

Gaining Insights into Data Sharing Process Among Multiple Stakeholders through Process Mining and Social Network Analysis

Lihua Wu ¹
Ying Wang ^{2,*}

*Corresponding author

¹ Student, Beijing Jiaotong University, Beijing, China, 23120637@bjtu.edu.cn

² Associate Professor, Beijing Jiaotong University, Beijing, China, ywang1@bjtu.edu.cn

ABSTRACT

As the data sharing process spans multiple stakeholders, its complexity increases significantly, making it increasingly difficult to track and understand the source, circulation, and usage of data. This not only limits the comprehensive utilization of data but also affects the establishment and maintenance of trust, thereby hindering cooperation efficiency. To address this, this paper integrates two analytical tools—Process Mining and Social Network Analysis (SNA)—to deeply explore the data sharing process among multiple stakeholders. By combining these two methods, we can more finely depict the data flow trajectories between different stakeholders and gain a deeper understanding of the structural characteristics of the data sharing network, such as the frequency, direction, and intensity of data sharing. This in-depth analysis helps identify potential barriers and opportunities within the data sharing process, providing new insights for promoting data compliance, efficient circulation, and utilization.

Keywords: Process mining, social network analysis, data sharing, multi-stakeholder.

INTRODUCTION

Data is the raw, unprocessed record of facts, whereas information is the knowledge that data, when interpreted and contextualized, can provide support for decision-making. Effective sharing and use of data is key to achieving information transparency and facilitating cross-organizational collaboration. The flow of data and the process of converting it into information are particularly important in the process of sharing data across organizations, as they affect not only the speed and quality of decision-making, but also trust and the efficiency of collaboration between organizations. Therefore, an in-depth look at managing and optimizing the data sharing process ensures that data can be used efficiently and compliantly throughout its lifecycle, thereby driving digital transformation in organizations and industries.

In the current trend of digital transformation, data sharing processes have become a significant driver for cross-organizational collaboration, accelerated decision-making, and innovation stimulation. Effective data sharing is essential to facilitate intelligent decision-making and collaboration, which is particularly evident in areas such as environmental decision support, building information modeling (BIM) projects, and hazardous waste management (Laurila-Pant et al., 2019; Song et al., 2021; Wang et al., 2020). However, the complexity of data sharing increases significantly with the number of stakeholders involved, which complicates tracking the origin, flow, and use of data, which in turn limits the full utilization of the data and may affect the establishment and maintenance of trust and ultimately the efficiency of collaboration.

To address this challenge, this study employs two powerful analytical tools, process mining and social network analysis (SNA). Process mining provides a clear way to understand the operation of complex systems by analyzing event logs to reveal and optimize business processes. Meanwhile, social network analysis (SNA) focuses on revealing patterns of interactions between individuals or organizations by visualizing key players and information flow paths in the network structure. Both approaches have shown significant results in supply chain management, customer experience enhancement, and team collaboration (Dorofeev et al., 2023; Schmitt et al., 2023; Sitova et al., 2023).

By integrating these two approaches, we can provide a finer-grained picture of data flow trajectories among different stakeholders and gain insights into the structural characteristics of data sharing networks, such as the frequency, direction, and intensity of data sharing processes. Such in-depth analysis can help identify potential barriers and opportunities in the data sharing process, providing valuable insights for promoting data compliance, efficient flow and utilization, and enhancing trust among collaborating parties.

The objective of this study is to delve into the data sharing process among multiple stakeholders by combining process mining and Social Network Analysis. We aim to uncover the intrinsic patterns of the data sharing process, promote data compliance and efficient circulation, enhance the effectiveness of collaborative networks, and provide scientific evidence and practical guidance for future data management strategies. Through this research, we expect to contribute to building a more efficient, transparent, and trustworthy data sharing environment, offering references and insights for subsequent work.

RELATED WORK

Data Sharing Process Among Multiple Stakeholders

Cross-organizational data sharing is a core issue in modern information management, aiming to optimize resource utilization, facilitate decision-making, and foster innovation. Early studies, such as those by Boh & Yellin (2006), investigated the influence of governance mechanisms on the use of enterprise architecture standards from an IT management perspective, providing a governance perspective on organizational-level data sharing. Barua et al. (2007) revealed the positive impact of information complementarity on organizational performance, indicating that high-quality data sharing can enhance the overall benefits of workgroups. Grandison et al. (2007) introduced the concept of sticky strategy execution to balance the needs of medical data sharing with privacy protection, while Plengsaeng et al. (2014) focused on non-technical barriers in transboundary water resource management, calling for a shift in organizational thinking to promote data sharing.

Alshwayat et al. (2021) focused on the construction of trust, collaboration, and knowledge-sharing cultures within formal organizations. Duncan et al. (2020) and Hazée et al. (2020) discussed the value and risks of data sharing in life sciences and product service systems, respectively. Naidu & Chakravarthy (2021) explored the security and privacy of data sharing from the perspectives of legal policy and local politics, revealing the challenges and barriers faced by data sharing.

Fernandez et al. (2020) proposed a data market platform to address the issue of data silos, increasing data value through market mechanisms. Guentherman et al. (2021) shared experiences in building an inter-organizational data-sharing management information system (MIS), emphasizing cross-institutional collaboration. Stalla-Bourdillon (2021) discussed the EU data governance act and the UK data standards board's recommendations, aiming to enhance data-sharing mechanisms and trust. Greene et al. (2021) developed the DSUA ontology to facilitate standardized cross-organizational sensitive data sharing. Herurkar et al. (2024) utilized representation learning and federated learning technologies to address privacy and competition issues in financial table data sharing, providing solutions and mechanisms for cross-organizational data sharing.

Process Mining

Process mining, as an interdisciplinary field that integrates data science, business administration, and information science, has continuously deepened its theoretical framework and broadened its application scope since its inception. The ongoing development of process mining technology enables it to demonstrate its unique value in more fields, becoming a key force in driving corporate digital transformation and process optimization.

The research by Diba et al. (2019) and van Zelst et al. (2021) elucidates the fundamental principles of process mining, which involves revealing the intrinsic structure, variability, and performance metrics of business processes through automated analysis of event logs.

As process mining technology matures, its application domains have rapidly expanded. The work by Dorofeev et al., (2023) in supply chain management demonstrates how process mining helps businesses identify process bottlenecks and potential risk points, enhancing overall efficiency and stability. Sitova et al. (2023) focus on customer relationship management, using process mining to analyze customer behavior and needs, optimizing customer experiences, and increasing customer satisfaction and loyalty. In manufacturing, Schmitt et al. (2023) show how process mining contributes to improving production efficiency and product quality. In the healthcare sector, Park et al. (2023) use process mining to enhance patient care workflows, allocate medical resources rationally, and significantly improve the quality and efficiency of medical services.

In recent years, process mining technology has become more intelligent with the advancement of artificial intelligence and machine learning. The research by Pasquadibisceglie et al. (2020) demonstrates how process mining can be combined with computer vision and deep learning to handle more complex data types, such as images and videos, significantly expanding its application range. The Multi-Perspective Modeling method proposed by Berti & van der Aalst (2020) allows process mining to more comprehensively consider various factors within processes, such as resource allocation and object interactions, to achieve more accurate process models.

Social Network Analysis

The application of Social Network Analysis has evolved from theoretical construction to the deepening of specific scenarios, and finally to comprehensive research on cross-domain complex systems, demonstrating its continuous expansion in both depth and breadth. Early studies focused on the construction of foundational theories and methodologies, which were subsequently deepened into applications in specific scenarios, and ultimately evolved into comprehensive research on cross-domain complex systems.

Lazega et al. (1995) were the first to apply SNA to stakeholder analysis, and Prell et al. (2009) laid the foundation for subsequent research. The theoretical framework and applied research further enriched the academic foundation of SNA. The methods of social network analysis have also been applied in specific scenarios such as construction and healthcare, achieving goals like workflow optimization (Durojaiye et al., 2022; Pan & Zhang, 2021). The research by Pan & Zhang (2022) demonstrated the application of SNA in cross-domain complex systems analysis, providing in-depth insights for business applications through multi-layer network modeling and clustering techniques based on event logs of constructed information models.

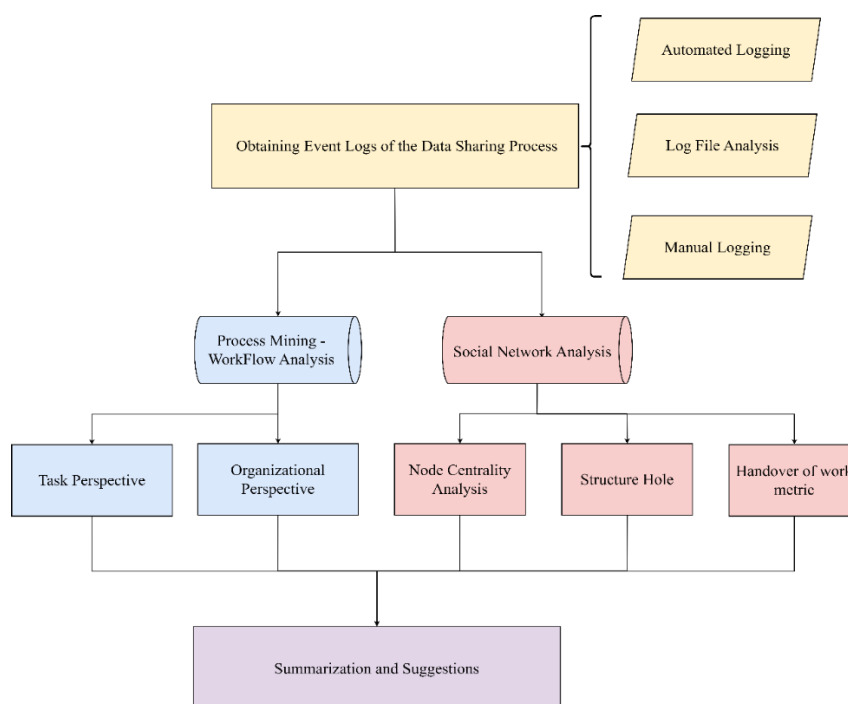
Recent studies have further shown the potential of SNA in automated process discovery, compliance checking, and social network mining. Pan & Zhang (2021) developed a framework for automatically discovering processes from BIM event logs to optimize construction workflows. Durojaiye et al. (2022) utilized social network analysis to reveal the day-night differences in pediatric medical teams, offering a new perspective in the healthcare field. Mustroph et al. (2024) proposed a method for detecting compliance deviations by combining natural language text and event logs.

The complex and subtle interactions between social structures and individual behaviors underscore the critical role of social network analysis in understanding social phenomena, predicting individual behavior, and designing effective social interventions. This has provided new research perspectives and empirical evidence for fields such as sociology, psychology, and public policy.

This paper aims to conduct an in-depth study of the data sharing process among multiple stakeholders through process mining and social network analysis. By integrating the ability of process mining to reveal business processes and the function of social network analysis to display key participants and information flow paths within network structures.

METHODOLOGY

This paper conducts an in-depth study of the data sharing process through the following steps: First, we obtained standardized data sharing process log records through strict data collection standards. These log records provide the foundational dataset for our research. Secondly, we analyze the workflow, and use the four mining methods of interactive data-aware heuristics (iDHM) to analyze the organizational structure, conduct in-depth mining of the data sharing process logs, and identify key patterns and trends in the data sharing process. Third, we utilized indicators and methods from Social Network Analysis (SNA) to analyze the data sharing process logs from an organizational perspective, revealing the social structures and dynamics within the data sharing process. Finally, we will summarize the research findings and propose corresponding strategies and recommendations.



Source: This study.

Figure 1: Methodology for Data Circulation and Ownership Optimization Framework

Step 1: Obtaining Event Logs of the Data Sharing Process

Obtaining and standardizing event logs for the data sharing process is the first step in our research methodology, which is crucial for understanding the dynamics of data sharing, identifying potential issues, and optimizing workflow. As the foundation of our study, the event logs provide detailed records of data sharing activities, including key information such as participants, actions, timestamps, and data sources. This information not only increases the transparency of the data sharing process, enabling all stakeholders to track the flow and usage of data, but also helps to clarify accountability, facilitating tracking and correction when issues arise. Moreover, log data can assist in analyzing bottlenecks and efficiency issues within the data sharing process, providing a basis for improvements and ensuring compliance.

To ensure the completeness and consistency of event logs, we designed a basic event log template that includes at least the following key fields: caseID, used to uniquely identify each data sharing case; eventID, used to uniquely identify each event;

actors_id, identifying the individuals or systems involved in the data sharing activities; activity_id, describing specific actions or events within the data sharing activities; Timestamp, recording the exact time the event occurred; and Data source, indicating the origin of the data, such as sensor data, manual records, or data from different systems. Considering that data may come from various sources, we added a "Data source" column to the template to distinguish the origins of data and ensure its accuracy and reliability.

Obtaining event logs for the data sharing process can be achieved through various methods, each with its own advantages and challenges. First, automated logging is one of the most common methods, which captures events of data sharing automatically through sensors, software systems, or application programming interfaces (APIs). The advantage of this approach lies in its ability to collect a large amount of data in real-time and accurately, reducing the likelihood of human error. However, the challenge with automated logging lies in ensuring the stability of the system and the integrity of the data, while also addressing privacy and security concerns.

Second, log file analysis is another effective method of recording, which involves analyzing log files generated by existing systems to extract information about data sharing. The benefit of this method is that it leverages existing log data without incurring additional recording costs. However, the difficulty lies in the potential for inconsistent formats, incomplete information, or difficulties in parsing the log files, which requires additional work to standardize and integrate the data.

Third, manual logging, which involves observing and recording data sharing events through human observation, is particularly useful in scenarios where automation is not feasible or where highly customized recording is required. The advantage of manual logging is its flexibility and in-depth understanding of specific situations. However, the challenge with manual logging is that it is time-consuming and can be subject to the subjectivity of the recorder, which may lead to inconsistencies and biases in the data.

Step 2: Process Mining- Workflow Analysis

This paper selects the Interactive Data-aware Heuristic Miner (iDHM) to mine and visualize the data sharing process from event logs of the data sharing process (Mannhardt et al., 2017). The reason for this choice is that iDHM offers a powerful interactive approach to understanding and revealing business processes. The iDHM is capable of handling complex event logs that contain noise and unstructured processes, and through its data-aware capabilities, it can leverage data attributes within event logs to enhance the discovery of process dependencies. Moreover, iDHM's visualization features, particularly the use of Causal Net (C-Net) representation, provide a clear and intuitive way to present process models, including the causal relationships and control flows between activities. This visualization method allows for a quick grasp of the structure and dynamics of the process, thus facilitating more effective analysis and understanding. Through iDHM, this paper transforms complex process data into visual representations that are easy to interpret and communicate, which is crucial for process discovery and knowledge sharing. The specific steps are as follows.

(1) Discover Dependency Relations. Identifying Dependency Links. Initially, the collection of dependencies is established. An individual dependency, noted as (a, b), signifies that activity a precedes and influences activity b, which is visually represented by an arrow connecting the two. For a dependency to be recognized, it must be robust, surpass a preset threshold for dependency strength, and appear more frequently than a specified observation frequency threshold. Various heuristic methods have been developed to uncover and measure these dependencies. We offer the option to utilize the Flexible Heuristics Miner, the Alpha Miner, Fuzzy Miner, or to integrate these methods by averaging the dependency values they identify.

(2) Identify Split and Join Points. The C-Net's distinct split and join points, represented by its I/O bindings, must be identified to ensure the model's semantic accuracy. These bindings appear as dots along the C-Net's flow paths, with separate dots indicating XOR splits and linked dots signifying AND splits.

(3) Uncover Decision Criteria. In this phase, uncover the criteria that will trigger the activation of specific output bindings. These criteria, or decision rules, are highlighted with a double border around the associated bindings. For a deeper understanding, detailed decision rule information is accessible through a simple right-click action on the binding. The decision rules are subject to a quality threshold, which can be used to filter them. The iDHM incorporates two distinct decision mining techniques: one that utilizes C4.5 decision trees and another that focuses on identifying overlapping decision rules.

Step 3: Social Network Analysis

This study employed social network analysis techniques to measure the cooperative efficiency among participants in drills. By utilizing various tools of social network miners, we can gain a deep understanding of the interaction patterns among participants and assess their impact on process execution. The focus of this stage of research is on conducting social network organizational analysis on the event logs of data flow to achieve effective identification of the roles and interactions of participants in the data sharing process. Through this method, we can not only reveal the potential cooperative patterns within the process but also identify and reinforce those cooperative factors that are crucial for the successful execution of the process.

Definition 3.1 Density

The density (D) serves as a prominent measure of the overall relationships within a network, reflecting the ratio of the actual number of direct connections between network actors to the total number of potential connections. The formula for density is given by:

$$D = \frac{L}{\frac{n(n-1)}{2}}, \quad (1)$$

where L denotes the total number of direct links observed in the network, and n represents the total count of nodes. In the context of an undirected graph, the maximum number of possible links is calculated by $\frac{n(n-1)}{2}$. Consequently, the density value falls within the range of 0 to 1, with 0 indicating no connections and 1 signifying that every possible connection is present.

Definition 3.2 Node degree

The node degree quantifies the number of connections a node has with its adjacent nodes. This measure reflects the comprehensive extent of relationships that a node, labeled as i, maintains with other nodes in the network. With respect to the orientation of these relationships, there are two distinct aspects: the in-degree and the out-degree of the node. The in-degree $s(D_i^{in})$ pertains to the number of edges that are directed towards the node, signifying the receipt of relationships, whereas the out-degree (D_i^{out}) pertains to the number of edges that originate from the node, indicating the transmission of relationships. Consequently, the calculation of the in-degree and out-degree for a node i can be expressed as follows:

$$D_i^{in} = \sum_k a_{ik}, \quad (2)$$

$$D_i^{out} = \sum_k a_{ik}, \quad (3)$$

where the in-degree of a node i is denoted D_i^{in} , which corresponds to the number of incoming edges (a_{ik}) that arrives at the node i; and the out-degree of a node i denoted D_i^{out} is the number of outgoing edges (a_{ik}) emanating from node i.

Definition 3.3 Structural holes

The Structural Holes Theory was proposed by sociologist Ronald Burt in 1992 and primarily focuses on the structural configurations within social networks and how these configurations influence the competitive advantages of individuals or organizations. In a social network, structural holes refer to the gaps or breaks between individuals or groups that lack direct connections. Individuals positioned at structural holes, due to their ability to bridge different communities or information silos, often gain access to unique information and resources, thereby securing a favorable position in competition.

Effective size: When contacts of a node are interconnected, its ego network (a first-level network centered around a node) exhibits redundancy. The effective size of a node's ego network is a measure of the non-redundant part of its relationships. p_{uw} refers to the normalized weight between u and w, while m_{vw} refers to the normalized weight between v and w divided by the largest normalized weight between v and its neighbors.

$$e(u) = \sum_{v \in N(u) \setminus \{u\}} \left(1 - \sum_{w \in N(v)} p_{uw} m_{vw} \right) \quad (4)$$

Efficiency is the effective size of a node divided by its degrees, a normalized measure of how non-redundant a node's relationship is.

Constraint measures the degree to which a node is constrained by its social network. A higher constraint near a node means higher network density and a lower number of structural holes.

$$c(v) = \sum_{w \in N(v) \setminus \{v\}} \ell(v, w) \quad (5)$$

A node's constraint is the sum of its local_constraint with all neighboring nodes

$$\ell(u, v) = \left(p_{uv} + \sum_{w \in N(v)} p_{uw} p_{vw} \right)^2 \quad (6)$$

The hierarchy indicates the degree to which all constraints are concentrated in a single connection. When a node's hierarchy is equal to 1.0, it means that all constraints are concentrated on a single connection.

Definition 3.4 Handover of work metric

This metric for (possible) causality is based on the frequency of work transfers among actors in the process.

The handover of work metric assigns a weighting W_{ij} to the relationship between actors i and j. For any two actors $V_i, V_j \in N$, where N is the set of the nodes in a network; if W_{ij} is above a given threshold τ , such a relationship (V_i, V_j) will be placed in a set R , i.e., $(V_i, V_j) \in R$. In this way, a weighted graph (N, R, W) can be built. Let L be an event log, then the basic handover of a work metric \triangleright_L defines $V_i \triangleright_L V_j = W_{ij}$ in $R = \{(V_i, V_j) \in N \times N | V_i \triangleright_L V_j \neq \mathbf{0}\}$ as follows:

$$-V_i D_L V_j = \left(\sum_{c \in L} |V_i \triangleright_c V_j| \right) \left(\sum_{c \in L} |C| - 1 \right), \tag{7}$$

$$-V_i \dot{\triangleright}_L V_j = \frac{(\sum_{C \in L \wedge V_1 \triangleright_c V_2} 1)}{|L|} \tag{8}$$

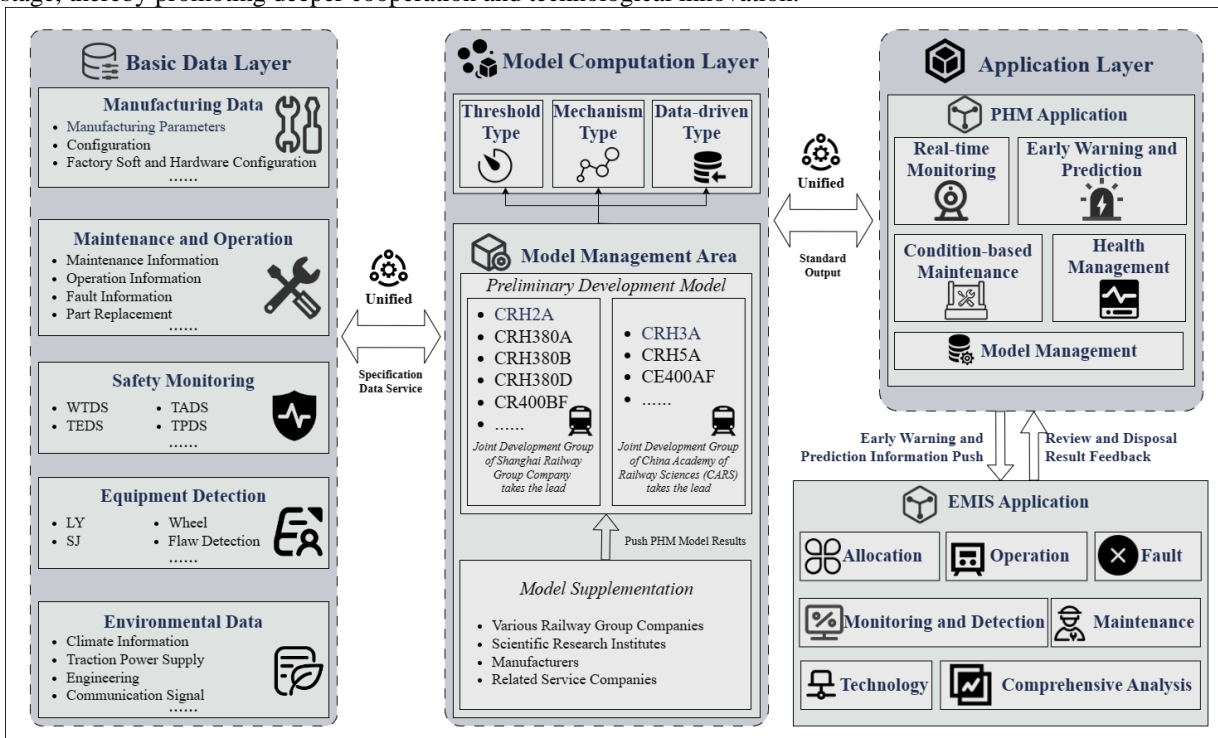
where $V_i \triangleright_L V_j$ means the handover of the work metric from V_i to V_j (based on the **logL**), obtained by dividing the total number of direct successions from V_i to V_j in a **log** by the maximum number of possible direct successions in the **log.C** indicates a case in the **logL**. The term $V_i \dot{\triangleright}_L V_j$ is the handover of a work metric ignoring multiple transfers within one instance .

CASE STUDY

Scenario

Considering the growing demand for safety, efficiency, and reliability in domestic railway transportation, especially in the context of the high-intensity operation of over 4,000 standard high-speed trains within the Chinese high-speed railway network, optimizing asset maintenance strategies to enhance operational efficiency, reduce costs, and strengthen safety has become an urgent priority. Currently, the transition from condition-based maintenance (CBM) to predictive health management systems (PHM) is gradually realizing intelligent and precise maintenance decision-making by leveraging a widely deployed network of onboard and ground sensors. This integration of diverse data sources, including manufacturing, maintenance, safety monitoring, equipment inspection, and environmental data, not only requires efficient management and analysis of vast and complex data streams but also highlights the issues of trust and collaboration across organizations, particularly between China State Railway Group Corporation, Original Equipment Manufacturer (OEM), maintenance service providers, and China Academy of Railway Sciences (CARS).

Therefore, we focus on the unified domestic high-speed train PHM big data platform, which has significant practical significance and challenges as a research subject. It not only bears the tasks of integrating, cleansing, and analyzing massive dynamic data but also serves as the core carrier for driving the optimization of maintenance strategies, real-time monitoring, early warning prediction, and health management functions. As shown in Figure 3, the construction and application of the high-speed train PHM platform involve various aspects of data flow and processing, from the collection of raw data to the dissemination of decision support information to end-users (such as EMU sections and depots), and the formation of a feedback loop. Each link in the entire chain requires a high degree of data transparency and trust mechanisms. Based on this platform, we incorporate event logs of the data sharing process for data attribution. Our goal is to explore and implement strategies to resolve conflicts of interest, significantly enhance data credibility, and optimize the flow of data among different stakeholders. This initiative aims to build an efficient, reliable, and future-oriented solution framework for the maintenance management of high-speed trains and the entire railway industry, ensuring clear visibility of data usage and responsibility at every stage, thereby promoting deeper cooperation and technological innovation.



Source: This study.

Figure 2: EMU PHM System Architecture

To precisely simulate the complexity of actual inspection and maintenance scenarios and to validate the potential application of theory in practice, this paper selects two core models from the Predictive Health Management (PHM) system for case analysis: the traction motor fault prediction model and the traction motor fault diagnosis model.

Obtaining Event Logs of the Data Sharing Process

Since there is no unified record of the data sharing process in the EMU maintenance process based on the PHM system, this paper can only be converted into a series of records describing specific events in the data sharing process by sorting out the research minutes and related literature formed by the investigation of Jinan EMU depot, Shanghai EMU depot and Guangzhou EMU depot. In order to further study the distribution of rights and responsibilities in data flow and the optimization of efficiency, this paper designs a series of simulated event logs for data sharing processes based on the established framework. These logs detail the interactions between participants (roles) and task activities (actions), following the specifications listed in Tables 1 and 2. The log system is built not only to track the flow of data between the various participating entities, but also to accurately record the details of task execution at each stage.

Table 1: Actors in Predefined Data Flow Logs

ID	Actor
A1	A EMU depot
A2	A EMU Operation Base
A3	A Original Equipment Manufacturer(OEM)
A4	Chinese Academy of Railway Sciences (CARS)
A5	EMU A001

Source: This study.

Table 2: Actors in Predefined Data Flow Logs

ID	Task	Actor
T1	Calculate the total operating mileage	A1
T2	Comparison shows the predicted result exceeds the threshold	A4
T3	Comparison shows the predicted result falls below the threshold	A4
T4	Detect real-time data of EMU	A5
T5	Discretize real-time related data of EMUs within a specific time period	A4
T6	Enhance the Model	A4
T7	Establish a Fault Diagnosis Model for Traction Motor	A4
T8	Establish a traction motor fault prediction model for prediction	A4
T9	Generate Fault History	A2
T10	Input data into the Fault Diagnosis Model for Traction Motorn	A4
T11	Input data into the traction motor fault prediction model for prediction	A4
T12	Integrate basic operational information of EMUs within a specific time period	A1
T13	Integrate real-time related data of EMUs within a specific time period	A1
T14	Issue a Warning Message	A4
T15	Normalize real-time related data of EMUs within a specific time period	A4
T16	Output Predictive Results	A4
T17	Output the fault diagnosis result as "Faulty"	A4
T18	Output the fault diagnosis result as "No Fault"	A4
T19	Process Warning Information	A1
T20	Push OEM-PHM Model Results	A3
T21	Record Basic Information	A1
T22	Record Train Set Description Information	A1
T23	Retrieve Basic Information	A1
T24	Retrieve basic operational information of EMUs within a specific time period	A4
T25	Retrieve Fault History	A4
T26	Retrieve real-time data of EMU	A4
T27	Retrieve real-time related data of EMUs within a specific time period	A4

Source: This study.

We first considered the extension of the time dimension, ensuring that the new data spans a three-year period, thereby reflecting comprehensiveness in the time series. On this foundation, we constructed more elaborate event chains for each case (caseID), with each event (eventID) corresponding to a specific action. From the capture of real-time EMU data to the deep

involvement of the China Academy of Railway Sciences (CARS), the application of fault prediction models, and the processing of warning information, each step is interconnected, forming a complete and logically coherent process. The repetition of these operations at multiple time points demonstrates a dynamic, cyclical iteration of health management processes, rather than merely an increase in activities. The expanded data situation is illustrated in Figure 5. The dataset comprises a total of 10,146 events. The table below provides a sample of the event log entries.

Table 3: Actors in Predefined Data Flow Logs

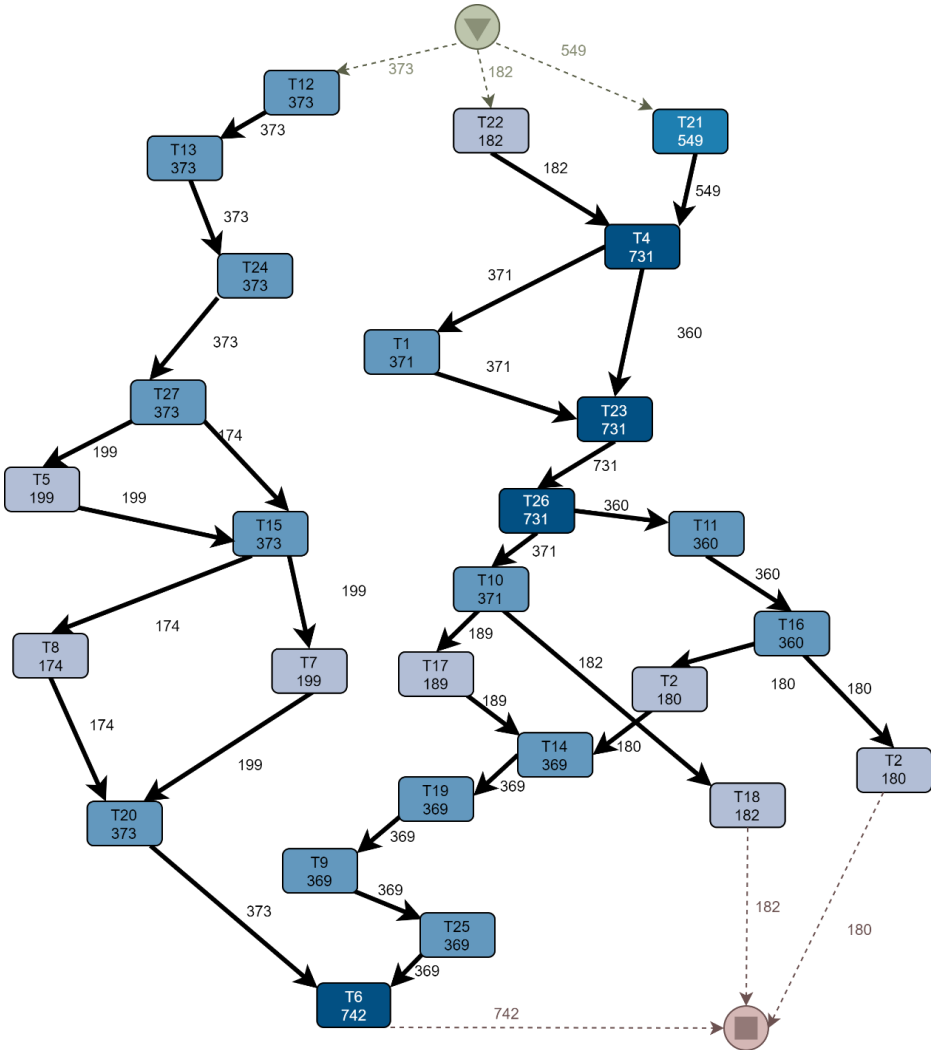
caseID	eventID	actors_id	activity_id	Timestamp	Data source	Stage
1	1001	A1	T12	2023/4/2 11:20:00	EMIS system	raw data
1	1002	A1	T13	2023/4/2 11:20:00	WTDS system	raw data
1	1003	A4	T24	2023/4/2 11:40:00	EMIS system	raw data
1	1004	A4	T27	2023/4/2 11:40:00	WTDS system	raw data
1	1005	A4	T15	2023/4/2 11:42:00	PHM system	Normalized data
1	1006	A4	T8	2023/4/2 12:00:00	PHM system	Data Model
1	1007	A3	T20	2023/4/3 12:11:00	A_OEM-PHM system	Model Computed Data
1	1008	A4	T6	2023/4/3 12:11:00	PHM system	Data Model

Source: This study.

Process Mining

Workflow Analysis- Task Perspective

In this study, to gain an in-depth understanding of the events and workflow structures within the simulated data flow logs, we first employed the process mining tool Disco for visual analysis. The visualization capabilities of Disco allowed us to observe the execution paths of workflows and the interrelationships between activities in an intuitive manner. The generated process models not only displayed the high-level structure of the workflow but also revealed detailed flow paths within the event logs. Through this intuitive visualization method, we were able to more clearly identify various stages of the process, including the sequence of activity execution, potential branches, and loops. Moreover, the abstract log data was effectively transformed into graphical process models by Disco, which not only enhanced the readability of the data but also facilitated further analysis. In conjunction with Figure 6, we found that the visualization results were consistent with the number of events in the data, validating that Disco provides accurate visual representations while maintaining data integrity.



Source: This study(obtained using ProM)..

Figure 6: Process mining using the fuzzy mining algorithm tool in Disco

We integrated four predefined cases to conduct a joint analysis of the actors and tasks within the process. The specific analysis is as follows. This analysis helps us understand how different actors collaborate in the process of traction motor fault prediction and diagnosis for electric multiple units (EMU), as well as how various tasks are interconnected and executed.

Case 1: The construction of a predictive model for traction motor faults in an EMU begins with the collection of operational information from the EMIS system and real-time data from the WTDS system by an EMU depot (A1). These data are raw and serve as the starting point (T12 and T13). Subsequently, CARS(A4) retrieves both basic (T24) and real-time (T27) data from the same sources. The real-time data is normalized within the PHM system (T15), leading to the construction of the predictive model (T8). After the model is built (T8), the original equipment manufacturer (A3) receives the prediction results through system push notifications (T20), enabling the model's application. Finally, CARS (A4) optimizes the model (T6), completing a full process from data to an optimized model that supports operational and maintenance decisions.

Case 2: The data processing flow focuses on the construction of a diagnostic model for traction motor faults in an EMU, involving several core stages. The EMU depot (A1) first integrates basic operational and real-time data (T12 and T13), providing the raw data source. CARS (A4) then extracts and processes this data (T24 and T27), performing discretization (T5) and normalization (T15) to optimize the model's input conditions. After the model is built (T7), the original equipment manufacturer (A3) receives the prediction results through system push notifications (T20). CARS (A4) optimizes its model based on the OEM-PHM model's results (T6), completing a full process from data to an optimized model that supports operational and maintenance decisions.

Case 3: The data processing flow starts with the EMU depot (A1) recording basic information in the EMIS system (T21) and real-time monitoring of EMU A001 by the WTDS system (T4). CARS (A4) then extracts basic operational data from the EMIS system (T23) and real-time monitoring data from the WTDS system (T26). These data are input into the PHM system's predictive model for traction motor faults (T11) for calculation, resulting in prediction results (T16), which are compared

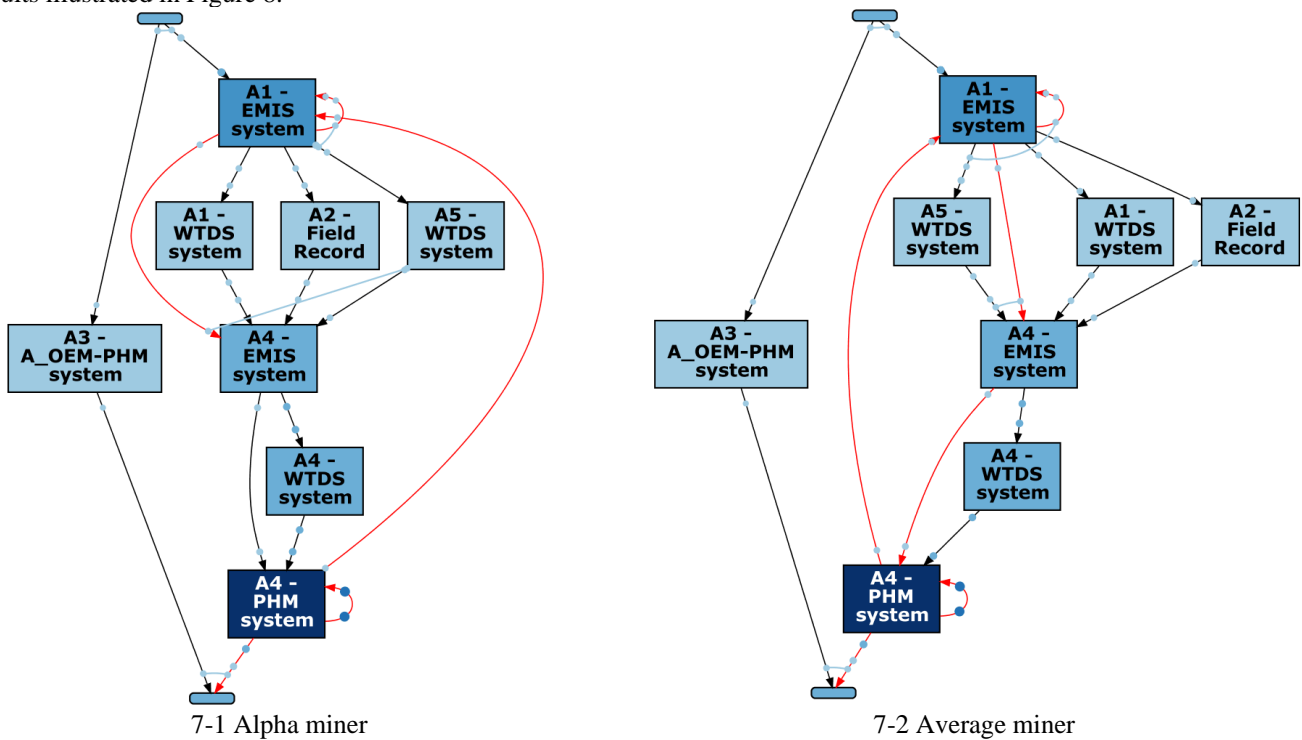
against a threshold (T2). If the prediction exceeds the threshold, CARS (A4) issues a warning (T14), and the EMU depot (A1) processes this warning (T19). The EMU Operation Base generates a fault handling record (T9), and CARS (A4) queries the fault handling record from the EMIS system (T25) and optimizes the model (T6), forming a closed-loop management process from data collection to decision support, effectively serving the operational maintenance of A Railway Bureau's EMUs. Case 6 is similar to Case 3, with the difference being that the model indicates a normal operational status below the preset threshold (T3), and the case concludes.

Case 4: The data processing chain begins with the EMU depot (A1) recording basic operational information in the EMIS system (T21), followed by the real-time data capture of EMU A001 using the WTDS system (T4). The EMU depot (A1) calculates the total operational mileage (T1), while CARS (A4) extracts basic operational data from the EMIS system (T23) and real-time monitoring data from the WTDS system (T26). These data are then input into the PHM system's diagnostic model for traction motor faults for analysis (T10). After the diagnostic model runs, it clearly identifies "fault present" (T17), triggering CARS (A4) to issue an immediate warning (T14). The EMU depot (A1) responds to the received warning (T19). The EMU Operation Base processes the fault, creating a fault history record (T9), laying the foundation for subsequent analysis. CARS retrieves the fault handling record from the EMIS system (T25) and iteratively upgrades the model based on the new situation (T6) to enhance the precision of predictions and diagnoses. Case 5 is similar to Case 4, with the difference being that the PHM system concludes "no fault" (T18), leading to no subsequent processing.

Workflow Analysis - Organizational Perspective

After conducting mining and analysis on business processes using Disco, event logs can be converted into XES and MXML formats. We then import the converted data files into ProM and use the iDHM plugin for further analysis.

Based on an organizational perspective, we utilized four mining algorithms within the iDHM plugin for analysis, with the results illustrated in Figure 8.



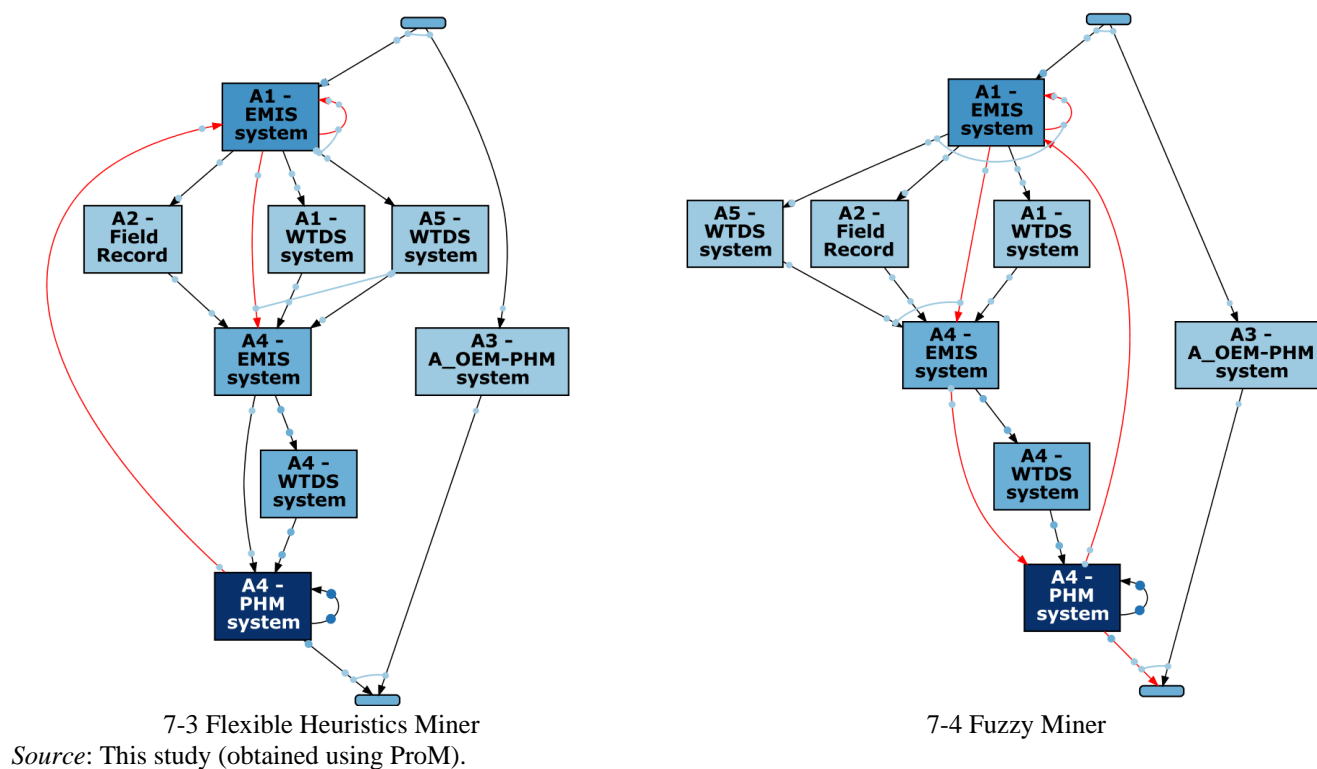


Figure 7: Process mining using the iDHM tool in ProM

Following the steps outlined in Section 3.3, we proceed to analyze Figure 8-1.

First, we establish a causal chain between activities by identifying dependencies. In Figure 8-1, Actor A1-EMIS System serves as the starting point for data collection, while A2 and A5 are associated with the WTDS System, indicating their potential role in real-time data collection and monitoring. These activities are connected by arrows, indicating the sequence of execution and dependencies between activities. For instance, activities performed by A1 in the EMIS System may depend on activities performed by A2 in the WTDS Field, and activities by A5 may depend on data collected by A1 and A2.

Next, we identify split and join points in the C-Net, which are represented as points in the diagram through I/O bindings. These points help us understand how data flows and transforms between different actors and systems. For example, the connection between A4 and both the WTDS and PHM systems may indicate data transfer between these two systems, where individual points may represent XOR splits based on specific conditions, and connected points represent AND splits for parallel processing.

Finally, we reveal decision criteria that determine when specific output bindings are activated within the data sharing process. In the diagram, the interaction between A4 and the OEM-PHM System involves the formulation of decision rules, which may be based on data retrieved from the EMIS and WTDS Systems.

Through a comprehensive analysis of the four diagrams, Figure 8-1 indicates that Actor A1 is directly connected to the EMIS System, while A2 and A5 are associated with the WTDS System, suggesting that data is directly entered into the recording system after being collected on-site. Figures 8-3 and 8-4 are similar, but they particularly emphasize the connection between Actor A4 and the PHM System, implying a critical decision or analysis phase. In contrast, Figure 8-4 shows the direct connections between Actors A1 and A2 with the WTDS Field and WTDS System, indicating a process that is more focused on data collection and initial recording. Figure 8-2 presents a more integrated view, with Actors A1, A2, A3, and A4 connected to multiple systems, demonstrating a complex process involving multiple actors and interactions across various systems. Moreover, the connection between Actor A4 and the OEM-PHM System in Figure 8-2 may suggest a broader business decision-making process involving external manufacturer systems. Overall, while these four diagrams share some commonalities, each highlights different data flow and decision points, revealing unique business processes and actor roles.

Social Network Analysis

Using the Social Network Analysis plugin in ProM, there are various metrics available to describe the network structure within the input event logs.

Density indicates the average degree of connections between organizations. The network has 8 nodes and 14 links. The density of the network is approximately 0.0714. In a high-density network, organizations interact more frequently, which should lead to more effective coordination. A high density value indicates a strong degree of cooperation and coordination between departments. This should increase the likelihood of social control over the network. If the social network is dense or cohesive,

the score is 1 (or 100%), meaning that all its members are directly connected, while a score of 0 indicates that they are completely disconnected. Considering the density of the entire network, the members of the EMU maintenance group have established 7.14% of the possible connections between them.

Task-Actor Matrix

Utilizing the 'Task-Actor Matrix Mining' feature in ProM to extract the relationship matrix between actors and tasks. The two-dimensional relationship matrix differs from a one-dimensional 'person-person' matrix, as it does not merely represent the connections between workers. Instead, it leverages the relationships between individuals and activities to simultaneously display the connections between tasks and the connections between workers. The matrix obtained using the 'Analysis-Originator by Task Matrix' plugin is shown in Figure 9.

originator	T1	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T2	T20	T21	T22	T23	T24	T25	T26	T27	T3	T4	T5	T6	T7	T8	T9
A1 - EMIS system	371	0	0	373	0	0	0	0	0	0	369	0	0	549	182	189	0	0	0	0	0	0	0	0	0	0	0
A1 - WTDS system	0	0	0	0	373	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2 - Field Record	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	369
A3 - A_OEM-PHM system	0	0	0	0	0	0	0	0	0	0	0	0	373	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4 - EMIS system	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	542	373	369	189	0	0	0	0	0	0	0	0
A4 - PHM system	0	371	360	0	0	369	373	360	189	182	0	180	0	0	0	0	0	0	0	0	180	0	199	174	174	0	0
A4 - WTDS system	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	542	373	0	0	0	0	0	0	0
A5 - WTDS system	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	731	0	0	0	0

Source: This study (obtained using ProM).

Figure 8: EMU Maintenance Case Activity-Actor Matrix Mining Results

In this matrix diagram, we can observe eight distinct organizational entities and their systems (A1-EMIS, A1-WTDS, A2-Field Record, A3-A_OEM-PHM, A4-EMIS, A4-PHM, A4-WTDS, A5-WTDS) participating in 27 different activities (T1 to T27). Through the element values in the matrix, we can understand the extent to which each actor participates in a particular activity. Each cell in the matrix represents the frequency with which a specific organization executes a particular activity, with the depth of color indicating the frequency level, and white cells indicating a frequency of zero.

From the diagram, it is evident that, for instance, A1 (A EMU depot) - EMIS system has a high frequency of execution for activity T21 (Record Basic Information), as indicated by the darker color of the corresponding cell, highlighting the importance of data collection and extraction within the overall activity set. Similarly, A4 (A EMU depot) - EMIS system has a high frequency of execution for activity T23 (Retrieve Basic Information), indicating the significance of data collection and retrieval in the workflow. The high frequency of execution for A4-PHM system in activities T10 (Input data into the Fault Diagnosis Model for Traction Motor) and T15 (Normalize real-time related data of EMUs within a specific time period) also underscores the critical role of data importation and processing in the workflow. On the other hand, A1 (A EMU depot) - WTDS system has a low frequency of execution or no execution across all activities, as it primarily serves to provide data.

Node Centrality Analysis

In the field of social network analysis, Ucinet is a powerful software tool that offers researchers a wealth of functionalities to parse and understand complex social relationship networks. In our research project, we utilized Ucinet to distill key information from multidimensional data, transforming a two-dimensional activity-actor matrix into a one-dimensional actor matrix.

To delve deeper into the status and influence of actors within the network, we needed to focus more closely on the actors themselves rather than their specific activities. Therefore, we converted the two-dimensional matrix into a one-dimensional actor matrix, a process that essentially projects the original matrix to obtain the connectivity of each actor within the entire network, i.e., the frequency and intensity of their interactions with others.

Degree centrality is a measure of the importance of a network node (in this case, an actor), based on the number of direct connections the node has. By calculating the degree centrality of each actor, we can identify which actors occupy central positions in the network and which individuals play key roles in information dissemination, resource allocation, and other aspects. This analytical step is crucial for understanding the workflow, power distribution, and communication patterns within an organization. It also provides valuable insights for optimizing team collaboration and enhancing overall efficiency.

Further conducting node centrality analysis on the EMU maintenance organization, in this subsection, degree centrality is selected as the indicator to measure node centrality. The degree centrality data table for the activity dimension is shown in the following Table 4.

Table 4: Node degree of the social network

	Degree	NrmDegree	Share
A4-EMIS system	2	28.571	0.5
A4-WTDS system	1	14.286	0.25
A1-EMIS system	1	14.286	0.25
A1-WTDS system	0	0	0
A3-A_OEM-PHM system	0	0	0

A4-PHM system	0	0	0
A2-Field Record	0	0	0
A5-WTDS system	0	0	0

Source: This study (obtained using UCINET 6.0 software).

Degree Centrality is a metric used to measure the importance of a node within a network, based on the number of connections a node has. The A4 - EMIS system exhibits the highest degree centrality with a value of 2, indicating that it has the most direct connections within the network and occupies a central position. Its normalized degree centrality (NrmDegree) is 28.571%, the highest among all nodes, meaning it participates in approximately 28.571% of the network activities. Additionally, its Share value is 0.500, indicating that it is involved in half of the total activities. Both the A4 - WTDS system and the A1 - EMIS system have high degree centrality values of 1, and their NrmDegree is the same at 14.286%, suggesting that these two systems have comparable activity levels within the network, with each participating in approximately 14.286% of the activities. The A1 - WTDS system, A3 - A_OEM-PHM system, A4 - PHM system, A2 - Field Record, and A5 - WTDS system all have a degree centrality of 0, indicating that these organizations and systems have no direct connections or activity participation within the network. Their roles in the current network structure may be peripheral or limited.

Through this analysis, we can understand the level of participation and importance of different systems or members within the high-speed train maintenance organization network. The A4 - EMIS system is clearly a key node in the network, demonstrating that in the high-speed train maintenance organization, the Chinese Academy of Railway Sciences (CARS) utilizes EMIS system data as a critical and central node.

Table 5: Descriptive Statistics of Node Degree in Social Networks

	Degree	NrmDegree	Share
Mean	0.5	7.143	0.125
Std Dev	0.707	10.102	0.177
Sum	4	57.143	1
Variance	0.5	102.041	0.031
SSQ	6	1224.49	0.375
MCSSQ	4	816.327	0.25
Euc Norm	2.449	34.993	0.612
Minimum	0	0	0
Maximum	2	28.571	0.5

Source: This study (obtained using UCINET 6.0 software).

The descriptive statistics presented in Table 5 offer key metrics for node degree centrality within the network. The average degree centrality is 0.500, indicating that nodes on average participate in half of the connections. However, the inconsistency in connections among nodes is highlighted by a standard deviation of 0.707 and a variance of 0.500, suggesting that some nodes have significantly more connections than others. The sum of degree centrality for all nodes is 4.000, and their cumulative proportion reaches 1.000, indicating that these connections account for the entirety of the network's connections. Additionally, the minimum degree centrality is 0.000, indicating that at least one node has no connections, while the maximum degree centrality reaches 2.000, suggesting that one node has twice the average number of connections. The Euclidean norm of 2.449 provides a measure of node importance in a multidimensional space, and the sum of squares of 6.000, compared to the average sum of squares of 4.000, further emphasizes the unequal distribution of connections among nodes. These data collectively suggest the presence of significant central nodes within the network, alongside some nodes that are more isolated, a distribution that may impact the network's stability and efficiency.

Structure Hole

In this section, we employ structural hole theory to analyze potential connections and gaps in information flow within the network. Based on the data from Table 6, we can observe the connection strengths between different systems. The A1-EMIS system exhibits a high degree of connectivity with the A4-EMIS system (0.56), once again indicating frequent information exchange and close collaboration between the EMU depot and the China Academy of Railway Sciences (CARS) in terms of EMIS system data.

On the other hand, the connectivity between the A1-WTDS system and the A2-Field Record, A3-A_OEM-PHM system, etc., is low or non-existent, indicating structural holes in the data sharing process among various organizations within the WTDS system, Field Record, and _OEM-PHM system. These structural holes lead to inefficient information transmission and increased coordination costs.

In particular, the A4-PHM system has a connectivity degree of 1, indicating its critical role as a liaison in the network and as a central node for information circulation. The A5-WTDS system also has a connectivity degree of 1, signifying a high degree of autonomy and independence within the network, but it also suggests that it has fewer direct connections with other systems.

Table 6: Inter-Organizational Dyadic Constraints and System Interactions

	A1-EMIS system	A1-WTDS system	A2-Field Record	A3-A_OEM-PHM system	A4-EMIS system	A4-PHM system	A4-WTDS system	A5-WTDS system
A1-EMIS system	0.56	0	0	0	0.56	0	0	0
A1-WTDS system	0	1	0	0	0	0	0	0
A2-Field Record	0	0	1	0	0	0	0	0
A3-A_OEM-PHM system	0	0	0	1	0	0	0	0
A4-EMIS system	0.2	0	0	0	0.44	0	0.2	0
A4-PHM system	0	0	0	0	0	1	0	0
A4-WTDS system	0	0	0	0	0.56	0	0.56	0
A5-WTDS system	0	0	0	0	0	0	0	1

Source: This study (obtained using UCINET 6.0 software).

Table 7 presents data on structural hole-related indicators in social network analysis across different systems. These indicators help us understand the position, influence, and interaction patterns of each system within the network. The Degree Centrality indicator reflects the number of direct connections each system has with other systems. The A4-EMIS system leads with a score of 3, indicating that it has the most direct connections with other systems in the network and plays a key role in information sharing and coordination. In contrast, the A1-WTDS system, A2-Field Record, A3-A_OEM-PHM system, and A5-WTDS system all have the same Degree Centrality score of 1, suggesting that they have fewer direct connections within the network. The Efficiency Size indicator shows that all systems have a score of 1, indicating that each system has a similar level of efficiency and influence within the network. In the Efficiency indicator, the A1-EMIS system scores the lowest (0.5), suggesting that this system has lower efficiency in task completion and information processing compared to other systems in the network. The Constraint indicator shows that all systems, except for the A4-EMIS system, have a score of 1, indicating that these systems face higher constraints within the network, with limited autonomy and flexibility. In the Hierarchy indicator, the A4-EMIS system scores the lowest (0.074), suggesting that it occupies a lower position in the network's hierarchical structure, while other systems have a more equal status in the hierarchy. Lastly, the Ego Betweenness indicator shows that only the A4-PHM system scores 1, which may indicate that the A4-PHM system exhibits a strong ego-centricity within the network, while other systems have a more average performance in this regard.

Table 7: Structural Hole Measures of Organizational Actors and Their Systems

	Degree	EffSize	Efficie	Constra	Hierarc	Ego Bet
A1-EMIS system	2	1	0.5	1.125	0	0
A1-WTDS system	1	1	1	1	1	0
A2-Field Record	1	1	1	1	1	0
A3-A_OEM-PHM system	1	1	1	1	1	0
A4-EMIS system	3	1.667	0.556	0.84	0.074	1
A4-PHM system	1	1	1	1	1	0
A4-WTDS system	2	1	0.5	1.125	0	0
A5-WTDS system	1	1	1	1	1	0

Source: This study (obtained using UCINET 6.0 software).

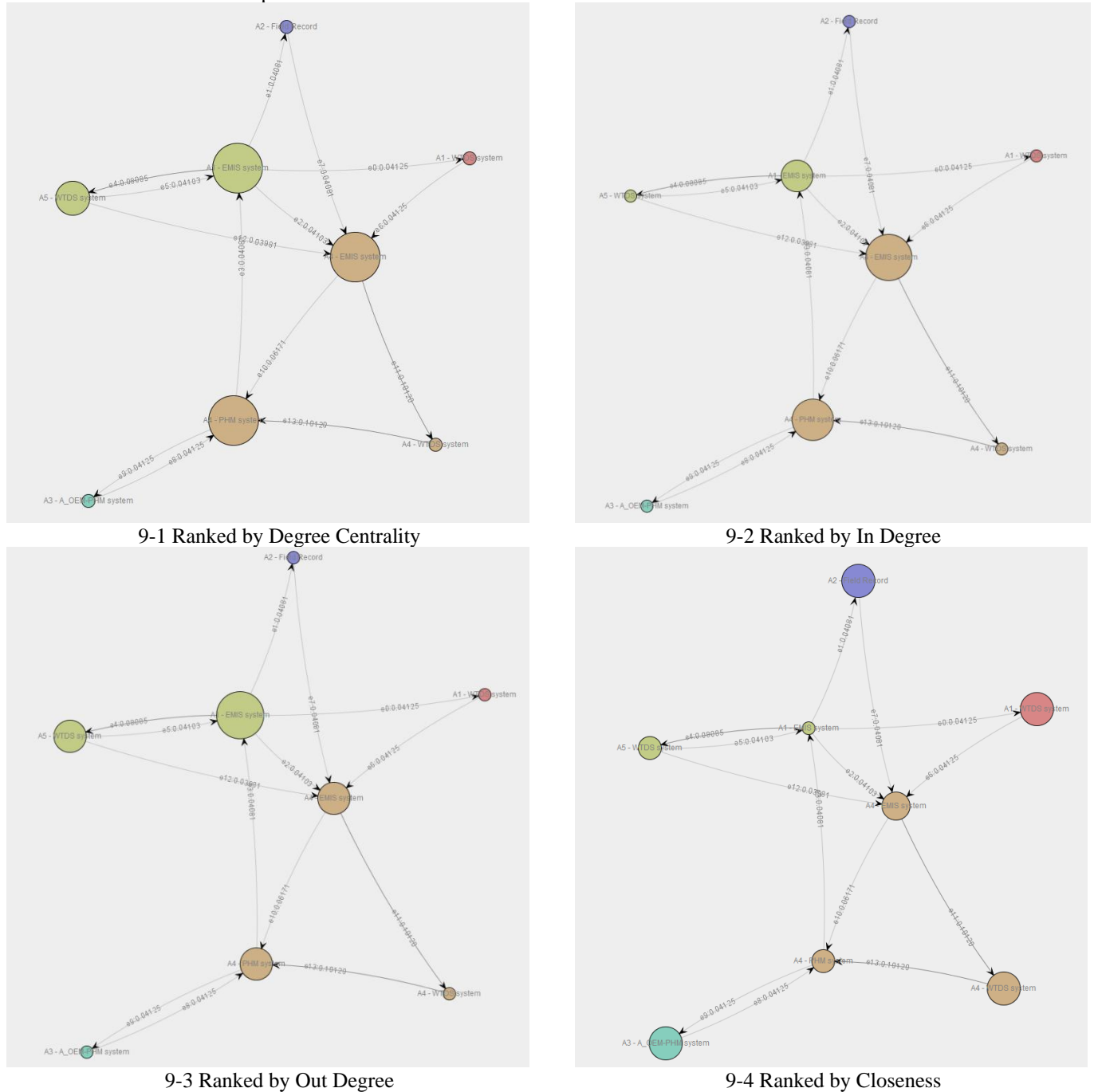
Handover of Work Metric

This section focuses on analyzing the handover of work, establishing the causal relationships between drill activities in the event logs, and constructing a social network graph that details the work transfer relationships among eight participants. This weighted graph not only illustrates the structure of interactions between participants but also reflects their levels of participation and task dependencies in the high-speed train maintenance process. Nodes in the graph represent different organizational units and systems used, while the connections between nodes, along with their weights and coefficients, quantify the relationships among these organizational units and systems.

Through a ranking analysis of four key indicators—Degree Centrality, In-Degree Centrality, Out-Degree Centrality, and Closeness Centrality—we gain profound insights into the network structure. Notably, the A4-EMIS system and A4-PHM system score significantly high in Degree Centrality and In-Degree Centrality, highlighting the central role of China Academy of Railway Sciences Group Co., Ltd. in the entire maintenance process. These systems are not only hubs for information and resource distribution but also play a crucial role in fault decision-making and coordination.

Concurrently, the high Out-Degree Centrality of the A1-EMIS system indicates that the EMU depot is a primary source of data and information output during the maintenance process. This characteristic underscores the importance of the depot in data collection and distribution, playing a key role in the transparency and control of the entire workflow.

In the analysis of Closeness Centrality, we observe that nodes with higher scores can quickly establish connections with other nodes, indicating their central positions within the network. However, the relatively low Closeness Centrality score of the A1-EMIS system suggests that it may be located at the periphery of the network, implying potential delays or obstacles in information transmission and process coordination.



Source: This study(obtained using ProM)..

Figure 9: Mining result for handover of work social network

Based on the data from Table 8, we can observe that the A1-EMIS system is associated with a large number of nodes, but its connections with the A1-WTDS system, A2-Field Record, and A4-EMIS system are relatively weak. This suggests that the EMU depot has limited direct interactions with the data sources of these organizations, indicating a low level of engagement in data sharing and work handover.

The A4-EMIS system has strong connections with both the A4-WTDS system and A4-PHIS system, with weight values of 0.06171 and 0.1012, respectively. This close relationship signifies that CARS not only facilitates efficient information flow

within its internal systems but also serves as a bridge connecting the EMU depot with other key stakeholders. It also demonstrates the frequency of data sharing and work handover among the EMIS, WTDS, and PHM systems during the EMU maintenance process.

Table 8: Edge weight values between nodes in the EMU maintenance network.

	A1-EMIS system	A1-WTDS system	A2-Field Record	A3-A_OEM-PHM system	A4-EMIS system	A4-PHM system	A4-WTDS system	A5-WTDS system
A1-EMIS system		0.04125	0.04081		0.04103			0.08085
A1-WTDS system					0.04125			
A2-Field Record					0.04081			
A3-A_OEM-PHM system						0.04125		
A4-EMIS system				0.04125		0.06171	0.1012	
A4-PHM system	0.04081							
A4-WTDS system						0.1012		
A5-WTDS system	0.04103				0.03981			

Source: This study

Combining the above discussions, we can draw three conclusions. (1) CARS occupies a central position in the entire network. Although the specific EMU A001 represented by A5 does not have a direct weight value in the table, it is the central object of the maintenance process, with all activities and data flows being directly or indirectly related to it. (2) The connection between the EMU Operation Base represented by A2 and the Original Equipment Manufacturer represented by A3 is not directly shown in the table, but their roles within the network are significant and cannot be overlooked. (3) The specific EMU A001 represented by A5, although lacking a direct weight value in the table, is the central object of the maintenance process, with all activities and data flows being directly or indirectly related to it

CONCLUSION

This paper primarily analyzes event logs of data sharing processes, yielding new insights, particularly focusing on the complex scenarios involving multiple stakeholders in a predictive health management (PHM) based high-speed train maintenance case. The study employs process mining and social network analysis techniques, using interactive data-aware heuristic miner (iDHM) to mine and visualize activity dependencies in data sharing processes, identifying key split and merge points, thereby revealing patterns and trends within the data sharing process. Additionally, social network analysis reveals the social structures and dynamics of data sharing processes, as well as the flow of data among different stakeholders.

The findings of this paper are significant for understanding the intrinsic mechanisms of data sharing, promoting data compliance, efficient circulation, and utilization, and enhancing the effectiveness of collaborative networks. Through this research, we have identified potential barriers and opportunities within the data sharing process, providing an empirical basis for the development of data management strategies.

The limitations of this study include, firstly, the fact that there is currently no specialized system or platform to record the actual data sharing processes in high-speed train maintenance. Although the method of extending simulated event logs used in this paper helps in understanding and exploring potential patterns of data sharing, the accuracy and comprehensiveness of the results are limited by the simulated conditions set. Future research should focus on developing or utilizing more advanced data collection tools to obtain more detailed and accurate records of data sharing processes, further deepening our understanding of data flow patterns in multi-stakeholder environments. Secondly, while iDHM can handle complex event logs, the diversity and incompleteness of data may still affect the accuracy of the model. With appropriate methodological adjustments, the research conclusions can be generalized to other types of data sharing scenarios. Therefore, future research could consider a broader range of data sources and more complex analytical techniques to comprehensively understand the data sharing process.

The findings of this paper will inspire future experimental designs, particularly in the areas of data governance, compliance, and stakeholder engagement. The results indicate that optimizing the data sharing process requires not only technological innovation but also a deep understanding of stakeholder behavior and social network structures. Therefore, future experiments should focus on developing data sharing frameworks that can adapt to the needs of diverse stakeholders and assess the impact of different strategies on data sharing efficiency and trust.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China [grant number 52172311, U2268202].

REFERENCES

- Alshwayat, D., MacVaugh, J. A., & Akbar, H. (2021). A multi-level perspective on trust, collaboration and knowledge sharing cultures in a highly formalized organization. *Journal of Knowledge Management*, 25(9), 2220–2244. <https://doi.org/10.1108/JKM-05-2020-0354>

- Barua, A., Ravindran, S., & Whinston, A. B. (2007). Enabling information sharing within organizations. *Information Technology and Management*, 8(1), 31–45. <https://doi.org/10.1007/s10799-006-0001-7>
- Berti, A., & van der Aalst, W. (2020). Extracting Multiple Viewpoint Models from Relational Databases. In P. Ceravolo, M. van Keulen, & M. T. Gómez-López (Eds.), *Data-Driven Process Discovery and Analysis* (pp. 24–51). Springer International Publishing. https://doi.org/10.1007/978-3-030-46633-6_2
- Boh, W. F., & Yellin, D. (2006). Using Enterprise Architecture Standards in Managing Information Technology. *Journal of Management Information Systems*, 23(3), 163–207. <https://doi.org/10.2753/MIS0742-1222230307>
- Diba, K., Batoulis, K., Weidlich, M., & Weske, M. (2019). Extraction, correlation, and abstraction of event data for process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10. <https://doi.org/10.1002/widm.1346>
- Dorofeev, A., Kurganov, V., Gryaznov, M., Egorov, R., & Filippova, N. (2023). Process Mining Platform of a Flexible and Sustainable Transportation Company. In K. S. Zokirjon ugli, A. Muratov, & S. Ignateva (Eds.), *Fundamental and Applied Scientific Research in the Development of Agriculture in the Far East (AFE-2022)* (pp. 340–347). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36960-5_39
- Duncan, S. E., Zhang, B., Thomason, W., Ellis, M., Meng, N., Stamper, M., Carneiro, R., & Drape, T. (2020). Securing data in life sciences— A plant food (edamame) systems case study. *Frontiers in Sustainability*, 1, 600394. <https://doi.org/10.3389/frsus.2020.600394>
- Durojaiye, A., Fackler, J., McGeorge, N., Webster, K., Kharrazi, H., & Gurses, A. (2022). Examining Diurnal Differences in Multidisciplinary Care Teams at a Pediatric Trauma Center Using Electronic Health Record Data: Social Network Analysis. *Journal of Medical Internet Research*, 24(2), e30351. <https://doi.org/10.2196/30351>
- Fernandez, R. C., Subramaniam, P., & Franklin, M. J. (2020). Data market platforms: Trading data assets to solve data problems. *Proceedings of the VLDB Endowment*, 13(12), 1933–1947. <https://doi.org/10.14778/3407790.3407800>
- Grandison, T., Ganta, S., Braun, U., & Kaufman, J. (2007). Protecting Privacy while Sharing Medical Data Between Regional Healthcare Entities. *Studies in Health Technology and Informatics*, 129, 483–487. <https://doi.org/10.3233/978-1-58603-774-1-483>
- Greene, M., Grabus, S., & Greenberg, J. (2021). DARSI: An Ontology for Facilitating the Development of Data Sharing and Use Agreements. *NASKO*, 1–7. <https://doi.org/10.7152/nasko.v8i1.15873>
- Guentherman, M., Hartman, E., Schlegelmilch, A., Brinck, E. A., & Anderson, C. A. (2021). Building a Management Information System with Inter-Agency Data Sharing to Improve Data-Based Decision Making Across Systems: Experiences From Wisconsin PROMISE. *Rehabilitation Counseling Bulletin*, 65(1), 6–22. <https://doi.org/10.1177/0034355220962182>
- Hazée, S., Vaerenbergh, Y.V., Delcourt, C., & Kabadayi, S. (2020). Service delivery system design for risk management in sharing-based product service systems: A customer-oriented approach. *International Journal of Operations & Production Management*, 40(4), 459–479. <https://doi.org/10.1108/IJOPM-08-2019-0581>
- Herurkar, D., Palacio, S., Anwar, A., Hees, J., & Dengel, A. (2024). Fin-Fed-OD: Federated outlier detection on financial tabular data (arXiv:2404.14933). *arXiv*. <https://doi.org/10.48550/arXiv.2404.14933>
- Laurila-Pant, M., Mäntyniemi, S., Venesjärvi, R., & Lehtikoinen, A. (2019). Incorporating stakeholders' values into environmental decision support: A Bayesian Belief Network approach. *Science of The Total Environment*, 697, 134026. <https://doi.org/10.1016/j.scitotenv.2019.134026>
- Lazega, E., Wasserman, S., & Faust, K. (1995). Social Network Analysis: Methods and Applications. *Revue Française de Sociologie*, 36(4), 781–783. <https://doi.org/10.2307/3322457>
- Mannhardt, F., de Leoni, M., & Reijers, H. A. (2017). Heuristic mining revamped: An interactive, data-aware, and conformance-aware miner: 15th International Conference on Business Process Management (BPM 2017). *Proceedings of the BPM Demo Track and BPM Dissertation Award, Co-Located with 15th International Conference on Business Process Management (BPM 2017), Barcelona, Spain, September 13, 2017*, 1–5.
- Mustroph, H., Winter, K., & Rinderle-Ma, S. (2024). Social Network Mining from Natural Language Text and Event Logs for Compliance Deviation Detection. In M. Sellami, M.-E. Vidal, B. van Dongen, W. Gaaloul, & H. Panetto (Eds.), *Cooperative Information Systems* (pp. 347–365). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-46846-9_19
- Naidu, K., & Chakravarthy, S. (2021). Legal Policies for Sustainable Security and Privacy in Big Data and Information sharing- A managerial perspective. *The Journal of Contemporary Issues in Business and Government*, 27(2), 1556–1571.
- Pan, Y., & Zhang, L. (2021). Automated process discovery from event logs in BIM construction projects. *Automation in Construction*, 127, 103713. <https://doi.org/10.1016/j.autcon.2021.103713>
- Pan, Y., & Zhang, L. (2022). Modeling and analyzing dynamic social networks for behavioral pattern discovery in collaborative design. *Advanced Engineering Informatics*, 54, 101758. <https://doi.org/10.1016/j.aei.2022.101758>
- Park, K., Cho, M., Song, M., Yoo, S., Baek, H., Kim, S., & Kim, K. (2023). Exploring the potential of OMOP common data model for process mining in healthcare. *PLOS ONE*, 18(1), e0279641. <https://doi.org/10.1371/journal.pone.0279641>
- Pasquadibisceglie, V., Appice, A., Castellano, G., & Malerba, D. (2020). Predictive process mining meets computer vision. In D. Fahland, C. Ghidini, J. Becker, & M. Dumas (Eds.), *Business Process Management Forum* (pp. 176–192). Springer International Publishing. https://doi.org/10.1007/978-3-030-58638-6_11

- Plengsaeng, B., Wehn, U., & van der Zaag, P. (2014). Data-sharing bottlenecks in transboundary integrated water resources management: A case study of the Mekong River Commission's procedures for data sharing in the Thai context. *Water International*, 39(7), 933–951. <https://doi.org/10.1080/02508060.2015.981783>
- Prell, C., Hubacek, K., & Reed, M. (2009). Stakeholder Analysis and Social Network Analysis in Natural Resource Management. *Society & Natural Resources*, 22(6), 501–518. <https://doi.org/10.1080/08941920802199202>
- Schmitt, R. H., Kiesel, R., Buschmann, D., Cramer, S., Enslin, C., Fischer, M., Gries, T., Hopmann, Ch., Huebser, L., Janke, T., Kemmerling, M., Müller, K., Pelzer, L., Perau, M., Pourbafrani, M., Samsonov, V., Schlegel, P., Schopen, M., Schuh, G., ... van der Aalst, W. M. P. (2023). Improving Shop Floor-Near Production Management Through Data-Driven Insights. In C. Brecher, G. Schuh, W. van der Aalst, M. Jarke, F. T. Piller, & M. Padberg (Eds.), *Internet of Production: Fundamentals, Applications and Proceedings* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-030-98062-7_16-1
- Sitova, I., Pecerska, J., & Merkurjevs, J. (2023). Customer-Centric Business Process Discovery Using Process Mining Techniques. *2023 IEEE 64th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, 1–5. <https://doi.org/10.1109/ITMS59786.2023.10317777>
- Song, X., Ali, M., Zhang, X., Sun, H., & Wei, F. (2021). Stakeholder coordination analysis in hazardous waste management: A case study in China. *Journal of Material Cycles and Waste Management*, 23(5), 1873–1892. <https://doi.org/10.1007/s10163-021-01258-9>
- Stalla-Bourdillon, S. (2021). A Maturity Spectrum for Data Institutions. *IEEE Security & Privacy*, 19(5), 90–94.. <https://doi.org/10.1109/MSEC.2021.3094985>
- van Zelst, S. J., Mannhardt, F., de Leoni, M., & Koschmider, A. (2021). Event abstraction in process mining: Literature review and taxonomy. *Granular Computing*, 6(3), 719–736. <https://doi.org/10.1007/s41066-020-00226-2>
- Wang, Y., Thangasamy, V. K., Hou, Z., Tiong, R. L. K., & Zhang, L. (2020). Collaborative relationship discovery in BIM project delivery: A social network analysis approach. *Automation in Construction*, 114, 103147. <https://doi.org/10.1016/j.autcon.2020.103147>