Bridging the Gap: A Decision Framework for Traditional Process Mining vs. Object-Centric Process Mining

Dominik Heinemann

June 28, 2025

Abstract

As business processes grow more complex and interconnected, organizations face increasing pressure to choose the right analytical tools to understand their operations. Traditional Process Mining (TPM), which relies on case-centric event logs, has long been the standard approach due to its tool maturity and ease of use. However, it often struggles to capture the nuances of multi-entity systems. Object-Centric Process Mining (OCPM) offers an alternative by preserving relationships between multiple object types, enabling more detailed insights into coordination and concurrency.

This thesis explores the comparative strengths and limitations of TPM and OCPM through a mixed-methods approach. A literature-based framework was developed and applied to two contrasting datasets: a structured administrative workflow from the BPI Challenge 2017 and a complex, dynamic object-centric log derived from Age of Empires II game telemetry. The comparison focused on eight analytical dimensions, including scalability, model complexity, and interpretability.

Based on these findings, a decision framework is proposed to help practitioners identify which technique is more suitable for their context. While TPM remains a strong option for straightforward processes and fast implementation, OCPM proves advantageous in capturing inter-object interactions and revealing deeper insights in complex environments. The framework aims to support more informed, case-specific method selection in both academic and applied process mining work.

KEYWORDS: Process Mining; Object-Centric Process Mining; Traditional Process Mining; Decision Framework

Contents

1 Introduction **5**
1.1 Research Questions . . . 5

2 Literature review **6**
2.1 Process Mining . . . 6
2.1.1 Bridging The Gap Between Process Science and Data Science . . . 6
2.1.2 Business Process Lifecycle . . . 7
2.1.3 The Three Categories of Process Mining . . . 8
2.1.4 Process Models: Notation, Purpose, and Applications . . . 9
2.1.5 Difference to Data Mining . . . 11
2.2 Object-Centric Process Mining . . . 12
2.2.1 Object-Centric Process Mining vs. Traditional Process Mining . . . 13
2.3 Event Logs . . . 14
2.3.1 The Role of Information Systems . . . 15
2.3.2 Object-Centric Event Logs . . . 15
2.3.3 The Potential for Process Mining . . . 16
2.4 Case Studies and Rationale for This Approach . . . 16
2.5 Decision Making in Method Selection . . . 17

3 Methodology **18**
3.1 Data Collection . . . 18
3.2 Dataset Background and Context . . . 18
3.3 Tool Selection . . . 19
3.4 Case Selection and Setup . . . 20
3.5 Comparative Framework Design and Evaluation Approach . . . 20
3.5.1 Dimension Definitions and KPIs . . . 21
3.6 Data Preprocessing . . . 22
3.7 Evaluation Approach . . . 23

4 Case Study **24**
4.1 Data Preparation and Model Generation . . . 24
4.2 Process Visualization . . . 24
4.3 Perspective Flexibility . . . 28
4.4 Analytical Depth . . . 29
4.5 Usability and Tool Support . . . 29
4.6 Summary . . . 30

5 Data Analysis and Results **30**
5.1 Comparative Evaluation 30
5.2 Observed Strength and Limitations 31

6 Decision Framework **32**
6.1 Framework Rationale 32
6.2 Decision Criteria 32
6.3 Tree-Based Decision Model 33
6.4 Technique and Tool Recommendation Table 33
6.5 Summary Table 34
6.6 Use in Practice 35

7 Discussion **35**
7.1 Key Findings and Interpretation 36
7.2 Implications for Practice 36
7.3 Limitations 37
7.4 Future Research Directions 38

8 Conclusion **38**

Bibliography **40**

Appendices **43**

A GPT-4 Labeling Prompt and Output **43**

B Python Code Snipped **45**

C Declaration of use of AI **45**

1 INTRODUCTION

As organizations strive to decode the complexity behind their daily operations, process mining has emerged as a flashlight in the dark, illuminating what really happens behind the curtain of business processes.

In today's data-driven landscape, organizations increasingly rely on process analytics to understand and enhance their operations. Among the most established techniques is process mining, which leverages event logs to reconstruct actual workflows. This enables the detection of bottlenecks, inefficiencies, and deviations that are often overlooked by traditional methods.

Yet, as organizational processes grow more interconnected and complex, the limitations of Traditional Process Mining (TPM) become more apparent. TPM operates under a single-case assumption, meaning each event is associated with one case identifier. While effective for straightforward workflows, this approach can oversimplify scenarios where multiple entities interact within and across events.

In response to these limitations, Object-Centric Process Mining (OCPM) has gained traction. Unlike TPM, OCPM preserves the relationships between multiple entities within event logs, offering a more accurate depiction of real-world processes. Although this method enables richer insights, it also introduces greater complexity in terms of tooling, preprocessing, and model interpretation.

This thesis is motivated by a practical question that remains underexplored: when should analysts choose TPM over OCPM, and vice versa? While the literature outlines the theoretical benefits of OCPM, there is little concrete guidance for practitioners tasked with selecting an approach based on their specific process and data context. This thesis addresses that gap by developing a decision framework that balances analytical capabilities with real-world constraints such as tool support, interpretability, and time-to-insight.

1.1 Research Questions

To address this gap, the following research questions guide this thesis:

- **RQ1:** What are the comparative strengths and limitations of Traditional vs. Object-Centric Process Mining in analyzing complex business processes?
- **RQ2:** How can a structured decision framework support the selection of an appropriate process mining approach based on process complexity and data characteristics?

By answering these questions, the thesis aims to contribute both conceptually and practically. On the one hand, it offers a comparative analysis of TPM and OCPM grounded in real-world data. On the other, it presents a decision framework to guide process analysts in selecting the right technique based on the structure and goals of their process mining initiatives.

2 LITERATURE REVIEW

This literature review provides the theoretical background for this study. It outlines the fundamentals of process mining, explains the main differences between traditional and object-centric approaches, and discusses the role of event logs in representing real process behavior. In addition, it describes the rationale for using case studies and introduces key concepts in decision making for method selection. Together, these insights support the comparative analysis and the decision framework developed in the following chapters.

2.1 Process Mining

The distinction between traditional process mining and object-centric process mining has gained practical importance as businesses increasingly manage complex processes involving multiple interconnected entities. Companies face growing challenges when selecting suitable techniques to analyze their operational data. Understanding the strengths and limitations of each approach, therefore, becomes essential to effectively capture and analyze real-world process complexities and ultimately enhance decision-making.

2.1.1 Bridging The Gap Between Process Science and Data Science

Process mining is a transformative discipline that balances the theoretical rigor of process science and the empirical power of data science. Founded upon the confluence of information technology and management sciences, process science seeks to optimize operational processes through trading off technical feasibility and organizational objectives (Van Der Aalst 2016). Unlike traditional process analysis that was often founded upon idealized models, the availability of event logs within modern information systems has enabled data-driven insights into actual process behavior. As Van Der Aalst (2023*b*) highlights, process mining techniques unlock actionable information from these logs, offering capabilities to discover, monitor, and improve processes in a wide range of domains.

This cooperation is not technological but philosophical. Van Der Aalst (2016) draws a parallel between process mining and the "yin and yang" of process science, where data-driven discovery and model-centric design coexist in a symbiotic state. While classical process mining has focused on case-centric analyses, new developments address the complexities of actual systems. For example, Object-Centric Process Mining (OCPM) goes beyond the classical boundaries by representing interactions among various entities (orders, invoices, and suppliers), thus recording the "fabric" of intertwined processes (Van Der Aalst 2023*c*, Berti, Montali & Van Der Aalst 2023). These developments are in line with the general move towards data-centricity, where artificial intelligence and machine learning improve pattern recognition and predictive features (Jarrahi et al. 2023, De Leoni et al. 2016*a*).

The consequences are profound. In healthcare, for example, the transformation of multi-dimensional data into object-based event logs has enabled combined analyses of patient care, resource allocation, and compliance (Park et al. 2024). Similarly, designs like IN-EXA (Benzin et al. 2024) demonstrate how abstraction mechanisms preserve vital information while condensing complex models into simpler representations for stakeholders to comprehend. These developments emphasize process mining's role as a bridge: it transforms raw data into process intelligence and maintains theoretical models strongly rooted in empirical facts.

2.1.2 Business Process Lifecycle

One way to optimize organizational workflows is the Business Process Management (BPM) lifecycle as it provides a structured framework through iterative phases of design, execution, monitoring, and refinement (Van Der Aalst 2016). Traditional BPM methodologies use sequential stages, beginning with process design and continuing through implementation, enactment, adjustment, and diagnosis. Modern approaches move towards the integration of process mining and AI-driven techniques to address the dynamic complexity of real-word systems.

During the design phase, processes are traditionally conceptualized through simplified, idealized models. However, advancements in automated Business Process Management (ABPM) now incorporate machine learning techniques, enabling data-driven creation of process designs based on historical event logs (Paschek et al. 2017). For example, AI-driven algorithms can analyze past activities to pinpoint inefficiencies or potential compliance issues early on, thus minimizing reliance on fixed assumptions (Jarrahi et al. 2023).

In the configuration and implementation phase, conceptual models are translated into operational practices. Here, object-centric process mining (OCPM) further enriches implementation by explicitly mapping interactions among different entities such as orders, invoices, and suppliers, ensuring the models accurately represent real-world complexity and interdependencies (Berti, Montali & Van Der Aalst 2023, Berti, Jessen, Park, Rafiei

& Van Der Aalst 2023).

During the enactment and monitoring phase, object-centric event logs provide deeper, more nuanced insights into process performance. Approaches like the INEXA framework (Benzin et al. 2024) support stakeholders by interactively abstracting complex process models, preserving essential details while simplifying visualization. This becomes especially valuable in intricate sectors such as healthcare, where examining multiple interconnected elements (like patient paths and resource utilization) can expose systemic inefficiencies (Park et al. 2024).

Adjustments within the diagnosis and requirements phase are increasingly moving beyond static or predetermined controls. Methods such as feature extraction from detailed, object-centric logs (Berti et al. 2024) and frameworks for correlation analysis (De Leoni et al. 2016*a*) facilitate dynamic responses by linking specific process behaviors like delays, cost variations, or compliance issues to their contextual circumstances. For instance, predictive analytics demonstrated by (De Leoni et al. 2016*a*) can foresee delays within administrative processes, allowing organizations to proactively adjust rather than reacting after the fact.

Ultimately, the iterative nature of the process lifecycle gains strength from AI-supported knowledge management practices (Jarrahi et al. 2023). Evolving requirements, prompted by regulatory updates or shifts in market dynamics, can now be detected sooner through ongoing log analysis, enabling proactive redesign instead of mere reactive corrections. This close integration of BPM with process mining indicates a significant shift from inflexible, purely model-driven approaches toward a more adaptive, data-informed evolution.

2.1.3 The Three Categories of Process Mining

Process mining techniques generally fall into three categories: discovery, conformance checking, and optimization, each addressing distinct challenges in operational analysis. Discovery involves extracting process models directly from event logs, creating a digital twin that mirrors the actual process. A digital twin is a virtual replica of the real process, built from actual event data, which shows how tasks are carried out step by step, including variations and interactions that occur in practice. This digital twin lets organizations visualize workflows from various angles, such as control-flow, resource distribution, or interactions between objects (Van Der Aalst 2016). For example, object-centric discovery techniques (Van Der Aalst 2023*b*) capture relationships among entities like orders, invoices, and suppliers, preventing oversimplifications typical of traditional, case-centric models. A comprehensive benchmark of process discovery algorithms by Augusto et al. (2018) provides an empirical comparison of traditional techniques such as Inductive Miner (IM), Evolutionary Tree Miner (ETM), and Split Miner (SM). Their study evaluates each method against quality dimensions including fitness, precision, generalization, complexity, and soundness. While IM excels in fitness, ETM performs best in precision, and SM offers a strong tradeoff with high F-scores and fast runtime. These findings highlight the strengths and limitations of commonly used algorithms in traditional process mining and reinforce the need for practical frameworks to support method selection. This thesis builds upon these insights by proposing a decision framework that aligns mining technique selection with specific data and process characteristics.

Conformance checking assesses how closely real-world execution aligns with the digital twin. By highlighting deviations, such as unauthorized process steps or compliance breaches, organizations gain practical insights into operational weaknesses. Frameworks such as the one proposed by De Leoni et al. (2016*a*) utilize correlation analysis to identify underlying causes of these deviations, such as delays caused by specific resource limitations. In healthcare contexts, object-centric conformance checking has uncovered systemic inefficiencies in patient care processes by correlating treatment stages with lab outcomes and staffing arrangements (Park et al. 2024).

Optimization builds upon these insights by simulating improvements directly on the digital twin. Techniques like interactive abstraction (Benzin et al. 2024) simplify intricate process models (e.g., condensing a manufacturing procedure from 1,489 components to a clear 58-element visualization) while retaining essential details for testing scenarios. Simulation tools prioritize interventions based on predicted benefits, like cost reduction or improved throughput. For example, extracting features from object-centric logs (Berti et al. 2024) allows machine learning models to forecast delays in administrative workflows, guiding proactive redesign efforts (Khayatbashi et al. 2024).

2.1.4 Process Models: Notation, Purpose, and Applications

Process models serve as vital tools for representing, analyzing, and improving business processes. They use specific notations to describe activities, their causal relationships, and other essential characteristics (Van Der Aalst 2016). The primary purpose of a process model is to clearly outline the sequence and conditions under which activities should be performed (Van Der Aalst 2016, p. 58).

Typical notations used in process models include:

- Activities and subprocesses
- Causal relationships and the ordering of activities

- Temporal constraints
- Data generation and utilization (such as decision modeling)
- Interactions of resources within the process

These notational elements allow for the depiction of sequential, optional, parallel, and repetitive activities within a process (Van Der Aalst 2016). Advanced approaches like Object-Centric Process Models (OCPM) capture more intricate relationships among objects, offering a comprehensive view of interconnected processes (Berti, Montali & Van Der Aalst 2023).

Process models fulfill various critical roles in organizational settings:

1. **Insight Generation**: Creating models reveals different perspectives on processes, uncovering hidden complexities (Van Der Aalst 2016).
2. **Stakeholder Communication**: Process models provide structured means for discussion among stakeholders, enhancing alignment and mutual understanding (Van Der Aalst 2016).
3. **Documentation and Compliance**: They serve as formal documentation for training, certification, and compliance, such as ISO 9000 standards (Van Der Aalst 2016).
4. **Verification and Error Detection**: Model analysis helps identify potential system errors, like deadlocks or bottlenecks (Van Der Aalst 2016).
5. **Performance Analysis**: Techniques such as simulation help organizations understand factors affecting response times and service levels (Van Der Aalst 2016). More sophisticated methods, such as those described by De Leoni et al. (2016*a*), can correlate different process aspects to gain deeper performance insights.
6. **Scenario Planning**: Process models allow users to explore hypothetical scenarios, providing valuable feedback for designers (Van Der Aalst 2016).
7. **System Specification**: These models act as clear agreements between developers and end-users or management, defining requirements for Process-Aware Information Systems (PAIS) before actual implementation (Van Der Aalst 2016).
8. **System Configuration**: Models guide the setup and configuration of IT systems to align closely with actual organizational processes (Van Der Aalst 2016).

9. **Process Mining and Improvement**: Modern techniques like OCPM facilitate discovering complex, multi-object process models from event logs, directly supporting ongoing process enhancements (Berti, Jessen, Park, Rafiei & Van Der Aalst 2023).

10. **Abstraction and Simplification**: Tools such as INEXA enable interactive exploration of process models at different granularity levels, improving comprehensibility of complex processes (Benzin et al. 2024).

By effectively applying these models, organizations gain richer insights into their operations, enhance stakeholder communication, and foster continuous improvement. Transitioning from traditional process modeling to object-centric approaches significantly boosts accuracy in capturing and managing complex, interconnected processes (Berti, Montali & Van Der Aalst 2023, Berti, Jessen, Park, Rafiei & Van Der Aalst 2023).

2.1.5 Difference to Data Mining

While process mining and data mining both focus on extracting valuable insights from data, their core objectives and methods differ notably. Data mining typically involves analyzing large datasets to uncover hidden patterns, summarize information, and present it in ways that are meaningful and useful to stakeholders (Hand 2007). However, unlike process mining, data mining does not usually concentrate explicitly on business processes.

The main distinctions between these two approaches are:

- **Focus**: Process mining specifically targets business processes and their execution, whereas data mining broadly searches for general data patterns without a process-oriented perspective.

- **Model Discovery**: Process mining uniquely enables the automatic generation of process models from event logs, a capability typically not available in traditional data mining (Van Der Aalst 2016).

- **Temporal Aspect**: Process mining explicitly considers the timing and order of events, aspects often secondary or less emphasized in data mining.

- **Organizational Context**: Process mining inherently incorporates organizational elements, such as roles and resources, into its analyses, while data mining may overlook these factors.

Despite their differences, process mining and data mining can effectively complement each other. For instance, after process mining establishes a clear control-flow structure (like a Petri net), data mining techniques can further enrich these models by discovering decision-making rules, adding predictive insights (Van Der Aalst 2016, p. 46). Recent developments in Object-Centric Process Mining (OCPM), as explored by Berti, Montali & Van Der Aalst (2023) and Berti, Jessen, Park, Rafiei & Van Der Aalst (2023), have further advanced analytical capabilities by providing more comprehensive insights into complex, interconnected business processes and associated data relationships.

2.2 Object-Centric Process Mining

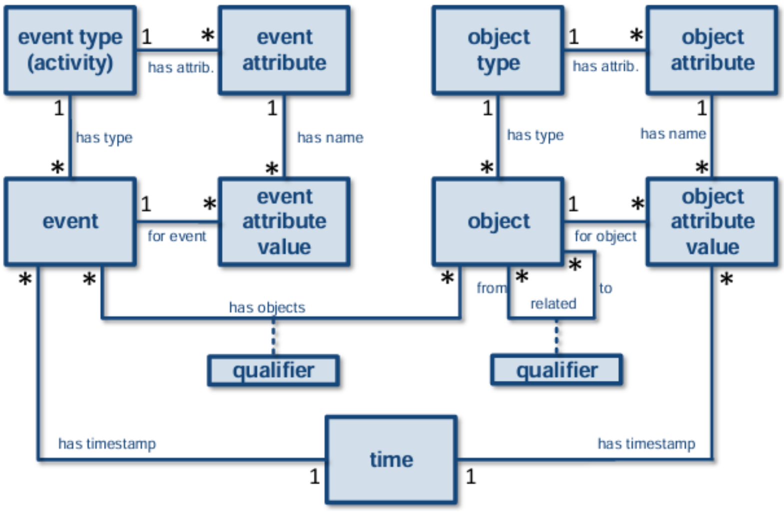

FIGURE 1: The object-centric event data meta-model (OCED-MM), adapted from (Van Der Aalst 2023*c*)

Object-Centric Process Mining (OCPM) has emerged as an innovative approach in the field of process analysis, addressing key limitations inherent in traditional case-centric methods. Unlike conventional methods that typically focus on individual cases (such as single orders), OCPM utilizes object-Centric Event Logs (OCEL) to model interactions involving multiple object types (like orders, products, or shipments) within a single event (Van Der Aalst 2023*b*). This shift enables more comprehensive, multidimensional analyses of business processes, capturing complex relationships among objects that traditional models often oversimplify (Van Der Aalst 2023*c*).

Academic research highlights OCPM's capability to effectively manage challenges like convergence (overlapping sequences) and divergence (fragmented process paths), common issues in traditional process mining (Berti, Montali & Van Der Aalst 2023). By accurately preserving object interdependencies, OCPM supports improved performance measurement and root-cause analysis, notably demonstrated in procurement and supply chain contexts (Berti, Jessen, Park, Rafiei & Van Der Aalst 2023). Recent advancements include predictive monitoring frameworks designed to forecast Key Performance Indicators (KPIs) and identify process deviations in real time (Ruffini 2023). For example, Ruffini (2023) developed techniques for diagnostic pattern extraction and constraint monitoring based on OCEL data, while Van Der Aalst (2023*b*) highlighted the scalability of OCPM in managing artifact-centric systems.

However, fully adapting existing algorithms to verify compliance effectively within OCEL frameworks remains challenging, as discussed by Berti, Montali & Van Der Aalst (2023). According to a systematic review conducted in 2023, notable gaps remain in standardized evaluation metrics, accompanied by practical barriers to broader implementation. This review emphasizes the importance of using granularity adjustment tools, such as drill-down and roll-up operations, to balance detailed insights with general overviews effectively Khayatbashi et al. (2024). These findings underline OCPM's significant potential in optimizing complex processes, particularly in advanced areas like blockchain technology and smart manufacturing Van Der Aalst (2023*c*).

2.2.1 Object-Centric Process Mining vs. Traditional Process Mining

Object-Centric Process Mining (OCPM) extends traditional process mining by addressing its limitations in handling complex, multi-entity systems. Traditional Process Mining (TPM) operates on case-centric event logs, where each event is linked to a single case ID. This assumption works well for linear, clearly structured processes but falls short in environments where activities involve interactions between multiple entities (Van Der Aalst 2023*c*,*b*, Berti, Montali & Van Der Aalst 2023).

In contrast, OCPM employs Object-Centric Event Logs (OCEL), enabling the representation of events involving several related objects. This facilitates a richer, more realistic analysis of organizational workflows. The main distinctions between OCPM and TPM can be grouped into five key dimensions:

- **Data Representation**: TPM uses flat logs centered on a single object type, which simplifies modeling but omits cross-entity relationships. OCPM logs allow one event to connect to multiple objects, offering a multidimensional process view (Park et al. 2024, Berti, Montali & Van Der Aalst 2023, Berti, Jessen, Park, Rafiei & Van Der Aalst 2023).
- **Process Discovery**: TPM tools tend to struggle with concurrency and loops due to their linear assumptions. OCPM overcomes this by modeling interactions among object types, improving the accuracy of discovered models (Berti et al. 2024, Benzin et al. 2024).

- **Conformance Checking**: While TPM checks compliance within single-case boundaries, OCPM enables cross-object conformance evaluation, identifying more complex deviations (Ruffini 2023, Khayatbashi et al. 2024).
- **Process Enhancement**: OCPM enhances diagnosis and optimization by tracing issues across object interactions. This leads to more targeted improvement strategies compared to TPM (Berti et al. 2024, Berti, Jessen, Park, Rafiei & Van Der Aalst 2023, Khayatbashi et al. 2024).
- **Scalability and Complexity**: TPM benefits from mature tools and fast processing for flat logs. OCPM is better suited for analyzing high-entity environments but demands more preprocessing and interpretive effort (Van Der Aalst 2023*c*, Berti, Montali & Van Der Aalst 2023).

Recent studies highlight OCPM's practical advantages. For example, Park et al. (2024) show that transforming healthcare data into OCELs allowed for a more nuanced understanding of treatment paths and resource dependencies. Similarly, Berti, Jessen, Park, Rafiei & Van Der Aalst (2023) demonstrate that OCPM revealed coordination inefficiencies in procurement processes that TPM could not capture.

Despite its promise, OCPM still faces challenges, most notably a lack of standardized evaluation metrics and limited tool support compared to TPM (Berti, Montali & Van Der Aalst 2023). This restricts its applicability in everyday process mining tasks. One of this thesis's core goals is to propose a practical framework for deciding when OCPM's added complexity is justified, based on process needs and data characteristics.

2.3 Event Logs

Event logs are structured digital records that document the sequence and details of activities, tasks, and workflows executed within an organization's information systems (Van Der Aalst 2016). Essentially, event logs serve as digital footprints of business operations, offering detailed records of how processes occur in reality (Van Der Aalst 2023*b*). Each event typically includes a unique case identifier, an activity description, and a timestamp (Marin-Castro & Tello-Leal 2021).

In process mining, event logs are crucial as they provide the foundational data necessary for analyzing real process behaviors. Rather than relying solely on idealized models, organizations use these logs to gain practical insights into actual process executions, revealing detailed flows, variations, and performance characteristics (De Leoni et al. 2016*a*, Van Der Aalst 2016).

2.3.1 The Role of Information Systems

Information systems are central to the generation and management of event logs. Modern enterprise solutions like Enterprise Resource Planning (ERP) systems, Customer Relationship Management (CRM) platforms, and workflow management software naturally record event data as part of their operations (Li et al. 2018). These systems capture extensive data, including user interactions, system responses, and process changes.

The effectiveness of event log analysis greatly depends on the quality of data collected by these information systems. While robust and well-integrated systems provide high-quality, comprehensive logs, legacy systems or poorly integrated IT environments can produce incomplete or inconsistent data (Suriadi et al. 2017).

To address these challenges, organizations frequently employ preprocessing techniques. These methods involve cleaning, standardizing, and enriching raw event logs to ensure the data is suitable for meaningful process mining analyses (Marin-Castro & Tello-Leal 2021). Such preprocessing may include data filtering, event correlation, and timestamp adjustments, significantly improving the reliability of subsequent analyses.

2.3.2 Object-Centric Event Logs

Object-Centric Event Logs (OCELs) represent a significant innovation in process mining, addressing limitations of traditional logs that focus solely on single cases (Ghahfarokhi et al. 2021). Unlike conventional event logs that record isolated cases, OCELs capture interactions among multiple interconnected objects, such as orders, invoices, and customers, in complex business processes (Van Der Aalst 2023*c*).

This advanced format enables a richer, multidimensional representation of processes, making it easier to understand intricate relationships and dependencies often overlooked by simpler models (Berti, Montali & Van Der Aalst 2023, Berti, Jessen, Park, Rafiei & Van Der Aalst 2023).

Key features of OCELs include:

1. **Multi-object perspective**: Logs include events associated with various object types simultaneously (Ghahfarokhi et al. 2021).

2. **Flexible relationships**: They allow many-to-many relationships between objects and events, reflecting real-world complexity (Van Der Aalst 2023*c*).

3. **Hierarchical structures**: OCELs effectively represent nested and multi-layered processes (Berti, Montali & Van Der Aalst 2023).

Adopting OCELs has expanded process mining capabilities, proving particularly valuable in sectors like healthcare, manufacturing, and supply chain management, where processes are highly interconnected (Park et al. 2024).

2.3.3 The Potential for Process Mining

Event logs, especially OCELs, provide significant opportunities for process mining applications. Leveraging these detailed data sources enables organizations to deeply understand their operational processes, enhancing efficiency, compliance, and strategic decision-making (Van Der Aalst 2016).

Event logs drive powerful process mining outcomes, including:

1. **Process Discovery**: Automatically constructing accurate process models from event data to visualize real operational flows and variations (De Leoni et al. 2016*a*).

2. **Conformance Checking**: Evaluating actual process execution against predefined models to detect compliance issues or deviations (Berti et al. 2024).

3. **Performance Analysis**: Identifying bottlenecks, inefficiencies, and improvement opportunities by examining process durations and resource use (Benzin et al. 2024).

4. **Predictive Analytics**: Utilizing historical data to predict future process outcomes, like expected completion times or potential deviations (De Leoni et al. 2016*b*).

The emergence of object-centric approaches has significantly enhanced these capabilities, allowing more nuanced analyses of complex processes involving multiple interacting elements (Berti, Jessen, Park, Rafiei & Van Der Aalst 2023). As a result, organizations gain deeper, holistic insights into their processes, accounting for detailed interactions among objects and stakeholders (Morelli et al. 2024).

As digital transformation continues and organizations generate increasingly sophisticated event data, the potential of process mining for driving operational excellence grows significantly. Leveraging event logs effectively through advanced process mining techniques empowers businesses to uncover critical insights, optimize operations, and maintain a competitive edge in an increasingly data-driven market (Santic 2023).

2.4 Case Studies and Rationale for This Approach

Case studies are a well-established method for investigating complex phenomena in depth and within their real-life setting (Yin 2017). One key advantage of this approach is that it allows researchers to capture the richness and context of processes that cannot be fully understood through abstract models or synthetic data alone (Paschek et al. 2017). By focusing on concrete cases, it becomes possible to examine not only the technical performance of analytical methods but also practical aspects such as data quality, tool usability, and interpretability of results.

In the context of this thesis, using case studies was particularly suitable because the goal is to compare traditional process mining and object-centric process mining under realistic conditions. Processes in organisations rarely follow perfectly linear paths; they often involve exceptions, noise, and overlapping activities. Analysing two contrasting cases, a structured administrative process and a complex, multi-entity system, provides a balanced view of how each method handles typical challenges found in practice (Leemans & Fahland 2020, Van Der Aalst 2023*a*).

Furthermore, the case study approach supports developing recommendations that are directly relevant to practitioners. Rather than relying solely on theoretical assumptions, this thesis grounds its decision framework in lessons drawn from actual data and real process behaviour. In this way, the use of case studies ensures that the findings are not only conceptually sound but also applicable to real-world scenarios where organisations must choose the most suitable process mining technique for their specific context.

2.5 Decision Making in Method Selection

Effective decision making is essential when choosing between technical approaches, especially in complex domains like process mining. Decision frameworks help structure the selection process, linking methodological rigor with organisational goals and context (Wątróbski et al. 2019).

Multi-Criteria Decision Analysis (MCDA) offers a systematic way to evaluate options against multiple, often conflicting, criteria. For example, Wątróbski et al. (2019) propose a hierarchical framework that guides users toward suitable MCDA methods based on the nature of the decision problem. Similarly, Cinelli et al. (2022) present a taxonomy-based decision support tool to recommend appropriate MCDA methods tailored to the decision context. These approaches demonstrate how to select analytical techniques rigorously and transparently, reducing bias and enhancing repeatability.

In this thesis, a structured decision model is proposed to determine when to apply traditional versus object-centric process mining. The model uses a multi-criteria approach, informed by attributes such as process complexity, data structure, and stakeholder requirements. Drawing from established MCDA principles, the model ensures that method choices are justified not by intuition but by clear, documented decision logic that aligns with organisational needs.

This decision approach supports more informed and defensible choices, ensuring that the selected process mining method maximises insight while remaining practical and interpretable in real-world scenarios.

3 METHODOLOGY

This study follows a qualitative case study design supported by exploratory event log analysis and software experimentation. Its central objective is to develop a structured decision-making framework that assists analysts in selecting whether Traditional Process Mining or Object-Centric Process Mining is more appropriate, depending on the characteristics of the process and the available data. By comparing two contrasting datasets and applying each method using relevant tools, the study evaluates the practical strengths and trade-offs involved. This methodology enables both conceptual understanding and applied insight, bridging theoretical foundations with real-world use cases.

3.1 Data Collection

The **BPI Challenge 2017** event log records the loan application process within a Dutch financial institution, containing events for over 30,000 applications with a single-case structure.

The **Age of Empires II** gameplay dataset, formatted according to the OCEL 2.0 standard (Liss et al. 2024), captures interactions from 1,000 game sessions, with more than 2.3 million events and over 360,000 objects across multiple types.

These datasets cover contrasting levels of process complexity and interaction, forming a suitable basis for comparative analysis.

3.2 Dataset Background and Context

This section introduces both datasets so that readers unfamiliar with the underlying domains can understand the context of the analysis.

Loan Application Dataset (BPI Challenge 2017). The first dataset originates from the BPI Challenge 2017 and reflects the loan application process of a Dutch financial institution. It includes 31,509 anonymized applications submitted between January 2016 and February 2017. Each event refers to a specific application and documents steps such as submission, validation, offer creation, and decision.

Applications are submitted either online or in person at a branch. Events include activities like *A_Create Application*, *A_Validating*, *O_Create Offer*, and *A_Pending*. In total, the log consists of 561,671 events and follows a case-centric format with one identifier per case. The process is relatively linear and administrative in nature, with limited concurrency or cross-entity dependencies, which makes it suitable for traditional process mining.

Age of Empires II Gameplay Dataset. The second dataset is based on gameplay data from 1,000 matches of the real-time strategy game *Age of Empires II*. In the game, players control units and structures to gather resources, construct buildings, and compete against each other. The early game centers around building an efficient economy, guided by so-called build orders that define optimized sequences of actions, similar to chess openings.

The dataset follows the OCEL 2.0 standard and contains over 2.3 million events involving more than 360,000 objects across 30 object types. Key object types include *Villager*, *Town Center*, *Resource Drop-Off Buildings*, and *Military Units*. Events describe actions such as *Command Build [Structure]*, *Gather [Resource]*, and *Complete Research [Technology]*. Many actions are triggered by players but others result from built-in automation in the game engine.

Each event can reference multiple objects, such as a villager constructing a building during a player session. The dataset captures a highly concurrent environment where object interactions are essential. Its multi-entity and non-linear structure aligns with the principles of object-centric process mining.

3.3 Tool Selection

Each dataset was analyzed using tools aligned with the respective process mining approach and to ensure fair methodological comparison.

The **BPI Challenge 2017** log was primarily processed using the **Celonis Execution Management System (EMS)**, a commercial platform specifically designed for traditional, case-centric process mining. Celonis provides robust automated process discovery, KPI reporting, and conformance checking, offering a mature and user-friendly environment for structured business workflows.

To ensure methodological consistency across the comparative analysis, the **BPI Challenge 2017** log was additionally analyzed using the **PM4Py** Python library. This allowed results from the traditional approach to be directly compared to the object-centric results generated with the same tool environment.

The **Age of Empires II** dataset was analyzed using **PM4Py** exclusively, as it supports advanced manipulation of object-centric event logs and provides flexible capabilities for parsing, exploring, and visualizing OCEL-based models. Although PM4Py requires more technical expertise than Celonis, it enables detailed control over object interactions and multi-entity representations.

This combined tool selection ensured that each process mining technique was applied using its most suitable platform while maintaining consistency for a robust comparative evaluation.

3.4 Case Selection and Setup

The two datasets were selected to test process mining approaches under contrasting conditions.

The **BPI Challenge 2017** dataset represents a structured administrative workflow with a clear single-case logic, suitable for traditional tools such as Celonis (Van Der Aalst 2016).

The **Age of Empires II** dataset involves high concurrency and multiple interacting objects, capturing the complexity relevant for object-centric analysis (Van Der Aalst 2023*b*,*c*, Berti, Montali & Van Der Aalst 2023).

Key selection criteria included:

- Number and variety of object types
- Level of interaction and concurrency
- Diversity of process variants

This combination supports a grounded comparison of method suitability for processes with different levels of complexity and interdependence.

3.5 Comparative Framework Design and Evaluation Approach

To assess the strengths and limitations of Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM), a comparative framework was developed and applied based on key dimensions identified in the literature (Van Der Aalst 2023*b*, Berti et al. 2024, Van Der Aalst 2023*c*, Berti, Montali & Van Der Aalst 2023) and refined through practical experimentation.

The framework evaluates both techniques across eight core dimensions that capture technical capabilities and user-oriented aspects: data representation, process discovery, model complexity handling, scalability, conformance checking, insight generation, ease of interpretation, and tool support. These dimensions reflect trade-offs commonly discussed in process mining research and were operationalized through a mix of qualitative analysis and quantitative indicators.

To ensure transparency, the BPI Challenge 2017 dataset was analyzed using both Celonis and PM4Py, while the Age of Empires II dataset was analyzed using PM4Py. This ensured consistent tool environments for comparing both methods.

This structured comparison provides the empirical foundation for the decision framework proposed in Chapter 6, supporting more informed method selection based on process complexity and data characteristics.

3.5.1 Dimension Definitions and KPIs

To ensure transparency and consistency, each evaluation dimension in the comparative framework is defined with specific criteria and example indicators used in this study:

- **Data Representation:** Assesses whether the method preserves relationships between multiple objects. Good representation means the event log structure allows analysts to explore interactions between entities without the need to flatten data.

- **Process Discovery:** Evaluates the ability to automatically generate accurate and understandable process models, including handling loops and concurrency. This was verified by comparing discovered control-flow with expected process behavior.

- **Model Complexity Handling:** Judges the method's capacity to handle complex, entangled processes without oversimplification. Clarity in distinguishing overlapping and concurrent paths was assessed qualitatively.

- **Scalability:** Measured by processing time and responsiveness when handling large logs. For example, runtime for importing, discovering, and visualizing models for the BPI Challenge log (561,000 events) and the Age of Empires II OCEL log (over 2 million events) was compared.

- **Conformance Checking:** Evaluated by how well deviations from expected process behavior were detected. An example KPI is the number of compliance violations identified per process variant.

- **Insight Generation:** Judged by the depth and variety of insights provided. Indicators include cycle time statistics, identification of bottlenecks, and root cause analysis. For OCPM, the ability to reveal inter-object coordination is a key advantage.

- **Ease of Interpretation:** Defined as the level of understandability for non-technical stakeholders. This was assessed based on visual complexity (number of nodes and edges) and the extent to which abstraction tools were needed. Simpler BPMN-style models scored higher in this dimension.

- **Tool Support:** Assessed by the maturity, usability, and available features of supporting tools. Strong support means intuitive interfaces, dashboards, and robust documentation.

Key performance indicators (KPIs) used in the comparison include cycle time ranges, variant frequencies, throughput times, number of detected compliance breaches, and model fitness and precision metrics, where applicable.

Figure 2 presents a radar chart summarizing the relative performance of TPM and OCPM along these eight dimensions, based on both literature benchmarks and experimental results.

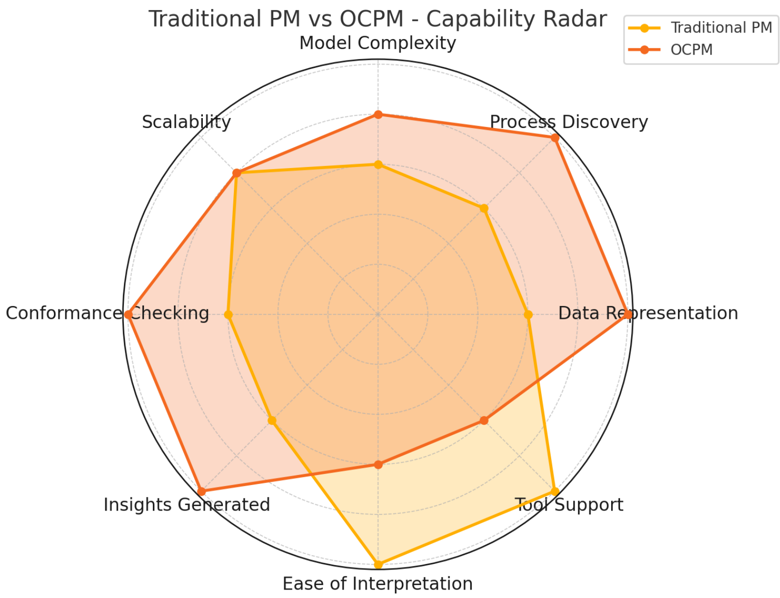

FIGURE 2: Radar chart comparing Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM) across eight evaluation dimensions.

3.6 Data Preprocessing

Each dataset was preprocessed to ensure clean, comparable inputs for analysis.

BPI Challenge 2017 preprocessing was done in Celonis EMS:

- **Data Cleaning:** Removed incomplete records and entries with missing timestamps or case identifiers.

- **Attribute Filtering:** Retained only relevant activities and attributes related to the application lifecycle.
- **Clustering:** Grouped process variants to support interpretation of throughput and bottlenecks.

Age of Empires II preprocessing was carried out in Python with PM4Py:

- **OCEL Parsing and Cleaning:** Filtered out technical events and redundant system-level interactions.
- **Object and Event Selection:** Reduced to a focused set of 8 object types and 12 key event types for clarity.
- **Entity Filtering:** Included only complete match sessions with full player data.
- **GPT-4-Aided Labeling:** Grouped similar actions into higher-level categories (e.g., resource_collection, building_construction) to simplify visualizations and highlight broader behavioral patterns. See Appendix A for an example.

These steps ensured that both logs were cleaned, filtered, and structured to support a consistent and fair comparison without oversimplifying their complexity.

3.7 Evaluation Approach

The evaluation of TPM and OCPM was conducted through a comparative framework developed specifically for this thesis. This framework was informed by the literature (Van Der Aalst 2023*b*,*c*, Berti, Montali & Van Der Aalst 2023) and refined through practical experimentation with the BPI Challenge 2017 and Age of Empires II datasets.

The dimensioned mentioned in 3.5.1 were operationalized through a combination of qualitative analysis (model interpretability, visual clarity, abstraction support) and basic quantitative indicators (number of discovered variants, processing time, completeness of conformance outputs).

For example, the BPI dataset was evaluated on throughput and bottleneck detection across high-frequency variants, while the Age of Empires dataset was assessed for its ability to visualize coordination between units and structures. In both cases, model outputs were compared against expectations from the literature and visual benchmarks derived from tool dashboards.

To support transparency, evaluations were also annotated with qualitative observations regarding usability and abstraction. The results were synthesized in the form of a radar chart and summary table (Chapter 5), enabling a side-by-side view of both approaches across all dimensions.

This structured evaluation serves as the empirical foundation for the decision framework proposed in Chapter 6.

4 CASE STUDY

To explore the practical implications of Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM), this chapter analyzes two contrasting datasets using a common Python-based approach with the PM4Py library. Additionally, the BPI Challenge 2017 dataset was also analyzed using Celonis to reflect industry-standard tool support for traditional process mining. This dual analysis allows for a more comprehensive comparison, particularly in terms of usability, depth, and tool limitations.

As mentioned in chapter 3. Methodology, the first case study examines a structured administrative process from the financial sector, while the second draws on gameplay data from the strategy game Age of Empires II, characterized by concurrent and interdependent actions across multiple object types. This contrast supports a grounded comparison of the strengths and trade-offs associated with TPM and OCPM.

4.1 Data Preparation and Model Generation

TPM (Celonis and Python, BPI 2017) For the TPM analysis, the log was flattened to a case-centric format using the Application ID as the main identifier. In Celonis, this enabled immediate process discovery, KPI generation, and conformance checking without code. In parallel, the same data was analyzed in Python using PM4Py to replicate core functionalities like variant discovery and performance analysis programmatically.

OCPM (Python, AoE II) In contrast, preparing the Age of Empires dataset for OCPM was more involved. The log was structured in OCEL 2.0 format, which retains relationships between different object types (such as *Villagers*, *Town Centers*, and *Resources*). This required writing custom scripts to parse and organize event-object mappings. While the process took more time, it allowed for a much richer representation of the interactions and dynamics within the system.

4.2 Process Visualization

TPM: The traditional loan application process from the BPI 2017 dataset was visualized using two complementary approaches. First, Celonis generated a BPMN-style model that was easy to interpret and highly suited for identifying bottlenecks and throughput variations across cases. This model, shown in Figure 3, illustrates the linear flow from application creation to completion, constrained to a single case perspective.

Second, a Petri net was discovered using the Inductive Miner algorithm implemented in the PM4Py Python library. The process log was flattened to a case-centric format using the Application ID, enabling model discovery consistent with traditional assumptions. While Petri nets are less intuitive for business users compared to BPMN diagrams, they offer a precise formalism ideal for detailed behavioral analysis and conformance checking. Figure 4 displays the resulting net structure, providing an alternative view on the control-flow logic of the same process. By comparing the two figures one clear advantage of Celonis can be observed, namely the readability. Figure 4, is hardly readable do to the nature of the process as it involves many steps and PM4Py is not designed for creating Petri nets that can be easily printed.

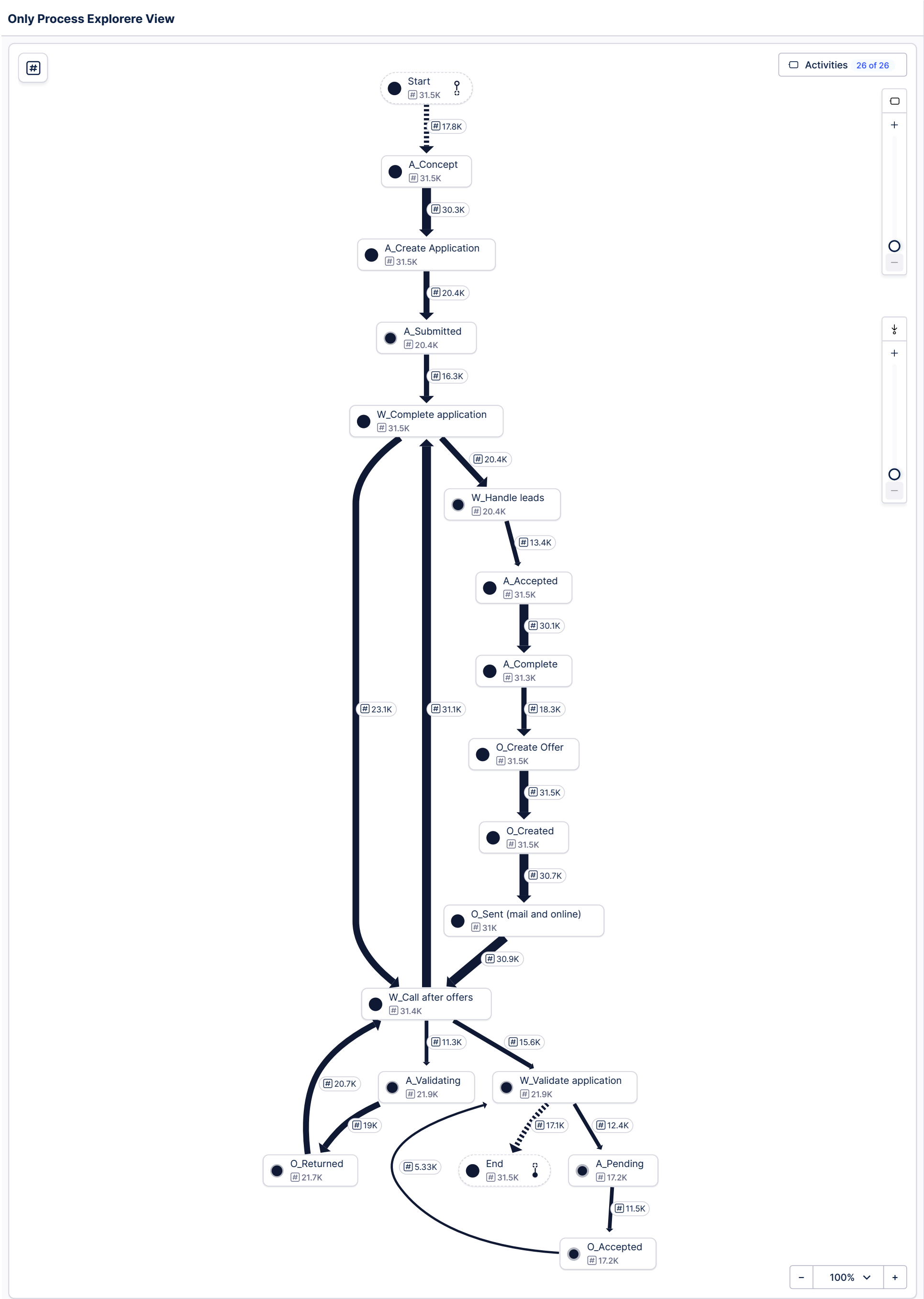

FIGURE 3: Traditional process model of the BPI 2017 loan application process visualized in Celonis. The model reflects a case-centric control-flow, showing the linear structure of the loan process from application creation to completion.

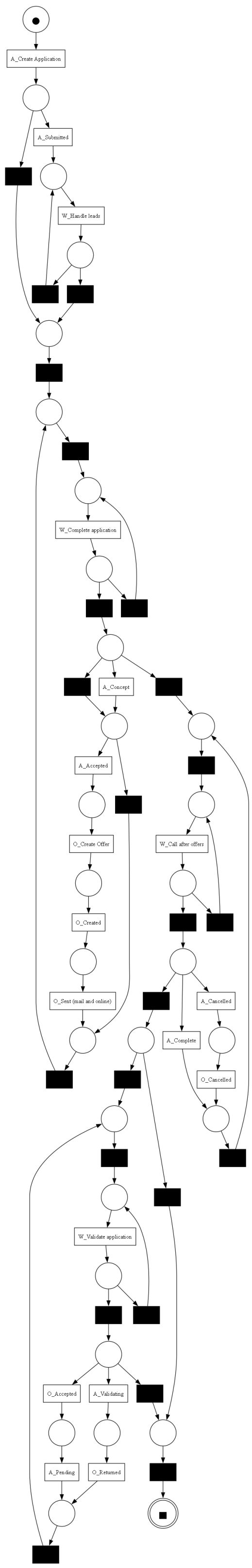

FIGURE 4: Petri net of the BPI 2017 loan application process generated using PM4Py in Python. The model reflects the discovered control-flow after flattening the event log to a single-case format.

OCPM: The Age of Empires II event log, represented in OCEL format, was visualized using object-centric graphs that capture concurrent interactions among different object types such as Villagers, Town Centers, and Resources. These visualizations revealed the complex coordination patterns typical in multi-agent environments. However, the resulting models were dense and required abstraction techniques, such as filtering event types and grouping objects, to ensure interpretability. An example of such a filtered model is presented in Figure 5.

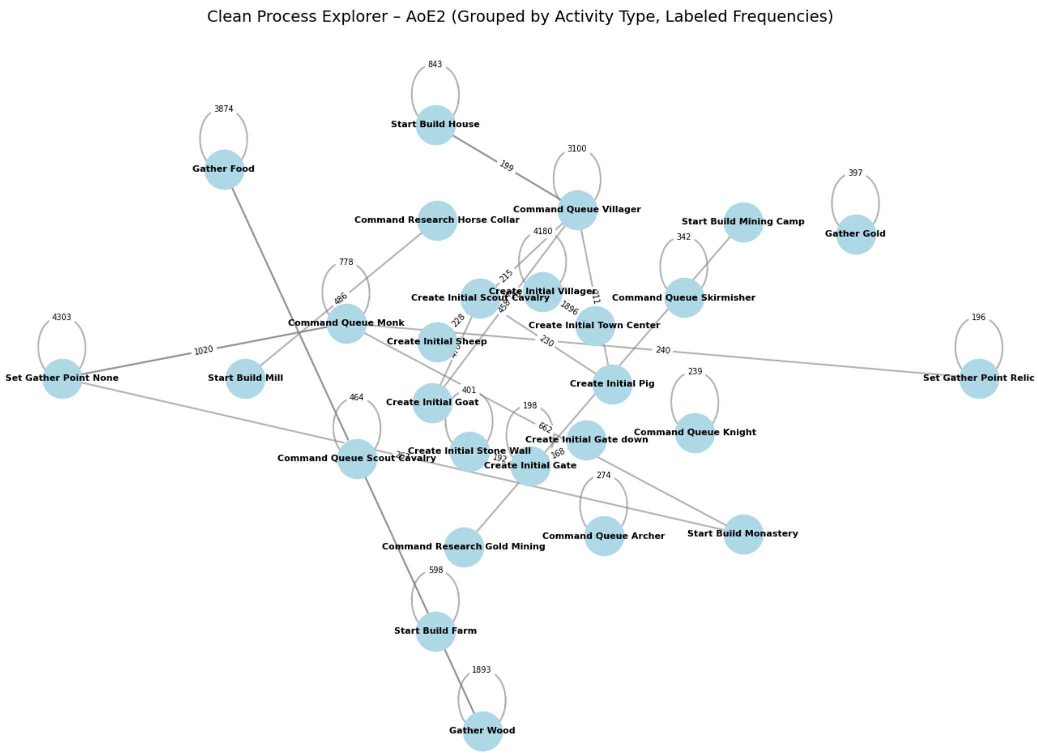

FIGURE 5: Object-Centric Process Mining visualization showing interactions between Villagers and Town Centers in the Age of Empires II dataset. The model illustrates multi-object concurrency and coordination, showcasing the richer process representation enabled by OCPM.

4.3 Perspective Flexibility

TPM: Analysis in Celonis was locked to the predefined case notion. Exploring other perspectives (e.g., per customer or per offer) would require re-importing or restructuring the log. PM4Py allowed for more flexibility, but still within the constraints of single-case logic.

OCPM: The object-centric structure made it much easier to explore the process from different angles. For example, I could generate one model focusing on *Villagers* and another on *Buildings*, without changing the underlying data. This flexibility turned out to be one of OCPM's biggest strengths, especially when analyzing complex systems with many interacting parts.

4.4 Analytical Depth

TPM: Both Celonis and PM4Py were effective in identifying throughput times, variants, and bottlenecks in the loan application process. However, they lacked the ability to detect inter-object coordination patterns.

OCPM: In contrast, the OCPM setup revealed a lot about the structure and flow of coordination between different object types. For example, it was possible to see how *Villagers* contributed to resource collection and how this behavior connected to the use of drop-off buildings. These relationships would be difficult to uncover in a traditional process mining environment.

4.5 Usability and Tool Support

TPM: Celonis is clearly designed for ease of use. The interface is intuitive, and the built-in dashboards are well-suited for business users. From loading the data to generating initial models, the entire workflow can be completed without writing a single line of code. PM4Py, while replicating similar analysis, demanded Python proficiency and manual configuration.

OCPM: Similiar to PM4Py for TPM, the OCPM workflow required a solid understanding of Python, OCEL structures, and visualization libraries. There is more manual setup involved, but also more control over what gets analyzed and how. Figure 7 shows an excerpt from the code used to create one of the object-centric models.

4.6 Summary

Table I summarizes the differences observed across the two approaches.

Dimension	**TPM (Celonis and Python, BPI 2017)**	**OCPM (Python, AoE II)**
Data Preparation	Flattened to case-centric; direct Celonis load; Python parsing with PM4Py	OCEL format; complex multi-entity filtering and grouping required
Visualization	BPMN-style in Celonis; graph-based in PM4Py; single-case focus	Dense multi-object graph; requires abstraction for clarity
Perspective Flexibility	Fixed in Celonis; limited in Python TPM	High; perspective switching between objects in Python
Analytical Depth	Strong for KPIs and variants; weak on coordination	High inter-object insight; coordination and dependency patterns visible
Usability	Very accessible in Celonis; PM4Py requires scripting	Requires technical skills; flexible but less user-friendly

TABLE I: Comparison of TPM and OCPM across analytical dimensions

5 DATA ANALYSIS AND RESULTS

The following section presents the results of the comparative analysis conducted using the evaluation framework introduced earlier. Based on two contrasting datasets, one representing a structured financial process and the other a dynamic multi-entity game environment, the analysis assesses how Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM) perform across eight analytical dimensions. Each method is applied to its corresponding dataset to highlight key strengths, limitations, and tradeoffs, providing empirical insight into the practical implications of choosing one approach over the other.

5.1 Comparative Evaluation

Both logs were analyzed using the framework introduced earlier, focusing on eight key dimensions as you can see in Table II.

TABLE II: Comparison of Traditional PM and OCPM

Dimension	Traditional PM (BPI 2017)	OCPM (AoE OCEL)
Data Representation	Flattened case-centric (one object per event)	OCEL format with multiple object types per event
Discovery Capabilities	Clear linear flows; some loops difficult to capture	Effective with parallel flows and multi-object transitions
Model Complexity	Limited to sequential task paths	Captures inter-object concurrency and entanglement
Scalability	High – Celonis processes flat logs efficiently	Moderate – pre-processing required but scalable for complex logs
Conformance Checking	Case-based deviations observable	Enables cross-object compliance and anomaly detection
Insight Generation	Strong in cycle time, bottleneck, and variant analysis	Enables multi-object insights and network analysis
Ease of Interpretation	High – simple models and intuitive UI	Lower – denser models require abstraction tools
Tool Support	Mature (Celonis, Disco)	Emerging (PM4Py, ProM, early Celonis OCDM support)

5.2 Observed Strength and Limitations

Using Celonis for the traditional case allowed for fast implementation and immediate insights. Analysts could detect where cases accumulated delays and which departments were involved in prolonged decision paths. The number of variants and their frequencies were easily retrievable, and performance metrics like cycle time could be compared across different application types.

The OCPM analysis, on the other hand, required more setup. Preprocessing the Age of Empires dataset into OCEL format involved careful mapping of relationships between units, players, and actions. However, the result was a much more nuanced process model that was capable of revealing indirect dependencies and resource contention. For example, the same event (e.g., a villager building a farm) could be associated with the villager object, the structure object, and the player session, providing multiple analytical angles.

While the visual complexity of OCPM models was significantly higher, tools like INEXA-style abstraction or object-specific filtering allowed for better manageability. In many ways, the depth of insight OCPM offered came at the cost of interpretability, particularly for stakeholders unfamiliar with the method.

6 DECISION FRAMEWORK

The increasing complexity of business processes has made it essential to identify the most appropriate process mining technique for a given context. While both Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM) offer valuable insights, their effectiveness depends on the nature of the process being analyzed. This section introduces a decision framework developed to guide practitioners and researchers in selecting the most suitable method based on key characteristics of their data and process landscape.

6.1 Framework Rationale

The rationale for developing a decision framework stems from the need to support practitioners in selecting between TPM and OCPM based on process characteristics. Rather than reiterating detailed differences between the methods, this section synthesizes insights from literature and empirical findings to identify key evaluation dimensions: object multiplicity, process complexity, and tool usability. These dimensions align with common analytical goals, such as scalability, interpretability, and insight generation, and help bridge the gap between theoretical capabilities and real-world applicability.

6.2 Decision Criteria

The core dimensions considered in the framework include:

- **Data Representation:** Whether the event log contains one or multiple object types.
- **Process Discovery Needs:** The degree of complexity in the flow, including concurrency and interdependent subprocesses.
- **Model Interpretability:** The importance of clear visualization for communication and decision-making.
- **Conformance Checking Requirements:** Whether cross-entity compliance checks are needed.
- **Scalability and Tool Support:** Availability of mature tools and the performance of those tools with the given data structure.
- **Data Quality:** The presence of well-defined object relationships and complete timestamping in the log.

6.3 *Tree-Based Decision Model*

To operationalize these criteria, a two-layer decision model was constructed. The first layer (Figure 6) guides users through a sequence of yes/no questions to suggest whether TPM or OCPM is more appropriate. For example, when dealing with multi-entity processes that exhibit high inter-object interaction and control-flow complexity, the framework recommends OCPM. Conversely, for flat, single-case processes with minimal interconnectivity, TPM is likely sufficient.

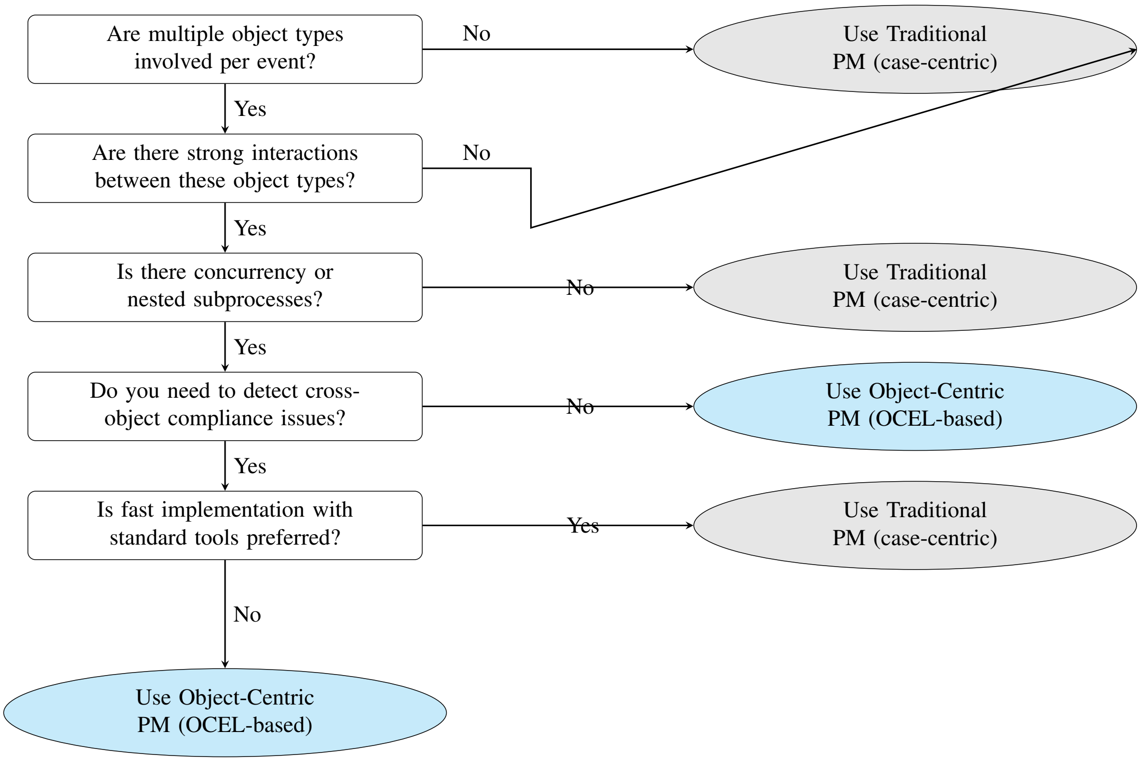

FIGURE 6: Decision tree framework for selecting between Traditional and Object-Centric Process Mining.

6.4 *Technique and Tool Recommendation Table*

Once a general method is selected, the second layer of the framework (Table III) provides practical guidance on choosing specific process discovery techniques or tools based on the identified scenario.

TABLE III: Recommended Techniques and Tools Based on Method Selection

Scenario	Recommended Technique or Tool	Rationale
TPM with need for high fitness	Inductive Miner (IM) in Celonis	Captures full process behavior with high replay accuracy; Celonis provides an intuitive interface for applying this
TPM with need for high precision	Evolutionary Tree Miner (ETM) in Celonis	Produces more restrictive models to avoid overgeneralization; supported within Celonis model variants
TPM with focus on speed and balanced results	Split Miner (SM) or Celonis Execution Management System (EMS)	High F-score with efficient runtime; Celonis EMS enables fast, practical deployment for business users
OCPM for complex coordination	PM4Py (graph-based algorithms)	Supports detailed modeling of inter-object behavior using OCEL logs in Python
OCPM with need for clear stakeholder visualisation	INEXA framework	Provides interactive abstraction to simplify dense object-centric models
OCPM for real-time or dynamic scenarios	Custom OCEL pipelines	Enables pattern detection and monitoring in continuously changing environments

6.5 Summary Table

Table IV presents a simplified overview of the conditions favoring each technique. This complements the decision tree by allowing quick reference during tool selection or process planning.

TABLE IV: Summary of Process Characteristics and Recommended Approach

Process Characteristic	Recommended Technique	Rationale
Single object per event	TPM	Simple structure; widely supported by mature tools
Multiple objects per event	OCPM	Captures inter-object relationships and preserves context
Low interdependency between entities	TPM	Flat models sufficient when object relationships are limited
High inter-object dependency	OCPM	Accurately represents realistic and interacting process behaviour
Linear or sequential flows	TPM	Easier to model, interpret, and communicate
Concurrent or nested subprocesses	OCPM	Better suited for capturing concurrency and complexity
Need for fast results or easy-to-use tools	TPM	Supported by intuitive platforms such as Celonis or Disco
Focus on relational insights (e.g., order–supplier–product networks)	OCPM	Enables network-based, multi-entity analysis

6.6 Use in Practice

This framework is designed for flexible application. Analysts can use it at the early stage of a process mining project to assess the appropriateness of available methods. Depending on the data structure and the business objectives, the framework facilitates a more grounded selection of analysis tools and models. By formalizing this choice, organizations can reduce implementation inefficiencies and align analytical effort with the real structure of their operations.

7 DISCUSSION

The following discussion reflects on the main findings of this study, interprets their practical implications, and outlines the limitations that should be considered when applying the results. It also highlights directions for future research to refine the decision framework and extend its use to a broader range of process contexts.

[0]For a detailed benchmark of traditional process mining algorithms, see Augusto et al. (2018).

7.1 Key Findings and Interpretation

The comparative analysis between TPM and OCPM confirms that both techniques offer distinct strengths depending on the nature of the process being examined. TPM remains a solid and accessible choice for linear, well-structured workflows that follow a single-case logic. It benefits from mature tools, fast implementation, and ease of interpretation, making it particularly suitable in business environments where clarity and speed are prioritized.

On the other hand, OCPM demonstrates a clear advantage in processes involving multiple entities and complex interdependencies. By preserving object relationships, it provides richer insights into coordination, concurrency, and multi-layered interactions that would otherwise be flattened or lost in a case-centric model. This strength, however, comes with a trade-off: the models are denser, the tools are less developed, and the learning curve is steeper.

The findings reinforce a central theme: complexity demands flexibility. While TPM is easier to use and explain, it sometimes oversimplifies the reality of operations. OCPM, although more demanding, is better equipped to reflect the intricacies of real-world systems. This insight shaped the development of the decision framework presented in Chapter 6.

7.2 Implications for Practice

From a practical perspective, the framework introduced in this thesis can serve as a valuable support tool for analysts, especially at the early stages of a process mining project. By asking a series of targeted questions about the structure of the process and the nature of the event data, practitioners can make a more informed decision about which approach to use. This is particularly useful when there is limited time or resources, or when working across departments that have different data capabilities.

For example, in a hospital environment, multiple entities such as patients, staff, equipment, and treatment stages interact simultaneously. Here, using OCPM is more suitable because it can preserve and analyze the complex relationships among these different objects. This allows analysts to uncover coordination issues, patient flow inefficiencies, or resource conflicts that a single-case log would obscure.

In contrast, for a straightforward and well-defined process like an order to cash workflow in a retail company, TPM remains the practical choice. The process typically follows a linear sequence from order creation to delivery and payment, with limited interaction between different object types beyond the order itself. TPM tools like Celonis can rapidly provide clear models, cycle times, and bottleneck insights without the overhead of modeling multiple entities.

Similarly, a public transportation company analyzing ticket sales and validation may benefit from TPM if the goal is to check adherence to a standard flow: from ticket purchase to validation and exit. However, if the same company wants to study how passengers interact with vehicles, stations, and different route options simultaneously, an OCPM approach would yield a more realistic view of how these entities influence each other.

Furthermore, organizations that are already invested in TPM tools may benefit from evaluating whether some of their processes could yield deeper insights through OCPM, even if that requires a temporary shift in tooling or mindset. For tool developers, these findings highlight the growing need for better support of multi-entity logs and more intuitive visual abstractions for complex models.

7.3 Limitations

Several limitations should be considered when interpreting the results and applying the proposed framework.

First, the comparative analysis relied on two specific datasets: the BPI Challenge 2017 event log and a game-based object-centric event log derived from Age of Empires II (see Section 3.4). These cases were deliberately chosen to represent contrasting levels of process complexity and object interaction, allowing for a clear evaluation of Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM). However, this focused selection limits the generalizability of the findings to other domains such as logistics, manufacturing, or healthcare, where process structures and concurrency patterns may differ.

Second, the evaluation framework (Section 3.6) combines literature-based benchmarks with practical tool usage, which introduces an element of subjectivity. Criteria such as ease of interpretation, visual clarity, and depth of insights partly depend on the analyst's experience and familiarity with the tools used. These subjective aspects may influence how the techniques are perceived and scored.

Third, technical constraints shaped the tool setup. While Celonis provides mature support for case-centric analysis, its academic version did not support generic object-centric event logs, which made it impossible to test OCPM within the same environment. Therefore, PM4Py was used exclusively for the object-centric analysis, which may introduce slight inconsistencies when comparing usability and tool features across approaches. Additionally, tool capabilities continue to evolve, and future updates may affect the practicality of implementing OCPM in mainstream platforms.

Finally, although the proposed decision framework offers structured guidance, it has not yet undergone extensive industry validation. Its effectiveness depends on practitioners' ability to accurately assess their own process structures and data quality. This may be more challenging in settings with limited technical expertise or incomplete event logs.

Overall, acknowledging these limitations helps define the scope of the conclusions and highlights areas for future research and practical refinement when applying or extending the framework to other contexts.

7.4 Future Research Directions

Several areas could benefit from further exploration. One important direction is to test the decision framework developed in this thesis on real-life industry cases. Applying the framework in operational environments would help validate and refine the decision points, especially in hybrid scenarios where both approaches might be partially suitable. Such practical testing could reveal additional conditions or exceptions not captured through the initial case studies.

Another promising path is to explore the use of AI to support automated model selection or preprocessing recommendations, particularly in situations where analysts face uncertainty about which method to apply. As object-centric logging becomes more common, better standardization and more intuitive visual simplification techniques will be important to support widespread adoption beyond academic or technical teams.

Lastly, future work could involve collaboration with organizations to measure the actual business impact of selecting one mining approach over another, not just in terms of model accuracy but also regarding decision-making effectiveness, process improvement outcomes, and time-to-insight.

8 CONCLUSION

This thesis set out to compare Traditional Process Mining (TPM) and Object-Centric Process Mining (OCPM) and to develop a practical framework that helps analysts choose between the two based on process and data characteristics. Through a combination of literature review, hands-on experimentation, and comparative evaluation, the study aimed to bring more clarity to an increasingly relevant decision in the field of process analytics.

The main findings of this thesis can be summarized as follows:

- **TPM remains an effective and accessible choice** for linear, single-case processes with clear control-flow and minimal inter-object interactions.
- **OCPM provides richer insights** for processes involving multiple related entities and high concurrency, capturing complex dependencies that TPM typically flattens or overlooks.
- **Each method has trade-offs:** TPM benefits from mature tool support, intuitive visualizations, and fast implementation, whereas OCPM requires more technical setup but delivers deeper, multi-entity perspectives.
- **A decision framework was developed** to guide practitioners in selecting the appropriate technique based on key factors such as data structure, process complexity, interpretability needs, and available tool support.
- **The framework bridges theory and practice** by translating comparative insights into a practical, step-by-step tool for analysts to apply in real projects.

Ultimately, the conclusion is clear: process mining is not a one-size-fits-all discipline. The best method depends not just on the tools available, but on the structure of the process being analyzed, the nature of the event data, and the goals of the analysis. By making these trade-offs explicit and actionable, this thesis contributes a small but practical step toward more informed and effective process mining practice.

References

Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F. M., Marrella, A., Mecella, M. & Soo, A. (2018), 'Automated discovery of process models from event logs: Review and benchmark', *IEEE transactions on knowledge and data engineering* **31**(4), 686–705.

Benzin, J.-V., Park, G., Mangler, J. & Rinderle-Ma, S. (2024), 'Inexa: Interactive and explainable process model abstraction through object-centric process mining', *arXiv preprint arXiv:2403.18659* .

Berti, A., Herforth, J., Qafari, M. S. & Van Der Aalst, W. M. P. (2024), 'Graph-based feature extraction on object-centric event logs', *International Journal of Data Science and Analytics* **18**(2), 139–155.

Berti, A., Jessen, U., Park, G., Rafiei, M. & Van Der Aalst, W. M. P. (2023), 'Analyzing interconnected processes: using object-centric process mining to analyze procurement processes', *International Journal of Data Science and Analytics* pp. 1–23.

Berti, A., Montali, M. & Van Der Aalst, W. M. P. (2023), 'Advancements and challenges in object-centric process mining: A systematic literature review', *arXiv preprint arXiv:2311.08795* .

Cinelli, M., Kadziński, M., Miebs, G., Gonzalez, M. & Słowiński, R. (2022), 'Recommending multiple criteria decision analysis methods with a new taxonomy-based decision support system', *European Journal of Operational Research* **302**(2), 633–651.

De Leoni, M., Van Der Aalst, W. M. P. & Dees, M. (2016*a*), 'A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs', *Information Systems* **56**, 235–257.

De Leoni, M., Van Der Aalst, W. M. P. & Dees, M. (2016*b*), 'A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs', *Information Systems* **56**, 235–257.

Ghahfarokhi, A. F., Park, G., Berti, A. & Van Der Aalst, W. M. (2021), Ocel: a standard for object-centric event logs, *in* 'European Conference on Advances in Databases and Information Systems', Springer, pp. 169–175.

Hand, D. J. (2007), 'Principles of data mining', *Drug safety* **30**, 621–622.

Jarrahi, M. H., Askay, D., Eshraghi, A. & Smith, P. (2023), 'Artificial intelligence and knowledge management: A partnership between human and ai', *Business Horizons* **66**(1), 87–99.

Khayatbashi, S., Miri, N. & Jalali, A. (2024), 'Advancing object-centric process mining with multi-dimensional data operations', *arXiv preprint arXiv:2412.00393* .

Leemans, S. J. & Fahland, D. (2020), 'Information-preserving abstractions of event data in process mining', *Knowledge and Information Systems* **62**(3), 1143–1197.

Li, G., de Murillas, E. G. L., de Carvalho, R. M. & Van Der Aalst, W. M. (2018), Extracting object-centric event logs to support process mining on databases, *in* 'Information Systems in the Big Data Era: CAiSE Forum 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings 30', Springer, pp. 182–199.

Liss, L., Elbert, N., Flath, C. M. & Van Der Aalst, W. M. (2024), Framework for extracting real-world object-centric event logs from game data, *in* 'International Conference on Process Mining', Springer, pp. 363–375.

Marin-Castro, H. M. & Tello-Leal, E. (2021), 'Event log preprocessing for process mining: a review', *Applied Sciences* **11**(22), 10556.

Morelli, F., Pintarič, N., Manfreda, A. & Radziszewski, P. (2024), 'The impact of object-centric process mining on business efficiency: A case study approach', *Anwendungen und Konzepte der Wirtschaftsinformatik* (20).

Park, G., Lee, Y. & Cho, M. (2024), 'Enhancing healthcare process analysis through object-centric process mining: Transforming omop common data models into object-centric event logs', *Journal of biomedical informatics* **156**, 104682.

Paschek, D., Luminosu, C. T. & Draghici, A. (2017), Automated business process management–in times of digital transformation using machine learning or artificial intelligence, *in* 'MATEC web of conferences', Vol. 121, EDP Sciences, p. 04007.

Ruffini, G. (2023), Object-centric process mining for process analysis and operational support., *in* 'ICPM Doctoral Consortium/Demo'.

Santic, J. (2023), '5 reasons why object-centric process mining is the fastest and most scalable way to realize process transformation'.
URL: *https://www.celonis.com/blog/5-reasons-why-object-centric-process-mining-is-the-fastest-and-most-scalable-way-to-realize-process-transformation/*

Suriadi, S., Andrews, R., ter Hofstede, A. H. & Wynn, M. T. (2017), 'Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs', *Information systems* **64**, 132–150.

Van Der Aalst, W. M. (2023*a*), 'Object-centric process mining: unraveling the fabric of real processes', *Mathematics* **11**(12), 2691.

Van Der Aalst, W. M. P. (2016), *Data science in action*, Springer.

Van Der Aalst, W. M. P. (2023*b*), *Object-Centric Process Mining: An Introduction*, Springer International Publishing, Cham, pp. 73–105.
URL: *https://doi.org/10.1007/978-3-031-43678-9*$_3$

Van Der Aalst, W. M. P. (2023*c*), 'Object-centric process mining: unraveling the fabric of real processes', *Mathematics* **11**(12), 2691.

Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A. & Zioło, M. (2019), 'Generalised framework for multi-criteria method selection', *Omega* **86**, 107–124.

Yin, R. K. (2017), *Case study research and applications: Design and methods*, Sage publications.

Appendices

A GPT-4 Labeling Prompt and Output

Prompt Used

Below is the prompt used to assist GPT-4 in semantically grouping event labels in the object-centric event log from Age of Empires II:

> You are given a list of event labels from a strategy game telemetry log. Each event represents a type of player action, such as gathering resources, constructing buildings, or training units. Please group these event labels into semantically meaningful categories based on the functional purpose of the action. For each group, provide a name and list the corresponding labels.
>
> Here is the list:
> `gather_wood, gather_food, gather_gold, build_town_center, build_barracks, build_archery_range, train_villager, train_archer, research_wheelbarrow, research_lumber_camp_upgrade, idle_villager, attack_unit, scout_area, move_unit, repair_building`

GPT-4 Output

- **Resource Collection**
  - `gather_wood`
  - `gather_food`
  - `gather_gold`
- **Building Construction**
  - `build_town_center`
  - `build_barracks`
  - `build_archery_range`
- **Unit Training**
  - `train_villager`
  - `train_archer`

- **Technology Research**
  - `research_wheelbarrow`
  - `research_lumber_camp_upgrade`
- **Idle and Maintenance**
  - `idle_villager`
  - `repair_building`
- **Combat and Scouting**
  - `attack_unit`
  - `scout_area`
  - `move_unit`

These grouped categories were then used to relabel or annotate event types in the OCEL log, improving abstraction and interpretability in downstream visualizations.

B PYTHON CODE SNIPPED

```
events = ocel['ocel:events']
match_player_events = defaultdict(list)

for event in events:
    timestamp = datetime.fromisoformat(event['ocel:timestamp'])
    activity = event['ocel:activity']
    omap = event['ocel:omap']
    match_id = next((obj for obj in omap if obj.startswith("M")), None)
    player_id = next((obj for obj in omap if obj.startswith("P")), None)
    if match_id and player_id:
        match_player_events[(match_id, player_id)].append((timestamp, activity))


transition_counts = Counter()
activity_counts = Counter()

for (match_id, player_id), event_list in match_player_events.items():
    sorted_events = sorted(event_list, key=lambda x: x[0])
    activity_seq = [act for _, act in sorted_events[:opening_event_limit]]

    for i in range(len(activity_seq) - 1):
        pair = (activity_seq[i], activity_seq[i + 1])
        transition_counts[pair] += 1

    for act in activity_seq:
        activity_counts[act] += 1


top_nodes = set([a for a, _ in activity_counts.most_common(max_nodes)])


G = nx.DiGraph()
for (src, tgt), count in transition_counts.items():
    if count >= min_edge_freq and src in top_nodes and tgt in top_nodes:
        G.add_edge(src, tgt, weight=count)

```

FIGURE 7: Python code snippet used to generate the OCPM visualization shown in Figure 5. The script parses event-object relationships and constructs an object-centric graph using the PM4Py library.

C DECLARATION OF USE OF AI

Declaration on the use of GenAI in this document

I have used the Generative AI applications – ChatGPT and Grammarly - as an aid for grammar correction with human oversight.

I have not used Generative AI applications to generate any content directly presented in the document. The content presented in the document is original and my own.

The used prompt was: "Can you help improve grammar, clarity, and flow without changing the original meaning of the following paragraph?"