involved in my work, and of the main processes involved in the use cases. Once described the processes, also the key elements and characteristics discussed within the result chapter are presented. Finally, the conclusions and limitations of the field work are presented.

3.1. METHOD

To develop this document and analyze the field work, a set of multiple cases was used – the objective was to explain in detail the Mining Process in different processes, aiming a consistence analysis with Data collected.

The main method used for data and frameworks' collections is observation: during the day-to-day internship tasks, the work processes, interactions, projects and strategies within Solvay have been directly observed. This is supported by the document review on the documentation provided by the Process Mining team to report the use cases' functional and technical activities.

Regarding data analysis, both quantitative and qualitative methods for collecting data were used: part of the support on evidence is given by numbers, KPIs, and figures. The qualitative approach complements the quantitative one when showing framework solutions to the use cases. Some quantitative figures must be hidden or left out due to privacy and/or confidentiality reasons.

When discussing Project Management, the main method used is Work Breakdown structure: using this method, we can divide a major project into all activities and components involved to have a 360 degrees visibility on all tasks needed for the successful outcome of the project itself.

3.2. TECHNOLOGIES

In this subchapter, the main technologies directly involved in the projects that will be analyzed and discussed are presented, to highlight their significance and impact on the organization's workflow. During the internship, I had the chance to learn in depth the capabilities of these technologies, as well as getting much experience on their usage.

3.2.1. Celonis Execution Management System

Celonis Execution Management System (EMS) is a Process Mining software provided by Celonis. Founded in 2011, Celonis emerged during the years as a leader in the field of Process Mining, Process Optimization and Digital Transformation; it earned a strong recognition within the industry and established key partnerships with other key players across various sectors, such as IBM, Accenture, AWS, Microsoft, Salesforce, and many others, as well as many academic partnerships.

Celonis EMS was launched in 2020, and it is a cloud service: it contains different sections, differing from each other for their usage and capabilities. The ones described in detail are the ones most used during my internship.

Data Integration is the section to perform the end-to-end ETL activities. It offers pre-built different connectors to different data sources to extract the data. Once the data is extracted, data-

preprocessing can be performed. To do that, the programming language used is SQL; the database is hosted by Vertica database management company. After the pre-processing activity is performed, the data is loaded to the front-end; to do that, data modelling is required.

Celonis Studio is the development front-end used by analysts and admins to create dashboards, action views, action flows, skills, knowledge models and reports. The programming language used in Studio is Process Query Language (PQL); PQL is a domain specific language, created to suit the need of having a process tailored language. It is inspired by SQL, and it is a rather simple language. Some of the functions, such as PROCESS_LIKE, MATCH_PROCESS, allow users to simply translate process patterns into queries.

Apps is the application used by business users. It allows them to navigate through the dashboards and all the solutions implemented. It is read-only privilege.

3.2.2. SAP ECC

SAP ERP Central Component (ECC) is an on-premises ERP system. SAP is one of the largest ERP providers in the world.

SAP ECC is a software that works to support the core systems of the organizations, to store data, to manage the key transactions for the company, and many other use cases, within different departments. It usually is used by middle and large companies, due to their need to centralize operations and follow the digitalization that is marking our era.

3.3. PROCESSES

In this subchapter, a high-level description of the general characteristics of the processes involved in the analysis of application of Process Mining will be presented. The focus is on the Purchase to Pay and Order to Cash processes, being the ones that will be later presented in terms of use cases.

In addition to those two processes, the team also implemented other processes, such as Record to Report and Inventory Management (currently in development phase).

3.3.1. Purchase To Pay

In general terms, the Purchase to Pay (P2P) process is an end-to-end core process, common to most organizations in different markets; it consists in all activities involved in the process of acquiring the necessary goods and/or services to make businesses run (e.g., for a manufacturing company acquiring the necessary raw materials).

P2P generally starts with the creation of a Purchase Requisition order and ends with the clearing of the Invoice related to the Purchase Order. Figure 9 shows on a high level of abstraction the desired flow of a Purchase Order; as seen in the figure, the P2P process can be segmented into two different subprocesses, the first one being Purchase to Deliver (also referred to as Provisioning), and the second one being Accounts Payable. In this document the two subprocesses will be considered combined, forming the P2P process.

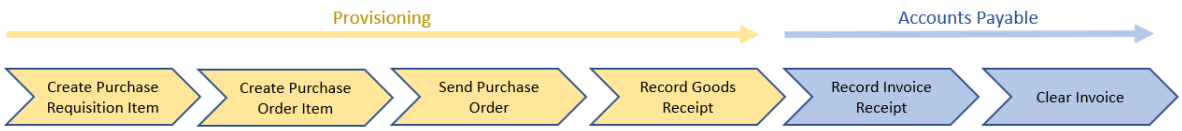

Figure 9 - Typical P2P Process Flow

In a context of medium and large companies, the process will likely present a high variability: this is due to the high number of actors involved, complex business rules, different compliance rules to follow in different countries and/or markets, and many other possible reasons.

The main documents involved within the P2P process are:

- Purchase Requisition (PREQ): it is an internal document establishing the need to purchase a good or a service. Once reviewed and approved, it is used to create the Purchase Order.
- Purchase Order (PO): it is a document showing the different goods and/or services needed, created by a company to its suppliers. A single PO might contain multiple goods and/or services; each one corresponds to a PO Item. The relationship between POs and PO Items is 1:n.
- Invoice: it is a document recording the business transaction, presented to the company by the supplier after the goods and/or services have been delivered/provided, and contains all information regarding the payment.

3.3.2. Order To Cash

The Order to Cash (O2C) is another end-to-end core process, common for all businesses providing services and/or goods; it follows all activities involved in between receiving an order from a customer to fulfilling that order to receiving the payment.

It is the opposite of the P2P process: in O2C the organization is providing the services/goods (it is the supplier), while in the P2P process the organization is requiring services/goods (customer). Once one company starts a P2P instance, the second company will start an O2C instance, and vice versa.

Figure 10 shows the typical high abstraction level process flow for O2C; as observed for P2P, also O2C can be divided in two different subprocesses: Order Management (OM) and Accounts Receivable (AR). As mentioned for P2P, in a medium/large context the process will likely show high variability.

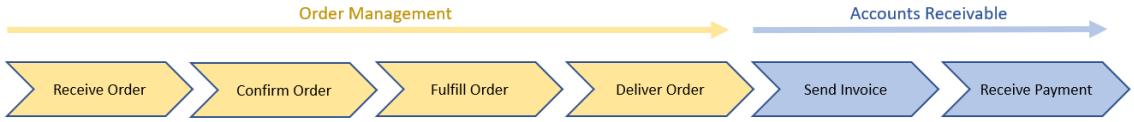

Figure 10 - Typical O2C Process Flow

The main documents involved within the O2C process are:

- Sales Order (SO): it is an internal document created by the seller to process the customer's order. It may be formed by multiple SO Items. The relationship between SOs and SO Items is 1:n.
- Shipping Document: document created internally to process the customer order's shipping.
- Delivery Document: document created internally to process the customer order's delivery.
- Invoice: it is a document recording the business transaction, presented by the company to the customer.

3.4. ANALYSIS

In this subchapter, we will go in depth for each process by providing the context of implementation of the process inside Celonis. Afterwards, it will be presented the data integration necessary to develop an efficient event log, and the activity encoding. After that, the focus will be brought to the evidence shown in the Process Discovery phase. The use cases identified for Conformance Checking and Process Enhancement will also be presented; the solutions' frameworks to tackle the process' gaps will be presented in the Discussion of Results chapter.

3.4.1. Data Integration and Activities Encoding

The first steps for integrating different processes in Celonis require a deep understanding on how the business processes run, who are the people involved in the team, where the necessary data is stored and how to build the necessary connections to be able to extract into Celonis the data from the source system(s). When I joined the company, some processes implementations' were already completed and those projects already were at an intermediate level; during my stay I had the opportunity to see how the processes' implementations in Celonis evolved, and how different use cases were built from scratch.

Once the process walkthrough phase together with the people involved has been carried out, the next step is to extract the data from the source systems into Celonis. To extract SAP table into Celonis, it is first necessary to build the connections between Celonis and SAP. This requires using a pre-built connector set-up, installing an on-premise extractor in the SAP systems, and inserting all the information required to guarantee a safe connection bridge while extracting, and to guarantee that only authorized users can access SAP data; this of course, in a business framework, is vital to preserve the data confidentiality and cyber security.

Once the connection is created and running, the next step is to configure all tables to be extracted, and their relative set-up. For each data connection, a data extraction job is created, and it is possible to select the tables to extract between all tables available for extraction. Within the data job it is possible to use parameters; they are useful to, statically or dynamically, keep track of the main extraction parameters such as starting date, active company codes, and so on.

Within the Celonis extractor data job for each table there are several ways in which the scope of the data to be extracted can be set-up.

First, it is possible to select which columns to extract (by default, all columns are considered), and, when necessary, to modify the combination of fields used to create the table's primary key, or to choose the columns to be pseudo-anonymized. Moreover, it is possible to limit the scope by setting up one or more JOIN configurations with another table(s) (for example, when extracting table A and configuring a JOIN with table B, only the rows from table A having at least a corresponding row in table B are extracted). Those joins are done at SAP level, not on the tables already extracted.

It is possible as well to limit the extraction using a DATE column; we can set up a start date and an end date, and only rows with values of that specific column falling within that timeframe are extracted. This is often used to reduce the scope only to a recent past, to avoid issues in terms of performance and data storage limitations, as well as defining a specific business scope.

Another way to limit the scope is by using the filter statement; this allows to explicitly write SQL-like code, to create a set of filters that will reduce the rows to be extracted to only the rows that adhere to the filter in question (for example, "WERKS = 'ABCDEFG'" will be used to extract only the rows of the table where the plant value is set to the string 'ABCDEFG').

Similar to the filter statement, the delta filter statement is executed only when the extraction is executed in delta mode: this means that instead of refreshing the whole dataset, only the rows being added after the last execution are extracted, and are merged with the rows present before-hand. This allows a faster data refresh and a more dynamic approach to have the data present in the dashboards as updated as possible.

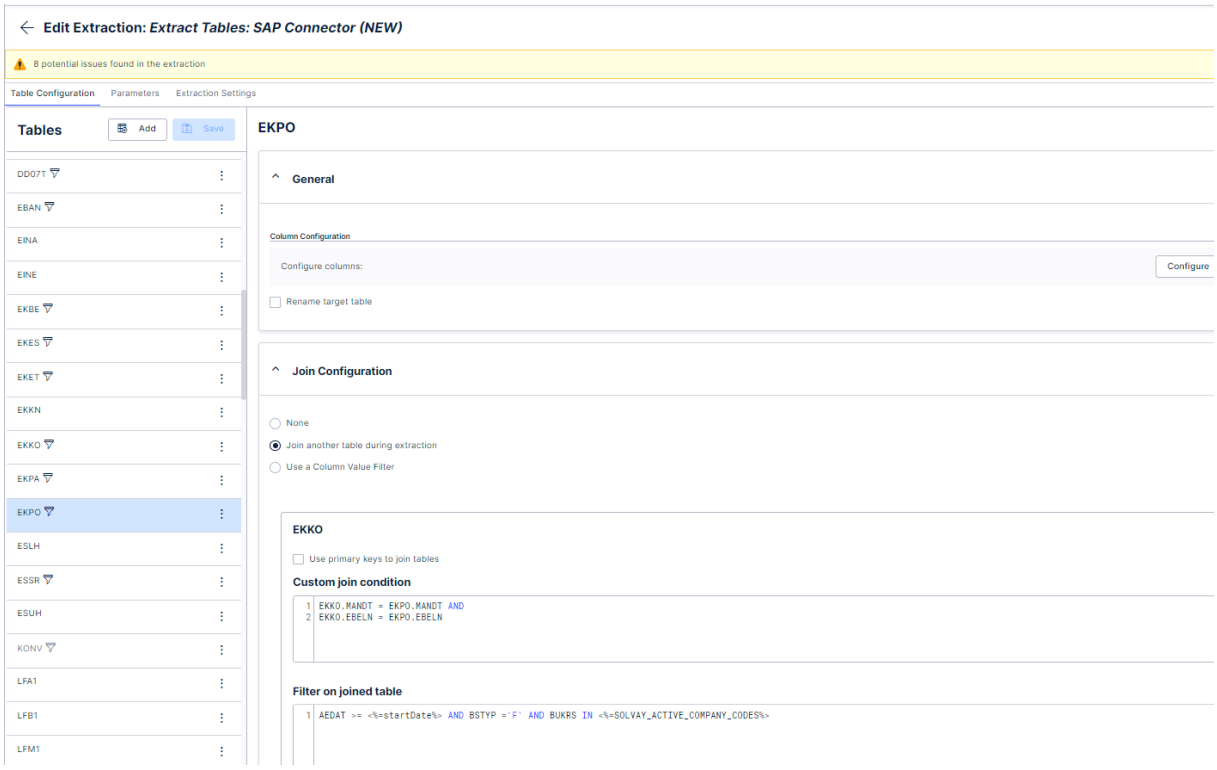

Figure 11 - EKPO Extraction Data Job

In Figure 11, it is possible to observe how the EKPO extraction data job has been set-up. Specifically, only the rows to EKPO having a correspondent in EKKO, joined via the EKKO primary keys (MANDT

being the source system code and EBELN being the Document number) are extracted. In Figure 12 it is shown the filters applied to the EKPO extraction. It is important to notice how, when both a join configuration and a filter statement is present, the first condition that must be met is the join, and only later the filter.

Filter Statement

```
AEDAT >= <%=startDate%> AND BSTYP IN ('F') AND BUKRS IN <%=SOLVAY_ACTIVE_COMPANY_CODES%>
```

Delta Filter Statement

```
AEDAT >= <%=startDate%> AND BSTYP IN ('F') AND BUKRS IN <%=SOLVAY_ACTIVE_COMPANY_CODES%>
```

Figure 12 - EKPO Extraction Data Job filter set-up

Once the data has been extracted, the data is segregated by the system they're extracted from; each system has its own schema, containing the data extracted for the system. This is where unification comes up, in order to bring all data into the same schema; this schema, called the Global Schema, is where once the data is unified, most data pre-processing activities happen. In Figure 13, an example of a SQL unification query is presented. Unification is usually rather simple and requires an individual Data Job, however when tables between different systems are different some challenges might come up to find a suitable table configuration in order to not lose information and at the same time limit data storage as much as possible.

```
Create VIEW "EBAN" AS(
SELECT
'<%=DATASOURCE:SAP_ECC_PF1%>' AS "SCHEMA"
,<%=DATASOURCE:SAP_ECC_PF1%>."EBAN".*
FROM <%=DATASOURCE:SAP_ECC_PF1%>."EBAN"
UNION ALL
SELECT
'<%=DATASOURCE:SAP_ECC_WP1%>' AS "SCHEMA"
,<%=DATASOURCE:SAP_ECC_WP1%>."EBAN".*
FROM <%=DATASOURCE:SAP_ECC_WP1%>."EBAN"
);
```

Figure 13 - EBAN Unification Script

After the unification, the data is to be transformed, in order to create the event logs and the tables that will be part of the data model to be loaded. Prior to doing that, some conceptual work needs to be done to define the lowest level of granularity; the lowest level of granularity will be indicator to which table will be the Case table in the data model. All activities present in the event log will be linked to the cases; this step is crucial to reach the desired outcomes of Process Mining.

In the Transformation Data Job, the activity table is created, and the rest of the data transformation is carried out. The data transformation phase is crucial to have the right information to be included in

the data model, to optimize the data storage performances and to prepare the data to be loaded in the front-end. Transformations, as the rest of the activities in the back-end, are done with SQL queries.

To load the data in the front-end, and make it available to the business users and the analysts, respectively in Apps and Studio, a new Data Job is created. In Figure 14, a screenshot of the data job to load the data models in the P2P data pool is shown. The data load can be done for all tables in the data model, or for a subset of them. The data load will throw an error if some of the Celonis requirements are not met.

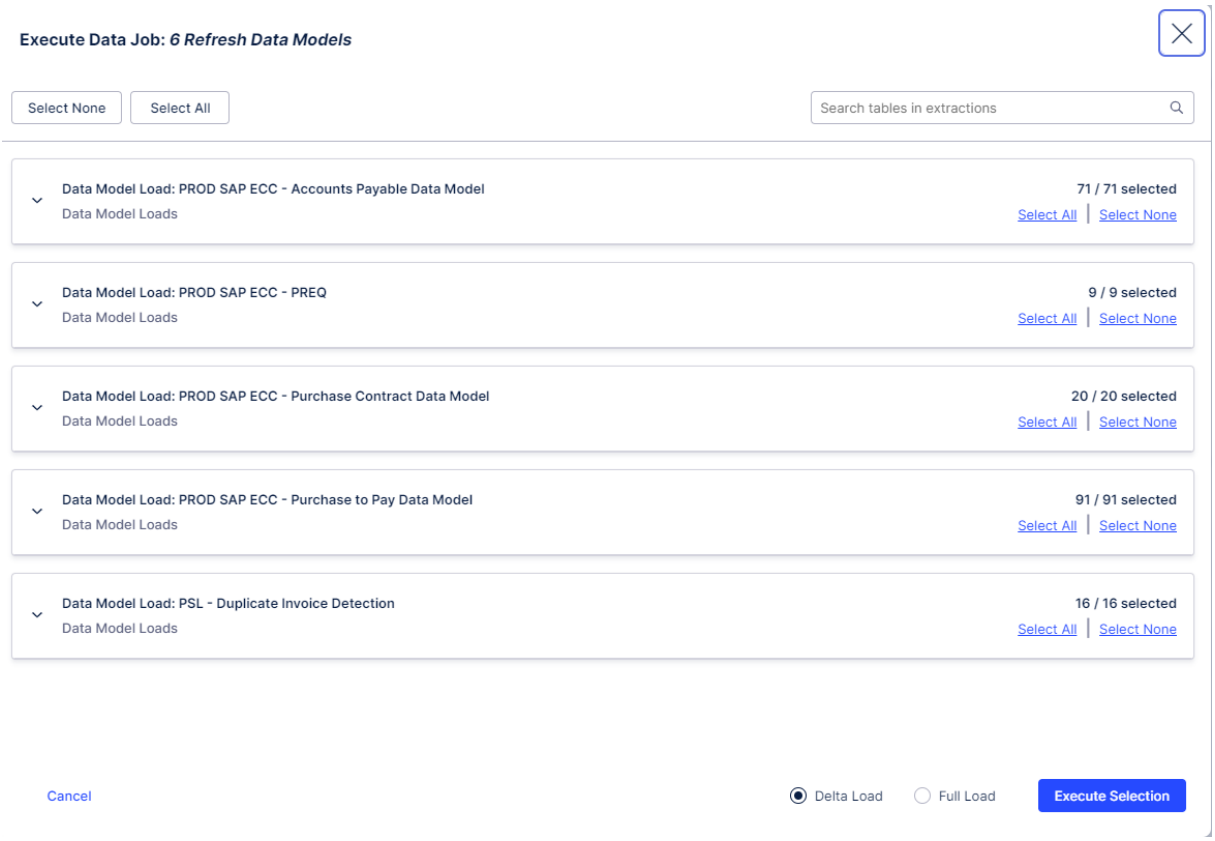

Figure 14 – Data Load for P2P data models

3.4.2. Process Discovery

The Process Discovery is one of the main purposes of Process Mining. In Celonis, there are mainly three components that allow doing that; Process Explorer, Variant Explorer and Case Explorer.

- Process Explorer
  This tool provides a user friendly DFG graph and it is used to explore what are the most common activities and connections between activities in our process. As a default, it presents the most frequent activities and connections; it is possible, however, to add as many activities and connections as there are present. By adding them, it is possible how they impact the overall process. There is a possibility of choosing which KPI to analyze in this graph, such as number of cases flowing through, average throughput time, automation rate. An important feature that this DFG graph offers is the possibility of filtering the graph on two different components: activities and connections. By filtering on activities, only the cases having at least one occurrence of that activity are taken in account. On the other hand, by filtering on a

connection, for example from activity A to activity B, only the cases in which the activity A is directly followed by activity B are considered. This graph, when a significantly large number of activities and connections are added will result in a *spaghetti-like* graph. Figure 18 shows the Process Explorer for the Provisioning process, limited to lowest number of activities and connections, without any filters applied.

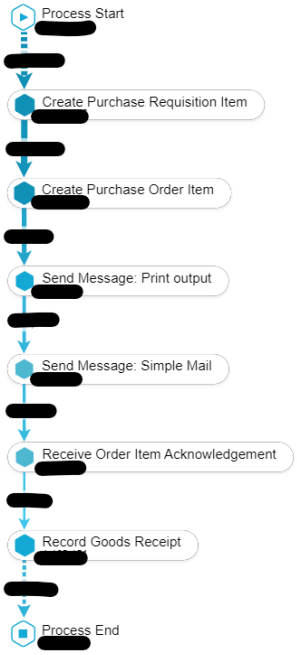

Figure 15 – Process Explorer in the P2P process[1]

- Variant Explorer
  On a first glance, the Variant Explorer might seem the same as the Process Explorer, however there is a significant difference between the two. The Variant Explorer also provides a user-friendly DFG graph; this however shows all process variants within the event logs, order by the most common ones. As in the Process Explorer, it is possible to analyse different KPIs, as well as filtering for specific activities. This analysis is very valuable to business users to have an easy to understand and very powerful way to see how the entire set of cases within the process is going in real time. Figure 19 shows the Variant Explorer in the Provisioning process; on the left there is the DFG graph, while on the right we can select the number of variants to consider, see how many different variants are present and what the percentage of cases covered by the variants in scope is.

[1] The image has been modified due to Company's confidentiality reasons

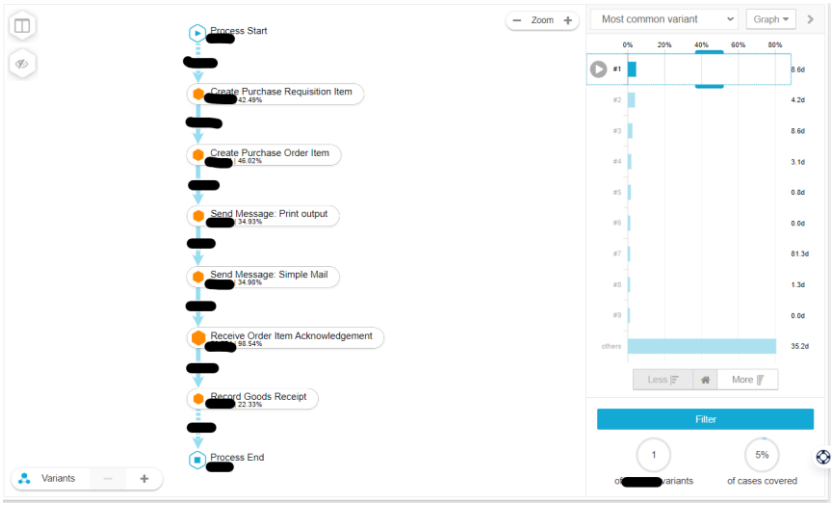

Figure 16 – Variant Explorer in the P2P process[2]

- Case Explorer

  The Case Explorer, differently to the previous two, is a tool for the user to analyse the activity history of a single case; this differs from the Process Explorer and the Variant Explorer since the scope of the analysis is on a single instance in the process, not all instances. This however is a very useful tool to understand on a micro level what the instance's process looks like, and get specific information regarding the instances' respective activities. The output is not anymore a DFG graph, but instead a table containing detailed information for all cases. By clicking on a case, a view on the left appears showing the activity history for the case being clicked on; furthermore, by clicking on the activity in the panel on the left more details regarding the activity show, such as timestamp, username, user type, changed from and changed to values, and all other fields in the activities table.

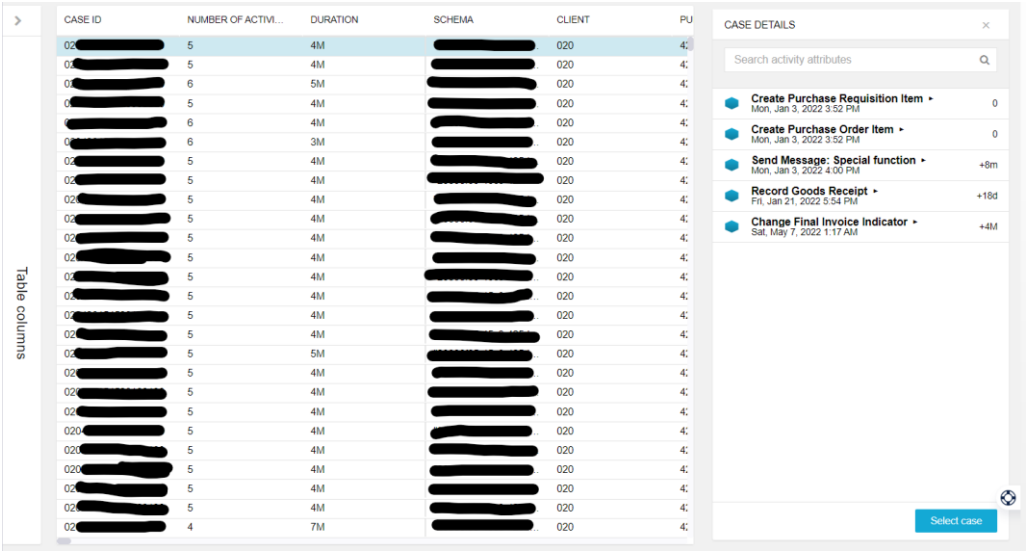

Figure 17 – Case Explorer for the Provisioning process [3]

[2] The image has been modified due to Company's confidentiality reasons
[3] The image has been modified due to Company's confidentiality reasons

3.4.3. Purchase To Pay

The P2P process was the first one to be implemented into Celonis by the team, early 2021. Due to the long-term adoption of the tool, the business users involved in using Celonis to provide value to the company show a great and deep understanding of Process Mining and the experience in using Celonis. Moreover, being the process implemented for a long time it evolved significantly, having many use cases and monitoring dashboards developed during the years.

The business users using this tool are part of the Procurement department of Solvay Business Services; the scope of their analysis therefore involves all GBUs. This translates in a key service provided by the PM team to the company. Being the process part of SBS, processes are standardized, meaning that POs will follow the same processes despite the GBU responsible for the PO.

I personally had the opportunity of following closely development and maintenance projects regarding this particular process.

During the years, the Purchasing Service Line has been using Process Mining via Celonis and our team to work on several use cases; due to limitations regarding quantity of content within this document and some confidentiality reasons, I will limit this chapter by presenting only one example of use cases being implemented.

3.4.3.1. P2P Data Integration and Activities Encoding

For the P2P process, the data is stored within three distinct SAP ECC servers, which from now on will be called WP1, PF1 and PI1; the reason of the distinction of the systems is the history of the adoption of the ERPs from different GBUs. Some GBUs use WP1, the others use PF1, and some functional transaction used are used in PI1. Although the systems are fairly similar, some inconsistencies exist; this inconsistencies often create a challenge to be able to merge the data together.

In Table 3 it is possible to see the list of the most important SAP tables for the P2P process mapping; in reality, the number of these tables is significantly higher, due to the high level of detail needed to map such a critical process, however to simplify the analysis only the main ones are being presented.

Table Technical Name	Description
EKPO	Purchasing Document Item
EKKO	Purchasing Document Header
EBAN	Purchase Requisition
LFA1	Vendor Master (General Section)
MARA	General Material Data
/COCKPIT/THDR	Invoice Header Data
/COCKPIT/TITEM	Invoice Item Data
BKPF	Accounting Document Header

BSEG	Accounting Document Segment
CDHDR	Change Document Header
CDPOS	Change Document Items

Table 3 – Main SAP Tables for P2P Process

All the relevant tables are extracted from the three SAP systems following the steps and the logics shown in chapter 3.4.1.

For the P2P data model, it was decided to use the EKPO table, containing data for Purchasing Document Items, as the central table. This means that all activities in the event log will be at Purchasing Order (PO) Item level; the activities that refer to Purchasing Order Headers, such as 'Create PO Header', will be carried to the Item level as well.

The activity table is then created, and contains multiple fields which understanding is key to deeply understand the following chapters. *MANDT, EBELN* and *EBELP* correspond to the system the EKPO record is coming from, the PO number and the Item number. Those are used to link the activity to the VBAP record; *_CASE_KEY* is a concatenation of those fields. *ACTIVITY_EN* is the descriptive name of the activity. *EVENTTIME* is the timestamp in which the activity took place; in case a PO Items has multiple activities happening at the same timestamp, *_SORTING* is used to chronologically sort which activity took place before, and needs to be defined beforehand for each activity. *USER_NAME* and *USER_TYPE* contain the username whom performed the activity and information whether that user is an automatic or manual user. *CHANGED_TABLE* and *CHANGED_FIELD* contain the technical SAP names of the table and field that were modified during the occurrence of specific activities, and it is useful to keep track of what is being considered. *CHANGED_FROM* and *CHANGED_TO* are used in change activities to know what it was are the before/after values being modified during a change activity. In the Figure below, the SQL script used to create the table is presented, showing all fields in the table.

```sql
--Creation of table Activities "_CEL_P2P_ACTIVITIES"

DROP TABLE IF EXISTS _CEL_P2P_ACTIVITIES;
DROP VIEW IF EXISTS _CEL_P2P_ACTIVITIES;
CREATE TABLE _CEL_P2P_ACTIVITIES (
    "_CASE_KEY" VARCHAR(50)
    ,"MANDT" VARCHAR(3)
    ,"EBELN" VARCHAR(10)
    ,"EBELP" VARCHAR(5)
    ,"ACTIVITY_DE" VARCHAR (300)
    ,"ACTIVITY_EN" VARCHAR(200)
    ,"EVENTTIME" DATETIME
    ,"_SORTING" INT
    ,"USER_NAME" VARCHAR(100)
    ,"USER_TYPE" VARCHAR(10)
    ,"CHANGED_TABLE" VARCHAR(20)
    ,"CHANGED_TABLE_TEXT_DE" VARCHAR(200)
    ,"CHANGED_TABLE_TEXT_EN" VARCHAR(200)
    ,"CHANGED_FIELD" VARCHAR(20)
    ,"CHANGED_FIELD_TEXT_DE" VARCHAR(200)
    ,"CHANGED_FIELD_TEXT_EN" VARCHAR(200)
    ,"CHANGED_FROM" VARCHAR (300)
    ,"CHANGED_TO" VARCHAR(300)
    ,"CHANGED_FROM_FLOAT" FLOAT
    ,"CHANGED_TO_FLOAT" FLOAT
    ,"CHANGE_NUMBER" VARCHAR(50)
    ,"TRANSACTION_CODE" VARCHAR(20)
    ,"_CELONIS_CHANGE_DATE" DATETIME
    ,"_ACTIVITY_KEY" VARCHAR(50)
);
```

Figure 18 – SQL Script used to create _CEL_P2P_ACTIVITIES table

The two SAP change logs' tables are CDPOS and CDHDR; the first one contains information such as the table and the field being changed, what are the before/after values, the SAP transaction that was used. All information regarding the timestamp of the activity and the user are stored in CDHDR.

All activities being defined using the change logs' tables for P2P process follow are Set, Change and Remove activities. When we consider a Set activity, a field that contained a NULL value is changed to something different value. A Change activity happens when a non-NULL value is changed to another non-NULL value; when it is changed to a NULL value, it is considered as a Remove activity.

_CASE_KEY	MANDT	EBELN	EBELP	ACTIVITY_DE	ACTIVITY_EN	EVENTTIME	_SORTING
[redacted]	[redacted]	[redacted]	[redacted]		Change Final Invoice Indicator	2022-10-21 16:01:43	1850
[redacted]	[redacted]	[redacted]	[redacted]		Change Delivery Indicator	2022-10-21 16:01:17	710
[redacted]	[redacted]	[redacted]	[redacted]		Change Final Invoice Indicator	2022-10-21 16:01:43	1850
[redacted]	[redacted]	[redacted]	[redacted]		Change Delivery Indicator	2022-04-02 01:19:15	710

Figure 19 – Fraction of _CEL_P2P_ACTIVITIES table[4]

Once the activity table has been created, as well as the other tables with relevant information, the data modelling phase starts. As mentioned above, the Central table for the data model (also referred to as the Fact or Case table) is EKPO. In Figure 16, it is possible to see the P2P data model currently in use.

[4] The image has been modified due to Company's confidentiality reasons

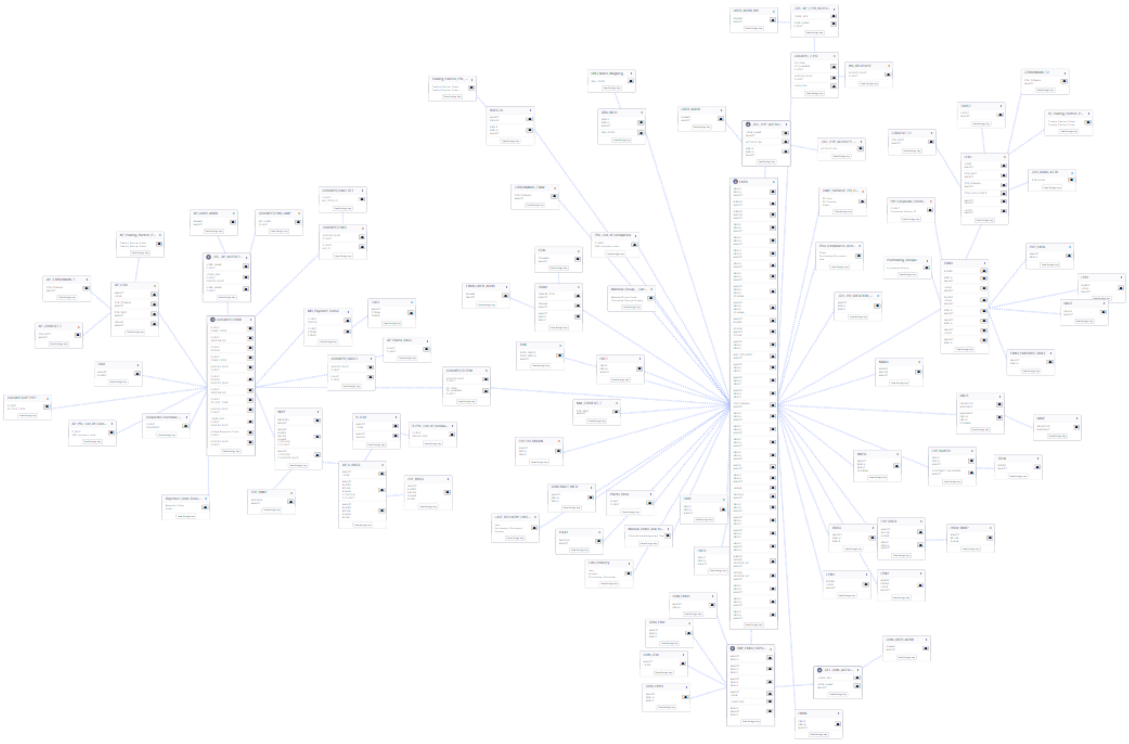

Figure 20 – P2P Data Model

Starting from EKPO, all tables need to be linked with foreign keys to the tables existing in the data model. In Celonis, all connections are exclusively 1:N. When creating a connection between tables, we need to define the fields to use as foreign keys, and which table is at the 1 side and which on the N side of the connection. In Figure 17, we can see how EKPO and _CEL_P2P_ACTIVITIES are linked within the data model. EKPO is at the 1 side of the connection, since each activity refers exclusively to a single PO Item, whilst a PO Item can have multiple activities associated. The two tables are linked via MANDT, EBELN and EBELP, respectively System, PO number and Item number.

Figure 21 – EKPO and _CEL_P2P_ACTIVITIES foreign key settings

Once the data model is been created, the next step involves loading the data to the front-end, in order to have always up-to-date data; this is crucial when running automations and using dashboards.

3.4.3.2. Use Case: Duplicate Invoice Checker

The first use case is called the Duplicate Invoice Checker. This use case is providing a lot of value to the company in terms of money saved; the solution has been implemented to front a common issue that moderately large and large companies are facing, that is receiving duplicate invoices from vendors. Those duplicate invoices are often not detected in the ERP systems, and either are paid twice by the company, resulting in a cash leakage, or there is a lot of manual work to be performed by employees to detect and resolve them, resulting in an inefficient labor productivity. Those invoices typically differ from each other from some fields, such as vendor name or the document date.

Celonis provides a pre-built solution to deal with this issue: it is standard for SAP ECC systems, and needed some customizations to make sure it would follow all the requirements to efficiently work within the Solvay business frame. This solution is a proprietary algorithm analyzing all invoices present in the data model, and flags them as potential duplicates if some conditions, which will be presented below, are met; whenever invoices are identified as possible duplicates, they are grouped together with the similar invoices within a duplicate group. The duplicate group will contain all invoices to be checked among themselves to investigate if they are indeed duplicates, or are different invoices but with similar information.

This Duplicate Checker algorithm consists in a python script, comparing the invoices between themselves on a customizable set of fields; these fields will be needed to map a particular name, such as vendor name, to the technical field in the data model corresponding to the vendor name. Some filters can also be applied, in order to limit this detection process only to the scope of invoices that the business identified as relevant to be included in the use cases. The search patterns can be also defined, to understand where the pattern of identification has been found. In Figure 21, we can see what the current search patterns are. The search pattern “Exact”, will look for invoices having the same vendor name, document date, reference, gross amount and currency. The search pattern “Different_Date”, contrarily to the previous one, will include the invoices having different document dates but the rest of the fields being the same. The different search patterns are very valuable also to have an historic overview on the number of occurrences by each search pattern.

This scripts runs each day, in order to daily have new invoice groups to be analyzed. Once the script runs and the new invoice groups are analyzed, the data loads into an Action View. An Action View, in Celonis, is a tool that empowers business users to not only analyze the data and KPIs, but also to take actions and use it as an operations tool; in this particular Duplicate Detection use case, the users can assign a status to each Invoice. These status reflect on the current state of the checking whether the necessary operations are being done by the team.

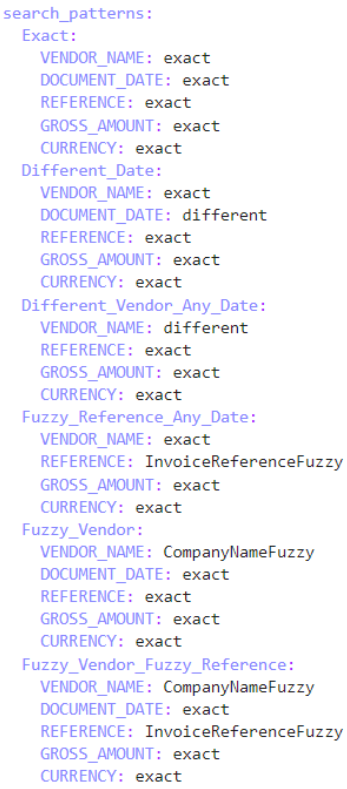

Figure 22 – Search Patterns for Duplicate Detection Checker algorithm

The possible statuses assigned to each invoice are:

- New (default status when the invoice is detected)
- In Progress – No Payment Performed
- In Progress - Payment to be Recovered
- Resolved – False Positive
- Resolved – No Payment Performed
- Resolved – Payment Recovered
- Resolved – Write Off
- Resolved – Original Document

In Figure 22, it is possible to see the Action View. In the part on top, the user can see KPIs such as Total Number of Potential Duplicates (meaning the invoices which status has not been set to Resolved), the total value of those invoices, the total number of invoices that have been analyzed, and the total number of invoices, with the total value as well, of invoices that have been identified as true duplicates.

Below that horizontal tab with the macro KPIs, there is information regarding each of the invoices such as the status, the company code for the invoice, the invoice document number, the duplicate group identifier, and many others. By clicking on a single invoice, there is the possibility of updating the status, as well as the root cause, via a dropdown menu: the root cause is used to provide more detailed information on the reason behind the current status.

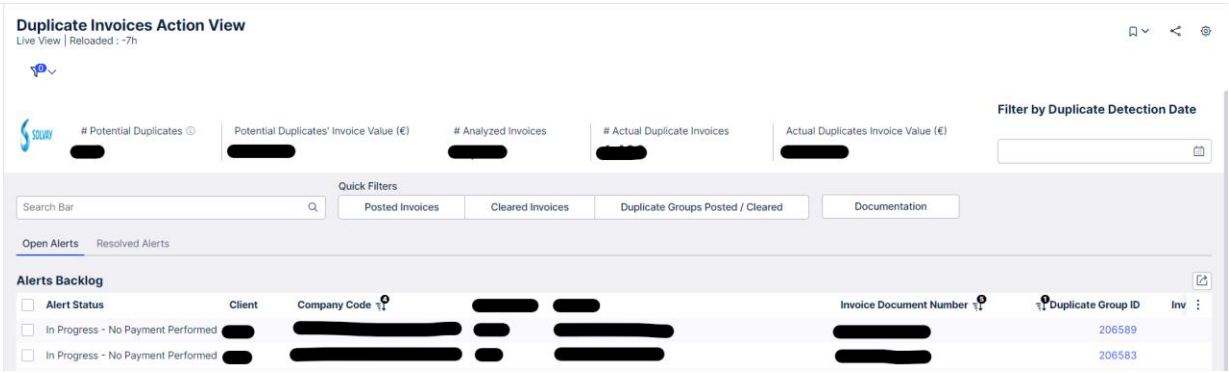

Figure 23 – Duplicate Invoices Action View[5]

The Action View is divided between Open Alerts, to see only the invoices currently in the operations' scope of the team, and Resolved Alerts, to see historic information regarding invoices that already have been analyzed and flagged as Resolved.

To measure the value and provide the users with analytics regarding this use case, a dashboard has been created to allow them to explore in great detail the workload they've been sustaining, the value that Celonis provided to their Service Line, and to find some improvements possibilities. For confidentiality's reasons, an image of the dashboard is not present.

This solution is providing a large value to the team and to the Solvay group in general; however, it can be further developed in the future to expand the utility of the Celonis tool for this specific team to solve this particular business challenge.

3.4.4. Order Management and Accounts Receivable

The O2C process PM implementation in Solvay has been divided in the two sub processes composing the process itself: Order Management and Accounts Receivable. The analysis' scope in the document will be limited to OM due to the higher maturity and work carried out of the project.

The Order Management PM use case has been started together with one GBU's O2C process excellence team in 2021; during the time the project has been growing, two other GBUs decided to use Celonis to improve their processes as well. Most processes are standardized across GBUs, however not on all aspects: this raised the challenge and the work to do in order to carry out customizations, both in event logs creation and use cases. Whenever a new GBU decides to implement Celonis as well, there is the need to go through deep dive sessions to show the AS-IS definitions and collect the requirements to customize definitions that not adhere with the AS-IS ones for the new GBU. Due to privacy, audit, and compliance reasons, especially in preparation for the Po2 project, the OM users' data permissions, belonging to different GBUs, cover only the data of the GBU each user belongs to.

Accounts Receivable (AR) process also presents two business cases implemented and hosted in Celonis: the first one, being Misdirected Payments, aims to simplify the work of people working in AR operations in automatically identifying payments received by customers that contain incorrect

[5] The image has been modified due to Company's confidentiality reasons

information regarding the company or the house bank indicated in the payment, and to act by contacting the customers and provide them with the right payment information. This use case has been developed in preparation of the Po2 project. The second use case aims to monitor and improve the Cash Allocation process. The users of these use cases belong to SBS; the visibility for them is therefore on all GBUs, and the processes across GBUs are standardized.

Due to limitations regarding quantity of content within this document and some confidentiality reasons, I will limit the O2C uses cases by presenting two examples of use cases being implemented.

3.4.4.1. O2C Data Integration and Activities Encoding

For the O2C process, the data is stored within two distinct SAP ECC servers, which from now on will be called WP1 and PF1; the reason of the distinction of the systems is the history of the adoption of the ERPs from different GBUs, similarly to the P2P process. Although the systems are fairly similar, some inconsistencies exist as well in the O2C process; this inconsistencies often create a challenge to be able to merge the data together, or due to the existence of custom fields or tables in only one of the systems.

In Table 4 it is possible to see the list of the most important SAP tables for the O2C process mapping; in reality, the number of these tables is significantly higher, due to the high level of detail needed to map such a critical process, however to simplify the analysis only the main ones are being presented.

Table Technical Name	Description
VBAP	Sales Document: Item Data
VBAK	Sales Document: Header Data
VBEP	Sales Document: Schedule Line Data
KNA1	General Data in Customer Master
LIPS	SD document: Delivery: Item data
LIKP	SD Document: Delivery Header Data
VTTP	Shipment Item
VTTK	Shipment Header
BSEG	Accounting Document Segment
CDHDR	Change Document Header
CDPOS	Change Document Items
VBFA	Sales Document Flow

Table 4 – Main SAP Tables for O2C Process

All the relevant tables are extracted from the two SAP systems following the steps and the logics shown in chapter 3.4.1.

For the O2CC data model, it was decided to use the VBAP table, containing data for Sales Order Items, as the central table. This means that all activities in the event log will be at Sales Order (SO) Item level.

The activity table is then created, with a structure very similar to the P2P activity table, and contains multiple fields which understanding is key to deeply understand the following chapters. *MANDT, VBELN* and *POSNR* correspond to the system the VBAP record is coming from, the SO number and the Item number. Those are used to link the activity to the VBAP record; *_CASE_KEY* is a concatenation of those fields. *ACTIVITY_EN* is the descriptive name of the activity. *EVENTTIME* is the timestamp in which the activity took place; in case a PO Items has multiple activities happening at the same timestamp, *_SORTING* is used to chronologically sort which activity took place before, and needs to be defined beforehand for each activity. *USER_NAME* and *USER_TYPE* contain the username whom performed the activity and information whether that user is an automatic or manual user. *CHANGED_TABLE* and *CHANGED_FIELD* contain the technical SAP names of the table and field that were modified during the occurrence of specific activities, and it is useful to keep track of what is being considered. *CHANGED_FROM* and *CHANGED_TO* are used in change activities to know what it was are the before/after values being modified during a change activity. In the Figure below, the SQL script used to create the table is presented, showing all fields in the table.

```
CREATE TABLE _CEL_O2C_ACTIVITIES (
  _CASE_KEY VARCHAR(50)
  ,ACTIVITY_DE VARCHAR(300)
  ,ACTIVITY_DETAIL_DE VARCHAR(300)
  ,ACTIVITY_EN VARCHAR(200)
  ,ACTIVITY_DETAIL_EN VARCHAR(300)
  ,EVENTTIME DATETIME
  ,_SORTING INT
  ,USER_NAME VARCHAR(80)
  ,USER_TYPE VARCHAR(20)
    ,CHANGED_TABLE VARCHAR(20)
    ,CHANGED_TABLE_TEXT VARCHAR(200)
    ,CHANGED_FIELD VARCHAR(20)
    ,CHANGED_FIELD_TEXT VARCHAR(200)
    ,CHANGED_FROM VARCHAR (200)
    ,CHANGED_TO VARCHAR(200)
    ,CHANGED_FROM_FLOAT FLOAT
    ,CHANGED_TO_FLOAT FLOAT
  ,CHANGE_NUMBER VARCHAR(50)
  ,TRANSACTION_CODE VARCHAR(20)
    ,MANDT VARCHAR(3)
    ,VBELN VARCHAR(10)
    ,POSNR VARCHAR(6)
    ,_ACTIVITY_KEY VARCHAR(50)
);
```

Figure 24 – SQL Script used to create _CEL_O2C_ACTIVITIES table

Most activities being defined using the change logs' tables for O2C process follow are Set, Change and Remove activities, similarly to the P2P process. Some other activities, however, refer to the Sales Document Flow, and few examples are 'Create Delivery', 'Create Shipment', 'Actual Goods Issue Date'.

ACTIVITY_EN	ACTIVITY_DETAIL_EN	EVENTTIME	_SORTING	USER_NAME	USER_TYPE	CHANGED_TABLE	CHANGED_TABLE_TEXT
Create Sales Order Item		2022-02-07 10:26:32	20	[redacted]	•		
System ATP Confirmed		2022-04-21 13:41:30	84	[redacted]	•		
Set Reason for Rejection		2023-08-09 09:56:44	61	[redacted]	•	VBAP	Sales Document: Item Data
Create Delivery		2023-06-16 00:04:06	71	[redacted]	•		
Material Availability Date passed		2023-06-15 00:00:00	150				

Figure 25 – Fraction of _CEL_O2C_ACTIVITIES table[6]

Once the activity table has been created, as well as the other tables with relevant information, the data modelling phase starts. As mentioned above, the Central table for the data model (also referred to as the Fact or Case table) is VBAP. In Figure 26, it is possible to see the O2C data model currently in use.

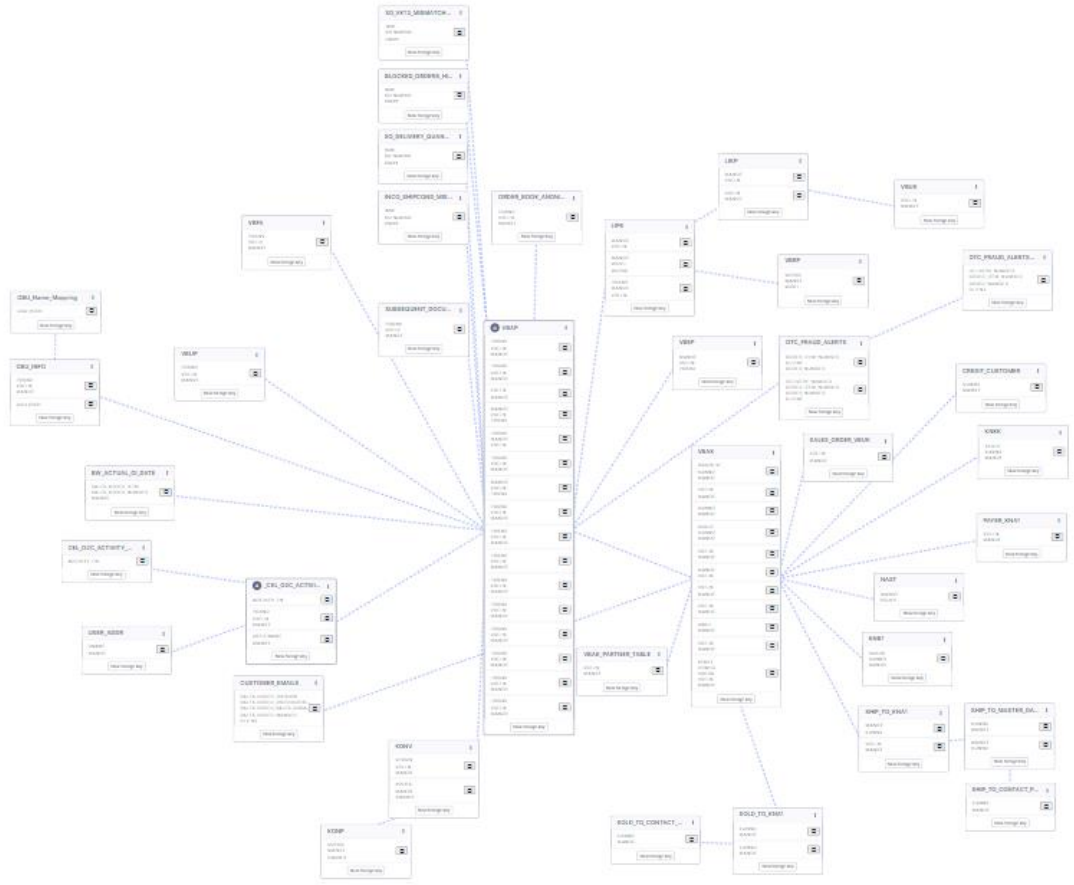

Figure 26 – O2C Data Model

3.4.4.2. Use Case: Perfect Order Rate

This use case was being developed to allow business users to monitor how the Sales Orders are being processed, and to compare the *de facto* processes versus the ideal flow. The business context is that some activities that are being carried out manually by operators are not within the ideal process flow of a Sales Order; these activities will be then considered as “unwanted”, or “blacklisted”. By analysing what ratio of orders has any unwanted activities, and keeping track of the number of unwanted activities for each Sales Order the business users have an opportunity to find improvements

[6] The image has been modified due to Company’s confidentiality reasons

possibilities to enhance the process, and therefore reduce the occurrences of those activities. Each occurrence of these unwanted activities is referred as a “touch”. A perfect Sales Order will therefore have zero touches in its activity history. This use case can be seen as a combination of Compliance and Process Enhancement’s types of Process Mining use.

Given the O2C framework, each GBU has defined the list of activities that count as a touch, as well as some exception rules, to adapt fully to their business needs and to ensure that they focus on their specific pain points. In order to do so, the Process Mining and the GBUs worked together to align on expected results and business rules to be implemented during separate workshops; this can be however seen as a continuous project, since business rules might change, as well as the scope of analysis to be performed. Generally speaking, the activities that are considered as unwanted are activities that are performed manually; those activities indeed require labour work and effect labour productivity.

The whole implementation of this use case has been done within the Studio section of Celonis EMS, and the data model being used is the Order Management one. The different classifications, calculations and KPIs have been developed within the Knowledge Model: the Knowledge Model can be seen as the “brain” of the Studio section of Celonis. The Knowledge Model refers to a single data model, and will be used by different dashboards to be able to call specific KPIs, store variables, apply filters, and many other functionalities, in order to have a centralized source for storing and referring to these elements across various analyses, action views, and automations. All these components (KPIs, filters, variables, records, etc…) are coded in PQL.

The way the number of touches are calculated within each SO Item is quite complex. First, a KPI goes activity by activity and flags the particular activity as either ‘Blacklisted’ or ‘Not Blacklisted’. In Figure 23 it is possible to observe the example of a partial section of this KPI; in particular, in the case when the activity is ‘Change Requested Delivery Date’ or ‘Change Requested Order Quantity’, and the GBU linked to the Sales Order Item is ‘ABCD’, then the activity is considered as ‘Blacklisted’. Once this KPI is built, a second KPI is used to count the number of ‘Blacklisted’ activities for each Sales Order Item: this assigns the total number of touches for each Item. Having this information, we can easily distinguish the Perfect Orders versus the Non Perfect Orders.

```
WHEN
  "GBU_INFO"."GBU_CODE" = 'ABCD' AND
  "_CEL_O2C_ACTIVITIES"."ACTIVITY_EN" IN ('Change Requested Delivery Date', 'Change Requested Order Quantity')
THEN 'Blacklisted'
```

Figure 27 – Partial PQL code for Perfect Order Rate

In Figure 24 it is possible to observe the screenshot of the Perfect Order Rate dashboard. On top some high level KPIs are present, such as the total number of Sales Order Items, the number of Perfect Sales Order Items, the ratio of Perfect Items, the average manual Change activities, the average number of Credit Change activities, and the average Solvay Driven Manual Changes activities per Sales Order Item.

Moreover, graphs are present to help the business user analyse the historical trends of Perfect Order Ratio, as well as the different values split by Region of the Shipment. A drilldown table is present to be able to choose which dimension to be analysed to group the data with, and the user can choose between several dimensions such as Vendor, Material, Activity Group, Sales Organization, and so on. A table is also present at the bottom of the page to have more detailed information regarding the

activities that are considered unwanted, by looking for example if specific patterns exists; if those patterns exist, it could be an indication that the master data might be outdated, and therefore needs an update to reduce manual intervention by the operators.

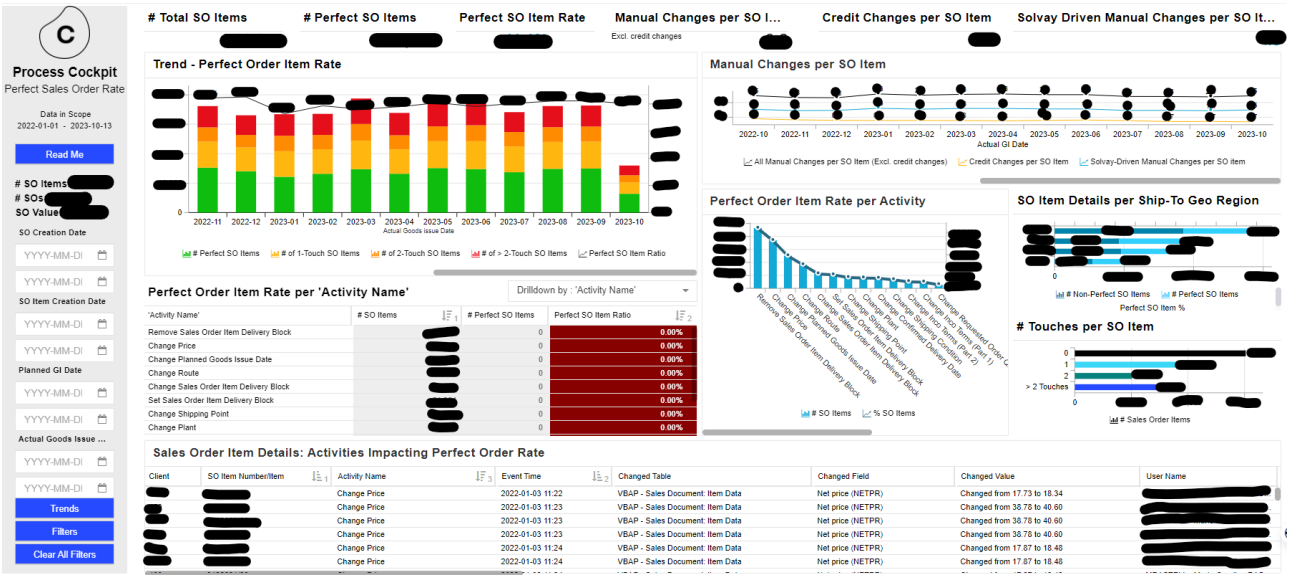

Figure 28 – Perfect Sales Order Dashboard[7]

The value that this use case creates to the business users of the dashboard is therefore having a greater visibility on how their processes are going versus what is the ideal flow of the processes, and giving them the possibility of mining some data to identify what are the root causes that mostly are impacting the Perfect Order Ratio, in order to establish what are the greatest pain points and later establish ways to act on these pain points and therefore improving the performances of their process. Once the Process Enhancement is being applied, the business users can use the dashboard to monitor that the KPIs after the improvement are aligned with expectations.

3.4.4.3. Use Case: On Time Delivery Monitoring

This use case was created to have more visibility on how well the GBUs are performing in terms of delivering the Sales Orders on time. This issue is common in the market, since the On Time Delivery is a shared challenge for most of companies providing goods or services across different markets. The expected goal for this use case is to identify what are the execution gaps resulting in not achieving the On Time Deliveries, and to build some specific automations or controls in order to improve the performances regarding this challenge.

What has been identified during the workshops with the business is the need for the analysis to be highly customizable by the business users; each GBUs measures the On Time Delivery KPIs in different ways, and therefore there is the need to build one analysis that can be used by all GBUs. To do so, a Settings tab in the dashboards has been created for the User to be able to modify the fields to be taken in consideration. In particular, we'll focus on the Due Date and the GI Date. These two fields are used to assess whether a Sales Order Item is considered early, on-time or late. In the dropdowns, the user can choose the settings adequate for its GBU. In the Due Date dropdown, for example, the possible

[7] The image has been modified due to Company's confidentiality reasons

choices are First Confirmed Delivery Date, Last Confirmed Delivery Date, First Confirmed Goods Issue Date, Last Confirmed Goods Issue Date and Requested Delivery Date.

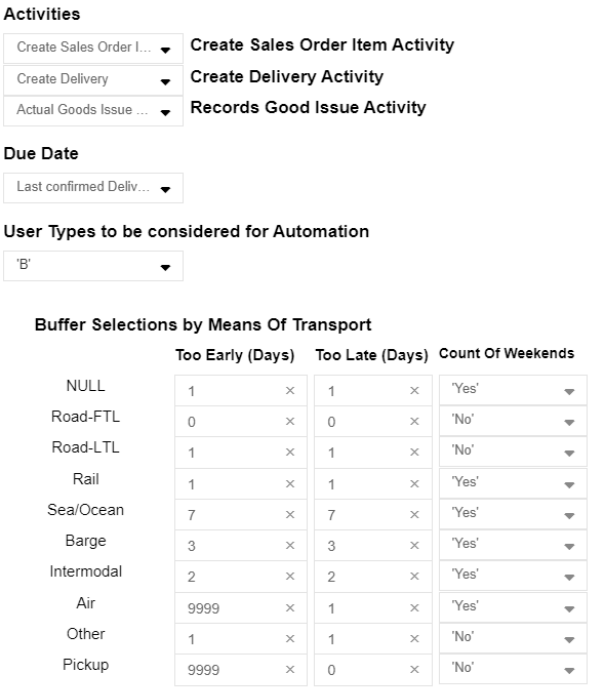

Figure 29 – Settings tab in On Time Delivery Dashboard

Some core concepts about the business need to be explained to fully understand this specific use case.

- Each Sales Order Item can be split between multiple deliveries, and each delivery can be split in multiple shipments. When multiple deliveries and shipments are present, we want to consider the last one in chronological order since we consider a delivery as complete only when everything has been delivered.
- A shipment can be assigned to different mode of transport: these mode of transportations can be AIR, ROAD, SEA, BARGE, PICKUP, and others. When calculating whether a Sales Order is early, on-time or late, a number of days of tolerance is taken in account. This number of days is different for each mode of transport, as well as a flag indicating whether the weekends need to be taken in account or not.
- The Actual Goods Issue Date is the date in which the Goods have left the plants and have been picked up to be delivered (or have been picked up by the Customer).
- The Route is the estimation in terms of number of days of how many days it takes for the shipment to arrive to the customer; by using the Route, we can calculate the Expected Delivery Date as the Actual Goods Issue Date plus the Route.

In Figure 26, it is possible to see what the PQL code for the On Time Delivery looks like. Starting from the first condition, if the Due Date variable is NULL then the Sales Order Item is classified as ‘No Confirmation’. When the Goods Issue Date variable is NULL, depending if the Due Date variable is in the future or in the past, it is classified as ‘No GI and Future Due Date’ or ‘No GI and Passed Due Date’. The Sales Order Item is classified as ’01 – Delivery on time’ whenever the difference between the Due Date and the Expected Delivery Date is within the tolerance, both the early and late one. When the

Sales Order Item is outside of the tolerances, will be classified as '02 – Delivery too early' when it is outside the early tolerance and '03 – Delivery too late' when it is outside of the late tolerance.

Once this classification has been validated and tested, the rest of the dashboard was built to be able to have significant graphs and tables to analyze the performances and identify the root causes. Most of graphs being developed and used are the historical trends of the KPI. A drilldown tab has been created for users to choose the dimension they want to cluster the data by, in order to spot inefficiencies, such as Customer, Ship-To Region, Plant, Carrier, and so on. A Details table has also been implemented in order to have detailed information at Sales Order Item level, and analyze single Items.

One of the root causes identified is the Delivery Blocks that are set up, both manually and automatically, and not released in time; this of course has a significant impact on performance, and is a key root cause to be analyzed and to act upon. To further analyze this topic, an *ad-hoc* tab has been created, to allow users to have deep mining and analysis sessions on Delivery Blocks, and their impact.

```
CASE
    WHEN
      ISNULL(<%=DUE_DATE%>) = 1
    THEN 'No Confirmation'

    WHEN
      ISNULL(<%=GI_DATE%>) = 1 AND DAYS_BETWEEN(<%=DUE_DATE%>,TODAY()) <= 0
    THEN  'No GI and Future Due Date'

    WHEN
      ISNULL(<%=GI_DATE%>) = 1 AND DAYS_BETWEEN(<%=DUE_DATE%>,TODAY()) > 0
    THEN 'No GI and Passed Due Date'

    WHEN
      ROUND(KPI("Days_Between_GI_DD")) >= COALESCE(<%=ROUTE%>*1.0,0.0)
        AND ROUND(KPI("Days_Between_GI_DD")) <= COALESCE(<%=ROUTE%>*1.0,0.0) + KPI("Too_Early_Buffer")
    THEN '01 - Delivery on time'

    WHEN
      ROUND(KPI("Days_Between_GI_DD")) < COALESCE(<%=ROUTE%>*1.0,0.0)
        AND ROUND(KPI("Days_Between_GI_DD")) >= COALESCE(<%=ROUTE%>*1.0,0.0) - KPI("Too_Late_Buffer")
    THEN '01 - Delivery on time'

    WHEN
      ROUND(KPI("Days_Between_GI_DD")) < COALESCE(<%=ROUTE%>*1.0,0.0) - KPI("Too_Late_Buffer")
    THEN '03 - Delivery too late'

    WHEN
      ROUND(KPI("Days_Between_GI_DD")) > COALESCE(<%=ROUTE%>*1.0,0.0) + KPI("Too_Early_Buffer")
    THEN '02 - Delivery too early'

   ELSE '04 - Other'
END
```

Figure 30 – PQL code for On Time Delivery Classification

The value that this use case creates to the business users of the dashboard is therefore having a greater visibility on how the performance in terms of On Time Delivery is going over time, as well frame the value on some possible improvements projects, either inside or outside Celonis, which can be carried out to improve KPIs and Cash Flow. Once the Process Enhancement is being applied, the business users can use the dashboard to monitor that the KPIs after the improvement are aligned with expectations.

3.5. Results and Limitations

The objective of this chapter, as mentioned before, was to show to the reader different ways and use cases in which Celonis has been implemented to front business requirements, specifically to the use

cases presented. As a result, a general yet technical framework has been presented to each use case, as well as the functional requirements. These business cases could be easily scaled to other companies within similar markets, being the challenges present frequent in the market, as well as companies operating in fairly different sectors.

This document will help the academic world by having a direct experience and documentation on what Process Mining can do in a real-case scenario, what the main challenges are, and what the expected results might look like. The document could also be a contribution to the peers in the Process Mining field in the industry, who could take inspiration and replicate the contents of the document.

Further work needs to be done in implementing the literature within this field, due to its rather recent existence and its low volume of literature regarding the real-world applications of Process Mining in the industry.

Due to the document size and the extended range of topics it could have covered, there is a significant limitation regarding possible use cases and implementations that Celonis offers or for which Celonis stood out as a valued solution.

4. DISCUSSION OF RESULTS

In this chapter, a discussion on the value provided by the use cases being implemented via Process Mining and Celonis will be carried out. Due to the company's non-disclosure-agreement, the monetary value won't be included within the document.

4.1. SOLUTIONS IMPLEMENTED BY THE TEAM

As mentioned in the previous chapter, only a small subset of use cases have been described in detail within this document; the results' discussion will not however being limited to those use cases, but all use cases will be analyzed in a qualitative way to explore what is the value that was created by implementing and carrying out Process Mining initiatives within the processes. A special highlight will however be present for the use cases being described in the previous chapter.

4.1.1. Purchase To Pay

The Purchase To Pay processes in Solvay have seen over the years many implementations being done in Celonis; at the moment, the team has built 12 Dashboards, 3 Automations, 1 Machine Learning and 2 Action Views for the Provisioning teams. For the Accounts Payable teams instead, 10 Dashboards, 2 Automations and 5 Action Views have been built.

The high volume of use cases being developed within Celonis are testimonials for the importance and the consideration that said teams allocate to Process Mining; over the years, a significant monetary value has been created for the group from these use cases. The range of use cases' nature range from Process Discovery, in order to give to the teams a more realistic and detailed visibility on what their processes look like, to Compliance, by identifying specific non-compliant behaviors by the users and automatically notifying them in order for the non-compliant tasks to be fixed, to Process Enhancement, helping the teams to identify root causes that were significantly affecting the performance of their figures and processes and fix them in order to have a process re-engineering and help them achieve better results.

The Duplicate Detection use case in particular during 2022 provided around 7,500 cases to the team in charge to analyze said invoices; out of those 7,500 invoices, around 650 of them have been correctly identified as duplicates, and therefore have not been paid twice. The value of those 650 invoices therefore has been saved for the company, and the monetary value created has been very significant and valuable.

4.1.2. Order To Cash

The Order To Cash processes have currently being implemented for three different GBUs, however a testimonial of the importance and the impact that the implementation in Celonis had is the interest by other GBUs to be involved in Process Mining as well, as well as the value brought by the projects to the GBUs currently in scope.

Currently, 8 different dashboards and 6 Automations have been built for all GBUs. In addition to that, some other customized projects have been valuable to GBUs, such as a Material Master Data dashboard and an Order Visibility project, which aims to be an operational tool for users to centralize

actions to be performed and have a greater visibility on the statuses of open orders throughout the Production, Quality and Logistic phases.

The Perfect Order Ratio use case has helped all three of the GBUs to customize what a Perfect Order would look like and what are the unwanted activities happening throughout the process, and to identify inefficiencies within the process; once said inefficiencies are found, a set of tools to identify the root causes is used to enhance the process and achieve a more compliant and reduce the change/rework ratios in order management.

The On Time Delivery dashboard instead has been used to frame the value that could be achieved by implementing new use cases within Celonis, and to prioritize said possible use cases.

4.2. Personal Contribution to the Projects

During the period of the internship within the Process Mining, I had the chance of closely collaborating with multiple projects, processes and teams. This allowed to get a very wide experience with all the projects experience, and a good glance of all the capabilities of Celonis.

For the Purchase to Pay processes I personally developed new elements of pre-existing dashboards, developed from scratch new automations and action view, created a new version of the Machine Learning algorithm used to forecast the Purchase Orders that are delivered late, as well as helped the team on the maintenance and optimization needed within the process data pool. In particular, for Duplicate Detection I had the chance of performing some fine-tuning, as well as checking daily whether all the data pipeline and script running went well, due to its crucial importance.

For the Order to Cash processes, instead, I had the chance of personally implementing a new GBU into the Celonis data model; this required performing all the necessary customization to the process needed for the specific requirements of the GBU. Moreover, I had the chance of creating from scratch multiple dashboards and automations, as well as performing fixes and maintenance tasks, as well as optimizing the backend in terms of data storage space and script runtime. Regarding the Perfect Order Ratio use case, I worked closely with the GBUs to define the customization needed, what the perfect order would look like, and to help them to carry some mining on the data itself. Regarding the On Time Delivery dashboard, instead, I created the dashboard from scratch after collecting the requirements for it to be built.

5. CONCLUSION AND RECOMMENDATIONS

This document aimed to provide an overview on how Process Mining, and Celonis in particular, are implemented, based on the experience collected during the internship carried out in Solvay.

Process Mining is a relatively recent field that aims to combine scientific fields such as BPM and Data Mining, in order to leverage Data Science techniques to obtain its main three objectives related to business processes: Discovering, Monitoring and Improving. The theoretical background, explained in detail in Chapter 2, highlights the state-of-the-art rules to be carefully followed, as well as a broad framework on all the different fields involved in the document.

Chapter 3 focuses on the field work and on a description of the different use case cases, in all the phases involved, starting from the process understanding, to the data extraction and transformation, until the data model creation and use case implementation. In particular, two different processes are discussed in detail: Purchase to Pay and Order to Cash. Within these two processes, three use cases are presented in the document: Duplicate Invoice Detection, On Time Delivery and Perfect Order Ratio. After presenting the use cases, and the solutions provided to face the business requirements, a quantitative and qualitative analysis is done to assess the impact of using Process Mining to tackle these particular business challenges.

Overall it can be concluded that, based on the evidence shown in the literature and in this document, Process Mining can be very beneficial for companies that have the need to improve the visibility on their processes, and can be a very powerful tool to drive digital innovation and process re-engineering, in order to have better efficiency in the business processes.

One limitation encountered during the internship is the challenge of data quality and reliability; the information systems used indeed store data in a non-process centric way. This means that a lot of work needs to be done to ensure data quality, and to transform data from a non-process-centric into process-centric data framework.

Another limitation that can be observed is linked to the Process Mining relatively early age; in some contexts, its capabilities might not be used into the full extent due to low common shared knowledge across business users. It is important to notice, however, how the interest around Process Mining grew and consolidated across time.

Some further research needs to be done in the Process Mining field, in order to explore and deepen the knowledge about the topic, explore further capabilities, gather some more evidence across different markets and industries, and have more people involved into the research and innovation of this field; this will allow Process Mining as a scientific field, and its market stakeholders, to grow and get the recognition it deserves.

BIBLIOGRAPHICAL REFERENCES

Celonis (2023), 2023 Gartner® Magic Quadrant™ for Process Mining Tools, https://www.celonis.com/analyst-reports/gartner-magic-quadrant-2023/

Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics*, *90*(2), 129–149. https://doi.org/10.1016/S0925-5273(03)00102-6

Aguirre, S., Parra, C., & Alvarado, J. (2013). Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach. In P. Cudre-Mauroux, P. Ceravolo, & D. Gašević (Eds.), *Data-Driven Process Discovery and Analysis. SIMPDA 2012. Lecture Notes in Business Information Processing, vol 162*. https://doi.org/https://doi.org/10.1007/978-3-642-40919-6_2

Batista, E., & Solanas, A. (2019). Skip Miner: Towards the Simplification of Spaghetti-like Business Process Models. *10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–6. https://doi.org/10.1109/IISA.2019.8900713

Becker, J., Rosemann, M., & Von Uthmann, C. (2000). Guidelines of Business Process Modeling. In W. van der Aalst, J. Desel, & A. Oberweis (Eds.), *Business Process Management. Lecture Notes in Computer Science* (Vol. 1806, pp. 30–49). Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/3-540-45594-9_3

Bigui, R. ', & Cho, C. (2017). The state-of-the-art of business process mining challenges. In *Int. J. Business Process Integration and Management* (Vol. 8, Issue 4).

Bose, R. P. J. C., Mans, R. S., & van der Aalst, W. M. P. (2013). Wanna improve process mining results? *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 127–134. https://doi.org/10.1109/CIDM.2013.6597227.

Derning, W. E. (1953). Statistical techniques in industry. *Advanced Management*, *18*(11), 8–12.

Diba, K., Batoulis, K., Weidlich, M., & Weske, M. (2020). Extraction, correlation, and abstraction of event data for process mining. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 10, Issue 3). Wiley-Blackwell. https://doi.org/10.1002/widm.1346

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of Business Process Management. In *Fundamentals of Business Process Management*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33143-5

Geyer-Klingeberg, J., Nakladal, J., Baldauf, F., & Veit, F. (2018). *Process Mining and Robotic Process Automation: A Perfect Match*.

Grimaila, M. R., Myers, J., Mills, R. F., & Peterson, G. (2012). Design and Analysis of a Dynamically Configured Log-based Distributed Security Event Detection Methodology. *Journal of Defense Modeling and Simulation*, *9*(3), 219–241. https://doi.org/10.1177/1548512911399303

Grisold, T., Mendling, J., Otto, M., & vom Brocke, J. (2021). Adoption, use and management of process mining in practice. *Business Process Management Journal*, *27*(2), 369–387. https://doi.org/10.1108/BPMJ-03-2020-0112

Hammer, M. (2015). What is business process management? In *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems* (pp. 3–16). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-45100-3_1

Hammer, M., & Champy, J. (1993). Reengineering the Corporation: A Manifesto for Business Revolution. In *Total Quality* (Issue 3). Harper Collins, New York.

Hashmi, M., Governatori, G., Lam, H. P., & Wynn, M. T. (2018). Are we done with business process compliance: state of the art and challenges ahead. *Knowledge and Information Systems*, *57*(1), 79–133. https://doi.org/10.1007/s10115-017-1142-1

Hompes, B. F. A., Verbeek, H. M. W., & van der Aalst, W. M. P. (2015). Data-Driven Process Discovery and Analysis. In P. Ceravolo, B. Russo, & R. Accorsi (Eds.), *Data-Driven Process Discovery and Analysis: 4th International Symposium* (Vol. 237). Springer International Publishing. https://doi.org/10.1007/978-3-319-27243-6

Jalali, A. (2020). Graph-Based Process Mining. In S. Leemans & H. Leopold (Eds.), *Process Mining Workshops* (pp. 273–285). https://doi.org/https://doi.org/10.1007/978-3-030-72693-5_21

Kharbili, M. E., A. K. A. D., Stein, S., & van der Aalst, W. M. (2008). Business process compliance checking: Current state and future challenges. *Modellierung Betrieblicher Informationssysteme*, *(MobIS 2008)*, 107–113.

Kluza, K., Wisniewski, P., Jobczyk, K., Ligeza, A., & Mroczek, A. S. (2017). Comparison of selected modeling notations for process, decision and system modeling. *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017*, 1095–1098. https://doi.org/10.15439/2017F454

Mannhardt, F., de Leoni, M., Reijers, H. A., Aalst, W. M. P. van der, & Toussaint, P. J. (2018). Guided Process Discovery – A pattern-based approach. *Information Systems*, *76*, 1–18. https://doi.org/10.1016/j.is.2018.01.009

Object Management Group. (2013). *Business Process Model and Notation (BPMN) 2.0.2*. https://www.omg.org/spec/BPMN

Peterson, J. L. (1977). Petri Nets. *ACM Computing Surveys*, *9*(3), 223–252. https://doi.org/10.1145/356698.356702

Qafari, M. S., & van der Aalst, W. (2020). Root Cause Analysis in Process Mining Using Structural Equation Models. In A. Del Río Ortega, H. Leopold, & F. M. Santoro (Eds.), *Business Process Management Workshops* (Vol. 397). Springer International Publishing. https://doi.org/10.1007/978-3-030-66498-5

Rudnitckaia, J., & Humby, C. (2014). *Process Mining. Data science in action*.

Sani, F. M. (2020). Preprocessing Event Data in Process Mining. *International Conference on Advanced Information Systems Engineering*.

Suriadi, S., Andrews, R., ter Hofstede, A. H. M., & Wynn, M. T. (2016). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems*, *64*, 132–150. https://doi.org/10.1016/j.is.2016.07.011

Szelągowski, M. (2018). Evolution of the BPM Lifecycle. *Communication Papers of the 2018 Federated Conference on Computer Science and Information Systems*, *17*, 205–211. https://doi.org/10.15439/2018f46

Taymouri, F., Rosa, M. La, Dumas, M., & Maggi, F. M. (2021). Business process variant analysis: Survey and classification. *Knowledge-Based Systems*, *211*. https://doi.org/10.1016/j.knosys.2020.106557

Tjahjono, B., Ball, P., Vitanov, V. I., Scorzafave, C., Nogueira, J., Calleja, J., Minguet, M., Narasimha, L., Rivas, A., Srivastava, A., Srivastava, S., & Yadav, A. (2010). Six sigma: A literature review. *International Journal of Lean Six Sigma*, *1*(3), 216–233. https://doi.org/10.1108/20401461011075017

van der Aalst, W. (2016). Getting the Data. In *Process Mining* (pp. 125–162). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4_5

Van Der Aalst, W., Adriansyah, A., Alves De Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., Chandra Bose, J., Van Den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., De Leoni, M., … Wynn, M. (2012). Process Mining Manifesto. *Lecture Notes in Business Information Processing*, 169–194. https://doi.org/https://doi.org/10.1007/978-3-642-28108-2_19

van der Aalst, W. M. P. (1998). The Application of Petri Nets to Workflow Management. *Journal of Circuits, Systems and Computers*, *08*(01), 21–66. https://doi.org/10.1142/s0218126698000043

van der Aalst, W. M. P. (2011). Process Mining. In *Process Mining*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19345-3

Van Der Aalst, W. M. P. (2018). *Process Mining and Simulation: a match made in heaven! In SummerSim*, 4–1. https://doi.org/10.22360/summersim.2018.scsc.005

Van Der Aalst, W. M. P. (2019). A practitioner's guide to process mining: Limitations of the directly-follows graph. *Procedia Computer Science*, *164*, 321–328. https://doi.org/10.1016/j.procs.2019.12.189

Von Rosing, M., White, S. A., Cummins, F., & De Man, H. (2014). Business process model and notation-BPMN. In *The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM* (Vol. 1, pp. 429–453). Elsevier Inc. https://doi.org/10.1016/B978-0-12-799959-3.00021-5

White, S. A. (2004). Introduction to BPMN. *Ibm Cooperation*, *2*(0).

Zairi, M. (1997). Business process management: A boundaryless approach to modern competitiveness. *Business Process Management Journal*, *3*(1), 64–80. https://doi.org/10.1108/14637159710161585