EVErPREP: Towards an Event Knowledge Graph enhanced Workflow Model for Event Log Preparation

Peter Filipp[1] [0009−0008−9163−0388], Rene Dorsch[1] [0000−0001−6857−7314], and Andreas Harth[1,2] [0000−0002−9421−8566]

[1] Fraunhofer Institue for Integrated Systems, Nuremberg, Germany
{firstname,lastname}@iis.fraunhofer.de
[2] Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
andreas.harth@fau.de

Abstract. Event data preparation is a critical yet time-consuming phase in process mining projects, often slowed down by complex relational data models and a lack of domain knowledge. This paper presents EVErPREP, a novel workflow model that leverages Event Knowledge Graphs to enhance event data preparation for event logs. EVErPREP uses Semantic Web technologies to improve the exploration, extraction, and processing of event data, ultimately improving the quality and interpretability of event data and event logs. The approach is evaluated through a case study at Munich Airport's Baggage Handling System, demonstrating its effectiveness in reducing complexity and improving explainability in event data preparation. By providing a more structured and semantically enriched foundation for process mining, EVErPREP showcases increased efficiency and effectiveness of process mining projects through a semantically enriched foundation.

Keywords: Process Mining · Event Log Preparation · Knowledge Graph.

1 Introduction

Process Mining (PM) has emerged as a research area that provides powerful, data-driven algorithms for improving and understanding processes. Process mining algorithms provide valuable insights into process behavior through automated discovery of process models, detection of deviations from designed processes, and analysis of key performance indicators [21]. As a result, PM has found applications in various domains, from business process workflows [22] to healthcare processes [14] and production [2].

PM algorithms are often based on event logs [20], structured records containing information about process instances, their activities, timestamps, and attributes. While event logs are crucial for PM, preparing them for analysis is a critical and often underestimated phase in PM projects, consuming up to 80% of the total project time [24]. Process Mining methodologies facilitating event log preparation have been proposed, to streamline activities that transition event data and process related domain knowledge to event logs. The Process Diagnostic Method (PDM) [1] was an early PM methodology designed to analyze processes from a single information system. However it did not address complex system environments and the planning phase of PM. The L* Life Cycle

Model [20] has been proposed to extend the scope to include a planning and justification phase at the beginning of a PM project. The Process Mining Project Methodology (PM^2) [22] further extends the preparation phase by separating it into three phases - planning, selection, and processing - and providing more detailed guidelines. Despite these methodological advances, activities related to explainability (e.g., poor documentation, unavailability of process experts), complexity (e.g., complex data structures), and data handling (e.g., identification of relevant data attributes) are still common and not sufficiently addressed by the methodologies [24], thus limiting their efficiency and effectiveness. As a result, many tasks in the preparation phase remain manual, unstructured, and knowledge-intensive [19].

To address knowledge-intensive tasks, researchers explored various processing approaches, data representations formats, and data integration methods. Event Knowledge Graphs (EKG) have demonstrated the ability to provide multiple views of event data, addressing data variety and complexity [12, 15]. RDF-based Knowledge Graphs (KG) have been widely used to facilitate data integration from multiple sources [7, 17, 18]. In addition, the application of ontologies within PM has shown promise in increasing the interpretability of event data by adding a semantic layer to the analysis [3].

Building on these advances, we propose EVErPREP, an EKG-enhanced workflow model for event log preparation. Our workflow model aims to support all event log preparation activities through the use of KGs. EVErPREP enables dynamic aggregation of data to handle different granularities, thereby reducing complexity; it supports the generation of dynamic views of event data, thereby reducing information overload in the preparation phase; and EVErPREP facilitates integrated knowledge acquisition through URIs used as identifiers for attributes and case IDs, thereby improving context availability where necessary.

Through EVErPREP, we aim to demonstrate how KG can facilitate persistent challenges in event log preparation, ultimately improving the efficiency and effectiveness of PM projects. Our contributions include:

- A workflow model for event log preparation using EKG.
- A case study from the workflow model at the Munich Airport.
- An evaluation of the workflow model's impact on event data preparation challenges in knowledge intensive tasks.

2 Background

In recent years, researchers have explored EKGs and Semantic Process Mining (SPM) to address issues related to knowledge-intensive tasks.

EKGs are studied within PM to represent and analyze event data. A KG is a system that captures and integrates information and applies a reasoner and ontologies to derive new knowledge [4]. EKGs extend this definition by including relationships between events, between events and entities, and between entities [8].

The Resource Description Framework (RDF) is widely used to represent KGs, facilitating data integration from disparate sources and multi-stakeholder environments [18]. RDF [13] provides a standardized framework for describing the instance and schema levels of a KGs. SPARQL [9] is used as a powerful query language for pattern matching within graph structures. Adopting these open standards has been shown to improve the discoverability, accessibility, interoperability, and reusability of data [7, 17].

EKGs have two levels to capture information: the instance level and the schema level. The instance level of EKG focuses on data graphs consisting of events and entities, and relationships as edges. The instance level has been studied and documented in several works [6, 11, 12, 15]. Fahland et al. [6] demonstrated that this approach enables complex analysis and facilitates the discovery of complex dynamics within processes, such as subprocesses with dynamic bottlenecks or high workloads. Khayatbashi et al. [11] compared the graph representation with object-centric event logs. They reported that data graphs can better handle relationships of complex, multifaceted processes. In addition, data graphs can overcome memory limitations by storing data on disk for processing and analysis, and can increase the efficiency of analysis by supporting different views. The schema layer specifies domain-specific meta models and the event model within ontologies. Ontologies define and formalize the knowledge of process experts through shared concepts, improving analysis through relationships and properties between and of concepts. SPM applies ontologies within PM to enable data analysis of complex data structures, and increase understandability and interpretability of process models [10, 5, 16]. Eichele et al. [5] presented a method for reasoning about and justifying of process activities, thereby improving process explainability. Nykänen et al. [16] proposed an approach that links ontology structures with event logs to achieve different levels of abstraction.

In summary, by incorporating rich metadata, EKGs and SPM provide a comprehensive framework for understanding, integrating, and analyzing complex event data, addressing many of the persistent challenges of event log preparation and analysis in PM projects.

3 EVErPREP Workflow Model

We present in this section EVErPREP. Our approach to PM preparation relies on EKGs to facilitate the discovery, accessibility, and usability of event data. Furthermore, EVErPREP assumes that data from systems and the knowledge of process experts are already integrated in the EKG. EVErPREP consists of three consecutive phases (see Figure 1):

1. Event Knowledge Graph Exploration: Explores the graph to discover entities, events, and attributes relevant to the specific process mining research questions (PMRQs).
2. Event Log Extraction: Extracts a referenceable minimal semantic event log based on the relevant entities, events, and attributes from the EKG.
3. Event Log Processing: Consolidates and aggregates the event log using the EKG to generate different views and levels of abstraction.

3.1 Phase 1) Event Knowledge Graph Exploration

The goal of the exploration phase is to discover the data sources and domain knowledge integrated into the EKG. The exploration supports the identification of relevant entities for analysis based on the PMRQ and its scope. The input to the exploration phase is the specific PMRQ, the scope of the PM projects, and the EKG. At the end of this phase, analysts have gained valuable knowledge about relevant entities and their context within a domain.

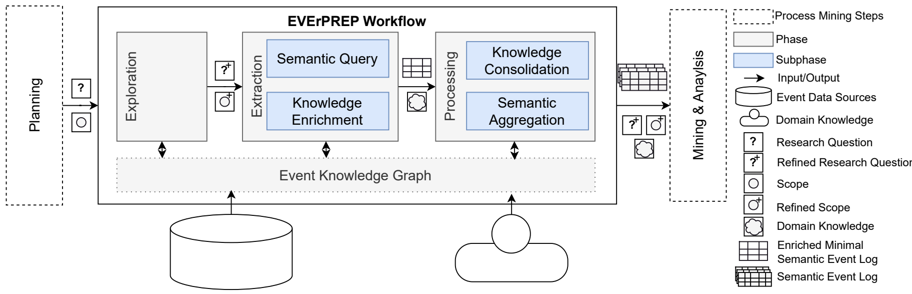

Fig. 1: EVErPREP Workflow

The exploration phase includes three main interconnected activities: graph structure analysis, knowledge retrieval, and creation of initial views. The activities are interrelated and iterative, each informing and refining the others. During graph structure analysis, analysts examine the instance and schema levels of the EKG to gather general information about the domain. Knowledge retrieval involves extracting process-related information relevant to the PMRQ and its scope from the EKG. As the analyst accumulates insights from both the graph structure analysis and the knowledge retrieval, they create and refine the initial views to guide the next steps in the process mining project.

3.2 Phase 2) Event Log Extraction

The extraction phase aims to generate an initial semantically annotated event log. The phase builds upon the results of the exploration phase and consists of two sub phases: semantic querying and knowledge enrichment.

Phase 2.1) Semantic Querying: This phase aims to specify views through semantic queries with semantic filters, simplifying the handling of complex event data early in the preparation process. This activity takes as input the domain knowledge, graph structure, retrieved entities, and PMRQ, resulting in a semantically annotated minimal event log. The event log consists only of columns for the relevant entities and the timestamps. The referencing enables knowledge acquisition through the EKG in subsequent steps.

During semantic querying, analysts formulate specific queries for the EKG, considering key objects identified in the exploration phase. SPARQL queries are applied for semantic querying and filtering of the minimal event log to reduce the amount of extracted data. The query scope is defined using the retrieved entities and the PMRQ. Cases and events of the event log are linked to their respective entities in the EKG using URIs.

Phase 2.2) Knowledge Enrichment: Following semantic querying, the knowledge enrichment phase focuses on obtaining precise descriptions of activities in the process context. This activity has as input the EKG and the minimal semantically annotated event log. The result is an enriched minimal event log extended with additional information for relevant attributes and event descriptions.

Precise descriptions can be obtained from a EKG within the instance level through attributes and entities, or within the schema level through classes. On the instance level, descriptive annotations (e.g., labels and comments) relating to the individual activities are obtained. Additional objects and attributes can be added as resources in PM or for further filtering. On the schema level, class specific information (e.g., class type, class comments and labels) can be obtained. This information aids in understanding the background of specific events or event constellations. The information at class level is particularly relevant for the semantics-based aggregation of events, which is discussed below.

3.3 Phase 3) Event Log Processing

The processing phase takes the enriched data from the extraction phase and refines it further through two sub phases: knowledge consolidation and semantic aggregation.

Phase 3.1) Knowledge Consolidation: The phase aims to integrate enriched information at the desired granularity level within the process context. This phase utilizes references from the semantic event log and the EKG as inputs. The output is a semantic event log incorporated with consolidated knowledge from the EKG, resulting in a more coherent, harmonized representation of key information. The primary challenge addressed in this step is the misalignment between EKG structure and process descriptions. EKGs typically contain varied granular knowledge, manifested as multiple labels and comments, due to inheritance hierarchies and varying detail levels in entity descriptions. This diversity, while rich in information, can complicate process analysis.

The consolidation process involves three key steps: identifying relevant knowledge, integrating diverse perspectives, and consolidation. Analysts identify relevant knowledge by determining which labels and comments from the EKG are most relevant to the PMRQ. They then integrate diverse perspectives by identifying different viewpoints on cases and events within the context of the analysis problem. Finally, the consolidation step involves choosing and combining labels and comments based on their availability and information content, ensuring activities are accurately described.

Phase 3.2) Semantic Aggregation: It aims to create an aggregated semantic event log that provides a higher-level view of the process, facilitating analysis more specific to the PMRQ while maintaining the semantic richness of the original event data. The semantic aggregation has as input the PMRQ and the enriched and consolidated information from the previous steps. The result of the semantic aggregation is an event log (e.g. in XES format) suitable for standard process mining algorithms.

The semantic aggregation process involves two main tasks: defining aggregation rules and performing contextual semantic aggregation. Aggregation rules establish semantic equivalence between event data, even when their syntactic descriptions differ. This allows for meaningful grouping of semantically similar events. Contextual semantic aggregation then applies these rules to the consolidated semantic event log data from the previous subphase, facilitating the creation of higher-level abstract activities that align with the analysis goal.

4 Case Study: Airport Munich

EVErPREP is a workflow model to facilitate the preparation phase of process mining projects in knowledge intensive processes where EKGs are already used. Our evaluation focused on answering the following research questions:

1. (RQ1) How can an Event Knowledge Graph be used to make an event log and its processing more explainable?
2. (RQ2) How can EVErPREP help to reduce complexity in event log handling?
3. (RQ3) How can EVErPREP support the accessibility of data and knowledge in PM projects?

To answer the RQs, we evaluated our approach within a case study of the Baggage Handling System (BHS) at the Airport Munich.

Characteristic of the Process: The BHS is a critical and highly complex system of the airport. It is critical because delays have a direct effect on other processes (e.g. boarding of passengers, departure flights). Complexity is created by thousands of sensors scanning the baggage at different locations within the BHS, multiple involved actors (airlines, baggage handlers), multiple modules transporting the baggage to their destination location, different business variants (handling inbound-, outbound-, and transferbaggage), and different resolution grades (baggage, passenger, container). Through the large amount of messages generated about a specific baggage, a syntactic analysis of the BHS without further domain knowledge is not suitable.

Process Mining Scope and Research Question: The Airline AirL (pseudonymized) offers a number of scheduled flights to a variety of destinations, with passengers checking in at different terminals and boarding gates. Once checked in, the baggage items are sorted transported within the different modules of the BHS until they reach the correct destination. To identify bottlenecks and understand the process of AirL better, the following PMRQ needs to be addressed: *"What is the actual baggage transfer process regarding automated baggage handling within Terminal 1 for Airline AirL?"*

Event Knowledge Graph: The EKG of the BHS uses for the schema level a domain-specific ontology, and the Simple Event Model Ontology (SEM) [23]. The domain-specific ontology describes the domain of the BHS with static concepts, such as sensor type, baggage types, airlines, and passengers. The SEM is used as a process metamodel to model events with their actors, activities, locations, and time. Actors are specific baggages, passengers, containers, and airplanes. Locations are described through the position of the sensors within the BHS. Activities are feedback from sensors that scanned baggages at specific positions within the BHS. RDF is used to instantiate the ontologies and the data from the BHS. Reasoning with OWL is used to derive hierarchies from the domain model and link differently named data points from the data sources with each other under a standardized description. Domain knowledge was derived from documentations, location specific information, and domain experts. The domain knowledge is attached to the instance and schema level concepts and relations of the EKG. An excerpt of the EKG can be seen in Figure 2. We used prefixes to shorten the URIs within the figure and the SPARQL query to improve readability. Common prefixes such as *rdfs:* can also be found on `https://prefix.cc`.

4.1 Execution of the Workflow

In this section we present the execution of EVErPREP on Munich Airport case study to answer their PMRQ. We describe the execution and the results of each step in the EVErPREP workflow.

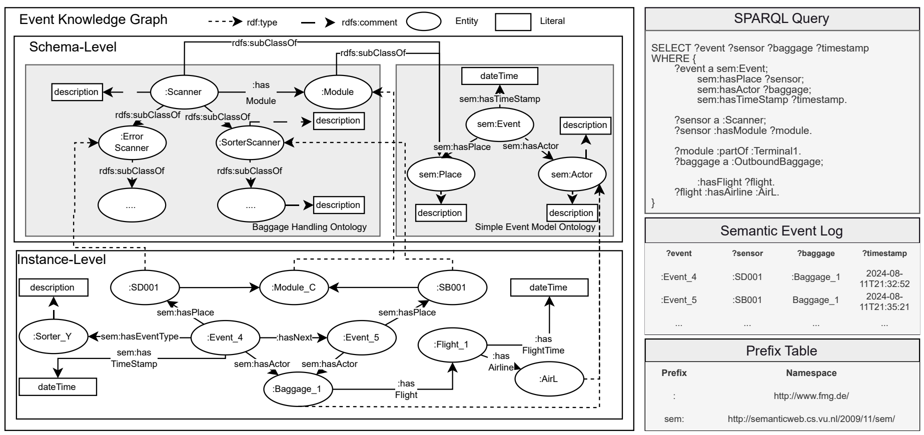

Fig. 2: Snippet of the EKG (left) with reduced relationships and basic SPARQL query and results for semantic querying (right)

Phase 1) Exploration The goal of this phase was to discover and understand the relevant entities, relationships, and context within the EKG that were relevant to answering the PMRQ. Therefore, the EKG was explored with SPARQL queries. The initial queries were primarily designed to discover and describe concepts to answer the PMRQ. The initial queries focused on schema-level descriptions from the domain ontology to support a high-level understanding of the EKG. Based on the schema-level concepts, instances such as terminals, modules, flights, airlines, sensors, and baggage types were discovered. In addition, annotations within the BHS, such as descriptions and provenance information, were retrieved to provide additional context to the concepts. Based on the metadata and the PMRQ, a selection of relevant entities was discovered. To describe the process, relationships between entities were retrieved based on the SEM ontology and the domain ontology. The result of the exploration step with the EKG of the airport contained four core concepts, the baggage, the sensors, the locations associated with the sensors, and the events associating the baggage with the sensors.

Phase 2.1) Semantic Querying: This phase focused on creating a minimal semantic event log through semantic queries with filters, based on the insights gained from the exploration phase. The scope of the semantic query was limited to the four main concepts identified in the exploration phase and the constraints from the PMRQ. The constraints included the location to Terminal 1 and the baggage handled by the airline AirL. The events modeled with SEM were retrieved with a SPARQL query to create the semantic event log. Figure 2 shows the query and a slice of its results, which include the timestamp and URIs for the sensors, baggage, and events.

Phase 2.2) Knowledge Enrichment: The enrichment phase aimed at enriching the minimal semantic event log with detailed descriptions and contextual information for activities and events using the data within the EKG. This was limited to information that contributed to a self-explanatory process model. Due to the scope of the PMRQ, only content related to the activity and background of the sensor activity was used. To do this, the enrichment scripts traversed the EKG using the knowledge gained from exploring the structure and context of the entities. Figure 3 shows the results of the knowledge enrichment, including multiple semantic descriptions for sensors at the instance and schema levels.

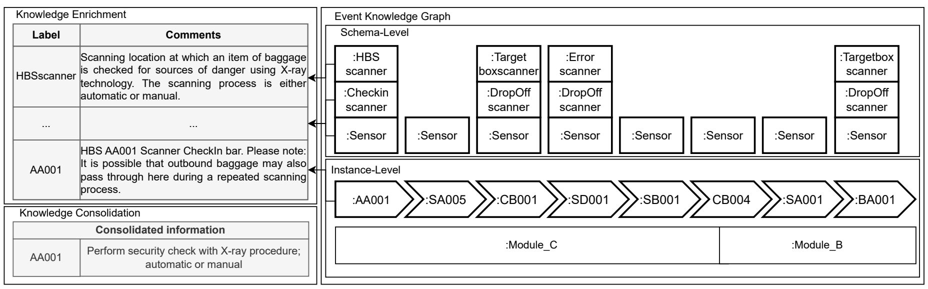

Fig. 3: Knowledge enrichment and consolidation for the sensor AA001 to combine different descriptions from the instance and schema level based on the scope and the PMRQ

Phase 3.1) Knowledge Consolidation: This phase involved integrating and harmonizing the enriched information at the desired granularity within the process context, PMRQ, and analyst understanding. Based on the results of Phase 2.2, the process model derived from the semantic event log would have contained several semantically similar variants. Consolidation was necessary because the EKG was not specifically designed to answer the specific PMRQ and therefore contains differently structured descriptions at different levels. Through consolidation, semantically similar descriptions were aligned using the OpenRefine[3] tool. An example of such an alignment is shown in Figure 3 for the sensor with the URI *:AA001*. The alignment for all activities has been recorded in the semantic event log in a new column containing the consolidated activity description.

Phase 3.2) Semantic Aggregation: Semantic aggregation aimed to provide a higher-level view of the process by aggregating events based on semantic similarities, facilitating PMRQ-specific analysis. Thus, the aggregation is based on the enriched and consolidated information from the semantic event log. Based on the PMRQ, hierarchies were created within the domain model for sensor types and baggage types. For each of the hierarchies, domain knowledge based on comments and labels was retrieved with SPARQL queries to create aggregations of events (see Figure 4). The final semantic event log was based on the PMRQ associated with different levels of abstraction of the BHS to facilitate dynamic views of the process. From the semantically enriched and

[3] `https://openrefine.org/`

aggregated event log, different views were extracted in XES format for the following process mining analysis.

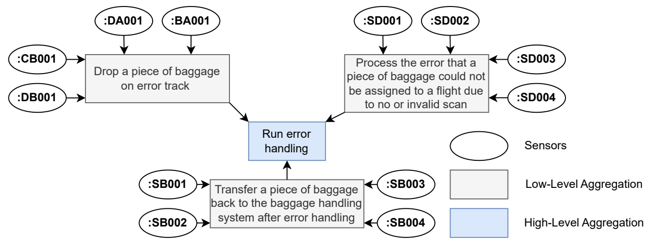

Fig. 4: Semantic aggregation to achieve a higher-level view of the process

4.2 Observations

This section presents key observations from its application to the case study regarding explainability, complexity, and data handling when EVErPREP is used for event log preparation.

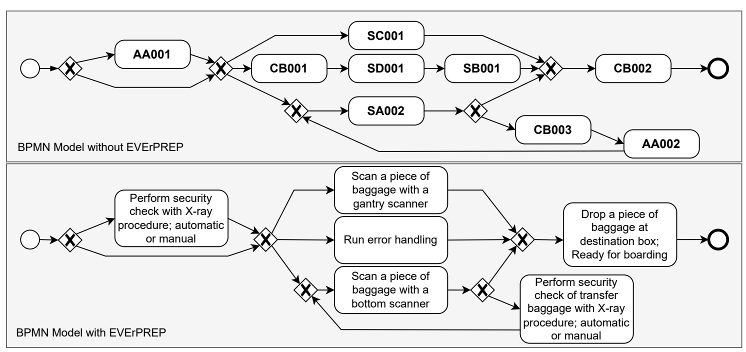

Fig. 5: Excerpt of the process model of a flight without and with EVErPREP

Explainability: A persistent challenge in PM is interpreting event logs with insufficient context, leading to misunderstandings of process behaviors [24]. Enrichment with knowledge from the EKG has significantly improved the clarity of activity descriptions within their overall context, making individual events and their sequences more comprehensible. Furthermore, the flexibility of this enrichment allows for adjusting the level of detail in process descriptions. For instance, Figure 5 shows for the activity *AA002* the more explainable semantic description of the activities aligned with the scope of the PMRQ. This adaptability enables tailoring the information to match the analysts' knowledge level, based on the content available in the EKG.

Complexity: Managing the inherent complexity of event logs, particularly those with numerous variants, often hinders effective process analysis and interpretation [24].

Semantic aggregation, built upon semantic enrichment, has effectively reduced complexity in the extracted event logs. This approach has streamlined activity descriptions to essential, concise components. Figure 5 shows the impact of EVErPREP, its associated knowledge consolidation, and the final semantic aggregation on the process model extracted from the AirL event log for the three activities *CB001*, *SD001*, and *SB001*. The results show a reduction in complexity and improved handling of the event log for subsequent analyses, particularly in relation to more concise case variants.

Event Data Handling: A well-known challenge in PM is the handling of large event logs that are merged from a wide variety of systems [20]. The pre-existing data integration of airport systems within the EKG provides valuable information, allowing early identification of attributes relevant to specific PMRQ. This integration enables semantic querying of only essential data for core event log requirements at an early stage. Consequently, attributes irrelevant to PMRQ can be excluded from queries and further data pre-processing, contributing to a more efficient workflow.

In summary, the use of EVErPREP in the case study demonstrates that the explainability of objects, resources, and activities; the reduction in complexity; and the simplified accessibility and usability of event data contribute to a notable enhancement in the effectiveness and efficiency of the data processing step.

5 Conclusion and Future Work

This paper introduces EVErPREP, an event data preparation workflow model designed to address challenges in knowledge-intensive domains. These domains often require time-consuming event data preparation to handle interpretability and complexity issues within the data. Our case study at Munich Airport's Baggage Handling System (BHS) demonstrates the effectiveness and efficiency of event data preparation and analysis with EVErPREP. EVErPREP leverages Event Knowledge Graphs (EKGs) to improve three key aspects of event log preparation. First, it enhances the explainability of event data by providing contextual meaning. Second, it facilitates improved comprehension of complex event data through the consolidation and semantic aggregation of information, based on the structure and semantics of the underlying EKG. Third, EVErPREP improves the handling of complex event data bases by dynamically integrating attributes into the event log through dereferenceable URIs. Our case study results show that EVErPREP improves the explainability for non-domain experts of event data within the BHS. It also reduces complexity by providing more concise and meaningful activity descriptions. Furthermore, the approach simplifies data accessibility and usability, contributing to enhanced effectiveness and efficiency in the data processing step.

Future research directions include investigating the influence of EVErPREP on the overall data quality of processed event logs. Another promising area is exploring methods to instantiate and integrate domain knowledge derived from EVErPREP and process mining analysis back into the EKG. This could include incorporating consolidated activity-based descriptions or derived process models as additional annotations. Furthermore, EVErPREP could be enhanced by including semi-automated approaches for consolidation and aggregation using natural language processing techniques to improve the efficiency of event data preparation. Such enhancements to the EVErPREP workflow would further streamline the process mining workflow and reduces key challenges of process mining [24].

In conclusion, EVErPREP offers a promising approach for streamlining event data handling, pre-processing, and interpretation in knowledge-intensive domains. By addressing key challenges in event log preparation, it paves the way for more efficient and effective process mining analyses, particularly in complex environments like airport operations.

Author Contributions and Acknowledgements

Conceptualization, P.F. and R.D.; methodology, P.F. and R.D.; writing - original draft preparation, P.F. and R.D.; writing - review and editing, P.F. and R.D; supervision A.H.; funding acquisition, A.H.

This work was funded by the Bayerisches Verbundforschungsprogramm (BayVFP) des Freistaates Bayern through the KIWI project (grant no. DIK0318/03) and by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through the Antrieb 4.0 project (Grant No. 13IK015B).

References

[1] Melike Bozkaya et al. "Process Diagnostics: A Method Based on Process Mining". In: *2009 International Conference on Information, Process, and Knowledge Management*. 2009. DOI: 10.1109/eKNOW.2009.29.

[2] Alexandre Checoli Choueiri and Eduardo Alves Portela Santos. "Multi-Product Scheduling through Process Mining: Bridging Optimization and Machine Process Intelligence". In: *Journal of Intelligent Manufacturing* 6 (2021). DOI: 10.1007/s10845-021-01767-2.

[3] A. K. Alves de Medeiros et al. "An Outlook on Semantic Business Process Mining and Monitoring". In: *On the Move to Meaningful Internet Systems*. Lecture Notes in Computer Science. 2007. DOI: 10.1007/978-3-540-76890-6_52.

[4] Lisa Ehrlinger and Wolfram Wöß. "Towards a Definition of Knowledge Graphs". In: *1st International Workshop on Semantic Change & Evolving Semantics*. 2016.

[5] Simon Eichele, Knut Hinkelmann, and Maja Spahic-Bogdanovic. "Ontology-Driven Enhancement of Process Mining With Domain Knowledge". In: 2023.

[6] Dirk Fahland. "Process Mining over Multiple Behavioral Dimensions with Event Knowledge Graphs". In: *Process Mining Handbook*. Cham, 2022. DOI: 10.1007/978-3-031-08848-3_9.

[7] Michael Freund, Julian Rott, and Andreas Harth. "FAIR Internet of Things Data: Enabling Process Optimization at Munich Airport". In: *ESWC: Extended Semantic Web Conference*. Hersonissos, Greece, 2024. DOI: 10.1007/978-3-031-60626-7.

[8] Saiping Guan et al. "What is Event Knowledge Graph: A Survey". In: *IEEE Transactions on Knowledge and Data Engineering* (2021).

[9] Steven Harris and Andy Seaborne. *SPARQL 1.1 Query Language*. W3C Recommendation. W3C, 2013.

[10] Fadilul-lah Yassaanah Issahaku et al. "An Overview of Semantic-Based Process Mining Techniques: Trends and Future Directions". In: *Knowledge and Information Systems* (2024). DOI: 10.1007/s10115-024-02147-x.

[11] Shahrzad Khayatbashi, Olaf Hartig, and Amin Jalali. "Transforming Event Knowledge Graph to Object-Centric Event Logs: A Comparative Study for Multi dimensional Process Analysis". In: *Conceptual Modeling*. Cham, 2023. DOI: 10.1007/978-3-031-47262-6_12.

[12] Eva L. Klijn et al. "Event Knowledge Graphs for Auditing: A Case Study". In: *Process Mining Workshops*. Cham, 2024. DOI: 10.1007/978-3-031-56107-8_7.

[13] Markus Lanthaler, David Wood, and Richard Cyganiak. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation. W3C, 2014.

[14] Ronny S. Mans et al. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*. SpringerBriefs in Business Process Management. Cham, 2015. DOI: 10.1007/978-3-319-16071-9.

[15] Milad Naeimaei Aali et al. "Clinical Event Knowledge Graphs: Enriching Healthcare Event Data with Entities and Clinical Concepts - Research Paper". In: *Process Mining Workshops*. Cham, 2024. DOI: 10.1007/978-3-031-56107-8_23.

[16] Ossi Nykänen et al. "Associating Event Logs with Ontologies for Semantic Process Mining and Analysis". In: *Proceedings of the 19th International Academic Mindtrek Conference*. Tampere Finland, 2015. DOI: 10.1145/2818187.2818273.

[17] Núria Queralt-Rosinach et al. "Applying the FAIR Principles to Data in a Hospital: Challenges and Opportunities in a Pandemic". In: *Journal of Biomedical Semantics* 1 (2022). DOI: 10.1186/s13326-022-00263-7.

[18] Julian Rott et al. "Breaking down Barriers with Knowledge Graphs: Data Integration for Cross-Organizational Process Mining". In: 2024. DOI: 10.1007/978-3-031-56107-8_38.

[19] Vinicius Stein Dani et al. "Towards Understanding the Role of the Human in Event Log Extraction". In: *Business Process Management Workshops*. Cham, 2022. DOI: 10.1007/978-3-030-94343-1_7.

[20] Wil van der Aalst et al. "Process Mining Manifesto". In: *Business Process Management Workshops*. Berlin, Heidelberg, 2012. DOI: 10.1007/978-3-642-28108-2_19.

[21] Wil M. P. van der Aalst et al. "Service Interaction: Patterns, Formalization, and Analysis". In: *Formal Methods for Web Services*. Berlin, Heidelberg, 2009. DOI: 10.1007/978-3-642-01918-0_2.

[22] Maikel L. van Eck et al. "PM^2: A Process Mining Project Methodology". In: *Advanced Information Systems Engineering*. Lecture Notes in Computer Science. Cham, 2015. DOI: 10.1007/978-3-319-19069-3_19.

[23] Willem Robert van Hage et al. "Design and Use of the Simple Event Model (SEM)". In: *SSRN Electronic Journal* (2011). DOI: 10.2139/ssrn.3199512.

[24] Moe Thandar Wynn et al. "Rethinking the Input for Process Mining: Insights from the XES Survey and Workshop". In: *Process Mining Workshops*. Cham, 2022. DOI: 10.1007/978-3-030-98581-3_1.