Knowledge Representation of Asset Information and Performance in OT Environments

Mukund Bhole
Institute of Computer Engineering
TU Wien
Vienna, Austria
mukund.bhole@tuwien.ac.at

Wolfgang Kastner
Institute of Computer Engineering
TU Wien
Vienna, Austria
wolfgang.kastner@tuwien.ac.at

Thilo Sauter
Institute of Computer Technology, TU Wien
Dep. of Integrated Sensor Systems
Danube University Krems, Austria
thilo.sauter@tuwien.ac.at

***Abstract*—To make a management decision for the Operational Technology (OT) environment, obtaining OT asset information from a plant is essential. Key Performance Indicators (KPIs) can play a crucial role in evaluating the performance of the manufacturing process, minimizing production costs, assessing process effectiveness, and investigating various settings to improve performance, operations, and quality in OT. This paper aims to provide a detailed examination of how Asset Information Mining Sources (AIMSs) support the process for acquiring both static and runtime OT asset information. Once we have obtained accurate OT asset information, we can calculate the KPIs and incorporate them into a knowledge representation using an ontology-based model. This model can assist the management team in making informed decisions and provide a roadmap for exploring further automation possibilities.**

***Index Terms*—Operational Technology, Asset Information Mining, Knowledge Representation**

I. Introduction

The cultural divide between Information Technology (IT) and OT is similar to that between modern and legacy systems. In the implementation of Industry 4.0 (I4.0), the convergence of IT/OT and the utilization of Asset Administration Shell (AAS) will play an essential role. This includes acquiring asset information, representing the information in a proper system model, and operating an actual system through communication protocols in I4.0 (e.g., OPC UA) [1]. However, legacy devices do not support most modern communication standards in I4.0. In 2022, the Intoware company conducted an independent survey among 1,030 UK-based industrial firms, revealing that nearly three-quarters (74%) of manufacturing and engineering companies still rely on legacy systems and communications [2]. As a result, the underlying OT environment currently faces security risks, such as open attack points due to unpatched asset inventory, insufficient asset monitoring methods, and the absence of an asset information inventory with scheduled vulnerability scanning [3]. These unresolved security issues can have a catastrophic effect on human safety in the real world, even leading to incidents resulting in the loss of human lives [4]. Many industries require assistance in managing their assets, especially when they are located in remote or geographically dispersed locations. Furthermore, managing assets using different tools to provide current asset status is necessary. However, most of these tools acquire asset information through static or one-time snapshots, lacking the current runtime status of OT assets. Maintaining runtime asset information is always challenging and requires manual updates. The company management team should have concrete and fine-grained information about each asset to better understand the system or sub-system and resolve any issues. Acquiring this information is essential regardless of the modern or legacy infrastructure capabilities.

KPIs are widely applied in industries to evaluate process performance and product quality by understanding the real-world data entities that are important to them. Literature, such as [5]–[7], provides definitions and case studies on KPIs. These indicators impact economic performance and assess compliance with essential objectives, such as minimizing production costs, improving process effectiveness, and other critical success factors within an industry [8]. The relationship between KPIs and asset information is closely intertwined, as gaining information about assets is a fundamental requirement for developing KPIs. To support this process, AIMSs are necessary. Furthermore, the data retrieved from AIMSs can be incorporated into an ontology-based knowledge representation. Ontology-based representations are shareable and reusable knowledge that can describe the relationships and interconnections in the environment, events, and actions, serving as the foundation for high-quality data. This data is intended to be understandable and processable by both humans and computers across various domains.

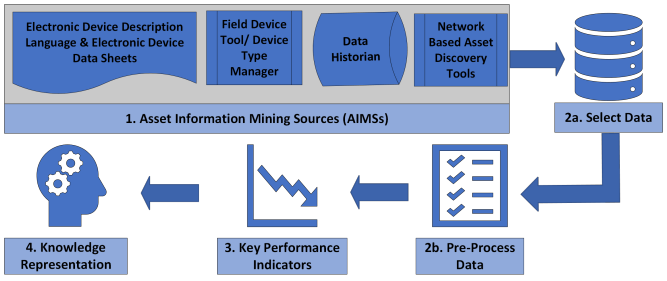


Fig. 1. Knowledge representation with AIMSs

The paper is structured as follows: Section II provides the related work, and the rest of the paper presents the proposed methodology for developing a knowledge representation, as depicted in Fig. 1. The methodology is divided into four steps,

outlined as follows: AIMSs (Section III) presents an overall picture of mining sources. This section focuses on retrieving asset information from the OT environment, regardless of the modern or legacy infrastructure capabilities. Select and pre-process data (Section IV) introduces approaches for selecting and pre-processing relevant information based on industry requirements. KPI (Section V) outlines the process for defining the KPIs. Knowledge Representation (Section VI) discusses the utilization of an ontology-based model for knowledge representation. Finally, Section VII concludes the paper with some remarks and outlines future work.

II. Related Work

The representation of knowledge of assets in Industrial Control Systems (ICS) through an ontology-based modeling approach covers essential aspects of any ICS infrastructure, including construction, function, performance, location, and business [9]. Bunte et al. [10] introduce an approach to a semantic knowledge base that enables smart services to access data from production resources. However, the introduction of smart services requires an IoT interface, which necessitates significant economic investments, particularly in brownfield industries. Hildebrandt et al. [11] demonstrate the semantic modeling of knowledge in the production domain and present a method for building a respective ontology, although they lack the sources to feed the relevant asset information into the ontology. A brief overview of recent research on the semantic and ontological approach used in ICSs to improve automation possibilities, enhance performance, and simplify maintenance can be found in [12]–[14]. Therefore, in this paper, the proposed work presents a promising and reasonable approach to overcome the problems above mentioned.

III. Asset Information Mining Sources

According to [15], Subsection III-A and Subsection III-B consist of static asset information used for configuring the system or sub-system during the initial stage. On the other hand, Subsection III-C and Subsection III-D comprise dynamic asset information utilized during runtime to gather the latest data from the asset. Subsection III-E gives an evaluation strategy of these AIMSs. This section provides a comprehensive expansion of AIMSs in the following manner.

A. Electronic Device Description Language & Electronic Data Sheets

The Electronic Device Description Language (EDDL) is utilized by device manufacturers to simplify device diagnostics, calibration, and device configuration, as analyzed by technicians. It is a standardized programming language that describes the specification features and functions of an OT device. These specifications define the information required by a host to operate the field device. An EDDL file consists of a core set of parameter definitions, user groups, and vendor-specific definitions (device configuration), including manufacturer, device type, and revision. Additionally, it contains device function information, function-related parameters, and special features. Prior to execution, this file is converted into a Device Description (DD) binary file. EDDL also addresses engineering requirements such as ensuring Operating System (OS) independence, supporting multiple OT/ICS protocols, employing standardized testing and validation procedures and tools within the system, and facilitating third-party device testing and registration [16]. The components of an EDDL file are shown in Table I-A.

In ICSs, manufacturers often provide device information through the Electronic Data Sheet (EDS) instead of using EDDL. The EDS serves as an alternative means to describe device information and configuration.

The configuration of an OT device can be accomplished through various methods, including printed datasheets, parameter objects and parameter object stubs, and an EDS. The EDS describes the hardware architecture and software operation characteristics, which electronics technicians can use to define electrical timing or mechanical specifications for the component. An EDS can be a simple ASCII text file that provides information on various aspects of the OT device's capabilities. Thus, an EDS file may contain information such as configuration details, supported connections, network ports, and object details (instances, attributes, and supported services) [20].

B. FDT / Device Type Manager (DTM)

The FDT technology is a standard communication architecture designed to establish an interface between field devices and the control system in OT. FDT is a technology that is compatible with widely used OT communication protocols, including HART, PROFIBUS, Foundation Fieldbus, Modbus, DeviceNet, Interbus, AS-Interface, Profinet, IO-Link, CC-Link, and many more. FDT consists of two components: the frame and DTM. The frame is a software application window that serves as the interface between the DTM and other tools. It initializes the DTM and establishes the connection to the appropriate communication gateway within the network. A DTM is a device-specific software component that defines the device's parameters and capabilities. It also provides comprehensive device access, enabling users to monitor, diagnose, and troubleshoot the system. The DTM has access to topology information, device type information, device status, and device data [17]. The components of an FDT/DTM are shown in Table I-B.

C. Data Historian

Process plants, such as oil fields, power plants, and petrochemical industries, generate a significant amount of data for process control, process monitoring, and manufacturing execution applications. In process control, data is typically stored in three levels (enumerated from bottom to top): (1) Controller Storage, (2) Distributed Control Systems (DCS) Storage, and (3) Plant Data Historian, as depicted in the storage hierarchy of process plants [19]. The data historian can be leveraged to develop an inventory of the OT system. It stores and captures real-time data from various automation

TABLE I
AIMSs RELATED INFORMATION

Sr.	Type	Information
A. Parts of EDDL file based on [16]		
1	File Header	The header consists of information such as copyright, product, device, communication, dictionary, etc.
2	Identification information	It contains the vital information to identify the device such as manufacturer, device type, device revision, Device Description (DD) revision.
3	Data Description	Structural elements such as VARIABLE describe the device parameters and different field data or a device object abstraction. The data types such as floating point, integer, byte, bit, date, time, text and other different types. "RECORD", "VARIABLE_LIST", "BLOCK_A", and "METHOD" organize data elements and describe device capabilities.
4	Communication Description	Communication tasks, fault handling and incident reporting text, related addressing information, message sequence and time conditions.
5	User Interface Description	User Interface Description includes the "MENU", "WINDOW", "TABLE", "CHART" and other visual elements.
B. Asset data from Field Device Tool (FDT) technology based on [17]		
1	Device Type	Online and offline identification, device information.
2	Device Status	According to NE107 [18].
3	Device Parameter	Parameter data range, access information, semantic info, parameter value.
4	I/O Data	Value range, limits, I/O usage
5	Network	Configuration information.
6	Other data	Manuals, technical documentation, certificates, device description, protocol list, ECLASS, etc.
C. Asset data from data historian based on [19]		
1	Analog Readings	Temperature, pressure, flow rates, levels, weights.
2	Digital Readings	Valves, limit switches, motors on/off, discrete level sensors.
3	Product Info	Product ID, batch ID, material ID, raw material lot ID.
4	Quality Info	Process and product limits, custom limits.
5	Alarm Info	Out of limits signals, return to normal signals.
6	Aggregate Data	Average, standard deviation, moving average.
D. Few network-based asset discovery tools		
Tool	**Supported Protocols**	**Description**
Forescout CounterACT[1]	Industrial protocols such as Ethernet/IP	Forescout CounterACT is used in the data collection and inventory management tool. The CounterACT actively collects data from the ICS environment. The ForeScout CounterACT can be deployed on virtual or physical appliances. The large networking sites require multiple physical and virtual, which can be managed on an enterprise level.
PLCScan[2]	Modbus and S7Comm	PLCScan is also a utility tool based on python for the collection of asset information from TCP devices such as Modbus and Siemens S7Comm protocol (it uses NSE script *s7-info*).
ZGrab[3]	HTTP, HTTPS, SSH, Telnet, FTP, SMTP, POP3, IMAP, Modbus, BACnet, S7Comm, and Tridium Fox	ZGrab is an open-source application scanner project which is also being used in the Censys search engine. The application supports different handshakes for protocols in the IT/OT environment.
Security Onion[4]	TCP, UDP, IPv4, IPv6, ICMP, telnet, FTP, HTTP, SMTP, IRC, X11, VNC, etc.	Security Onion is a Linux distribution based on Ubuntu used to audit security in network components. It includes Wireshark, Snort (IDS), and Kibana (for data visualization) tools. It allows for managing the inventory asset passively by capturing the network traffic.
Dragos Platform[5]	50+ ICS protocols	Dragos is a commercial, industrial control system cybersecurity-monitoring platform based on threat behavior and analytics. It is used in the build to provide asset discovery and monitoring.
GrassMarlin[6]	54 industrial protocols	GrassMarlin is an open-source software tool that provides a solution to discover and catalog SCADA and ICS on an Internet Protocol (IP)-based network. It uses various sources to generate the data, such as PCAP files, router, switch configuration files, Content Addressable Memory (CAM) tables, and live network packet captures. If the communication between the local network to another network is tagged suspicious, it triggers an alarm in an alert system.
Nessus [7]	IT/OT/IIoT Protocols	Nessus is a proprietary vulnerability scanner developed by Tenable Network Security. The tool protects IT environments by running vulnerability scans, configuration, compliance checks, malware detection, and Web application scanning; includes assessment capabilities, agent-less, and agent-based scanning; and runs reports and filters data.
Splunk[8]	150+ IT and OT protocols	Splunk is a Security Information and Event Management (SIEM) system that allows the collecting and parsing of logs and system information from multiple points.
OTORIO RAM^2 [9]	IIoT/IT/OT Protocols	OTORIO RAM^2 platform operates in the critical infrastructure to reduce the operating environment risks. It integrates the information from the various operating and security systems to develop the asset management window, which focuses on discovering, analyzing, and monitoring the OT components.
Shodan[10]	HTTP/HTTPS, FTP, SSH, Telnet, SNMP, IMAP, SMTP, SIP, RTSP, OPC UA, etc.	Shodan is a search engine that indexes industrial devices and services connected to the Internet. Shodan scans its target based on public IP address and retrieves the related asset information based on TCP packet ACK.
ISF[11]	Internet and industrial protocols	Industrial Control System Exploitation Framework (ISF) is an active identification software that is an exploitation framework. It is used to identify the OT assets actively. ISF is based on another open-source project called Routersploit - an exploitation framework for embedded devices. ISF can support standard protocols like Modbus and PROFINET.

processes within the plant. Depending on industry-specific configurations and requirements, the data historian may contain plant information, process information, asset information, asset location, asset specifications, and more [21]. Table I-C provides an example of the data that can be recorded from the data historian, which may vary depending on the industry.

D. Network-based Asset Information Discovery

Various network-based techniques and tools used to gain an overview of the real-time OT/ICS environment, which provides knowledge about the internal network and control system architecture and device interactions [22]. The following are some network-based approaches that can be employed in an OT environment:

- **Network Connection Enumeration:** Network connection enumeration involves using tools such as netstat, Nmap, and Wireshark, along with firmware combinations, to discover the communication patterns of devices in the network.
- **Network Sniffing:** Network sniffing is utilized to capture network information and traffic, including important data such as credential transfers to and from devices. Additionally, techniques like Address Resolution Protocol (ARP) and Domain Name System (DNS) poisoning, which exploit vulnerabilities in the network infrastructure, can be employed to capture asset information.
- **Remote System Discovery:** IP addresses, host names, or logical identifiers on a network can be used to discover information about other devices or systems, such as operating systems, firmware or further vendor details.
- **Remote System Information Discovery:** Obtaining information about a remote system and its configuration can provide insights into other asset devices within the network.
- **Wireless Sniffing:** Software-defined radio devices can capture radio frequency signals used in the OT/ICS frequency ranges (3kHz to 300GHz). Radio communication protocols like WirelessHART, ZigBee or WIA-FA can be monitored. If the communication is not encrypted, asset information can be discovered.

Based on these network-based approaches, several existing (automated) tools can be employed to retrieve OT asset information, as discussed in Table I-D.

E. Evaluation of Asset Information Mining Sources

The OT asset information obtained from one or multiple sources, as discussed from Subsection III-A to Subsection III-D, can also affect the time required to retrieve the relevant information. *precision and recall* can be utilized to determine the ranking of AIMSs in order to obtain the most relevant information regarding an asset, especially when selecting the pertinent data for further processing and analysis.

Precision P represents the fraction of relevant OT asset information among all the retrieved information,

$$P = \frac{|\ \{\textit{relevant asset info}\} \cap \{\textit{retrieved asset info}\}\ |}{|\ \{\textit{retrieved asset info}\}\ |}. \quad (1)$$

Recall R is the fraction of relevant OT asset information that was known [23], [24],

$$R = \frac{|\ \{\textit{relevant asset info}\} \cap \{\textit{retrieved asset info}\}\ |}{|\ \{\textit{known asset info}\}\ |}. \quad (2)$$

For example, let us assume that there are 75 known parameters of the asset, and the AIMSs retrieved approximately 30 parameters, out of which only 15 are relevant for drawing meaningful insights. In this scenario, the precision and recall can be calculated as follows: $P = 15/30 = 0.5$ and $R = 15/75 = 0.2$.

A high precision value suggests a significant amount of relevant information can be expected. However, a high recall value indicates that there might be insufficient information to generate comprehensive knowledge.

IV. Select and Pre-Process Data

The process of information mining can vary drastically depending on the infrastructure or availability of AIMSs. Literature [25], [26] provides an understanding of document and passage retrieval extensions, which can be utilized to increase precision and reduce recall by filtering relevant information from irrelevant information. Information retrieval of uncertain data, whether in a textual or non-textual format, involves identifying named entities and relationships within natural language sentences. Techniques such as statistical learning, pattern matching, and natural language processing can be employed for information preprocessing of mining sources. For instance, Amazon Textract, a machine learning service, can be used to automatically extract data from documents, such as datasheets, in key-value pairs [27]. However, handling such a large amount of data can introduce uncertainty into the acquired knowledge. Therefore, selecting specific entities and parameters for calculating KPIs can help in creating a meaningful knowledge representation. Additionally, to demonstrate the methodology, we manually select and preprocess the asset information for this specific part of the process, as outlined in Section V.

V. Key Performance Indicators

KPIs can be used to refine the process model and obtain a comprehensive performance overview. By converting measurable variables, such as minimizing production costs, into feedback control measures, one can analyze the overall architecture and assess the performance impact resulting from the implementation or non-implementation of specific measures.

The sub-process for calculating KPIs is shown in Fig. 2. It involves the following steps: (1) Defining the strategy and goal of KPI measurement, which is essential for monitoring the industrial goals to be achieved in the future. (2) Establishing the scope based on customer expectations, cost, schedule constraints, and management support. (3) Selecting relevant

data based on the scope and evaluation goals (see Section IV for more details). (4) Calculating the KPI based on the selected data.

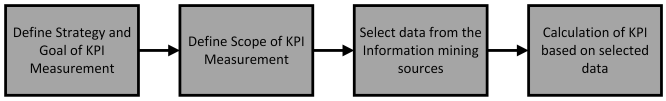


Fig. 2. KPI Calculation Process

For example, to calculate the system's throughput (the total number of items produced per unit of time), we require the number of items, start time, and end time. These parameters can be obtained from the product information stored in the data historian (refer to Subsection III-C). Let's consider an instance where 40 items were produced in five hours by the manufacturing process. The throughput of the process can then be calculated as $40/5 = 8$.

Table II presents a description of a few KPI methodologies in the OT environment, including their scope and the parameters necessary for their calculation, based on [28]. Additionally, [29] highlights the importance of implementing KPIs in the context of Industry 4.0. The concept behind production feedback-controlled KPIs involves utilizing runtime data from Section IV to optimize the system through continuous feedback loops.

VI. Knowledge Representation

A use case for demonstrating the knowledge representation of the data acquired from AIMSs is shown in Fig. 3. It involves a distribution group that is part of a Modular Production System (MPS) 403-1 didactic plant system manufactured by Festo. The distribution group functions as a feeder unit responsible for holding, sorting, and feeding workpieces within the production system. In this scenario, the MES orders the workpiece, the HMI specifies the color of the workpiece, and the RFID writes the data on the finished workpiece. Therefore, the entire process is focused on the distribution group [30].

Building and maintaining a knowledge representation concept requires significant time and effort, particularly in cases where a deeper level of understanding is desired. In such situations, an ontology-based representation becomes essential. Utilizing individual instances of OT components in an ontology allows for knowledge sharing, processing, reuse, and effective communication channels [31]. YAGO [32] is an example of an ontology-based approach that facilitates the development of a clean, human-readable taxonomy while enforcing semantic constraints to ensure logical consistency in the ontology-based representation. This approach also enables further automation possibilities, pipeline execution, and constraint checking. In Section IV and Section V, we extracted network connections, power supply information, relationships, data entities, and KPIs specific to the distribution group using the available AIMSs format. We employed Amazon Textract, as shown in Fig. 4, for this retrieval process.

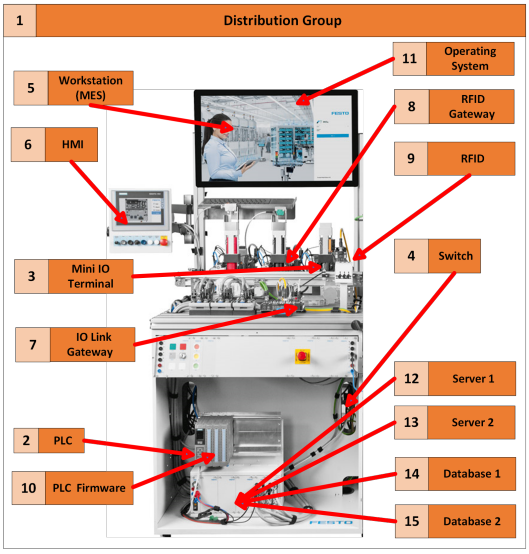


Fig. 3. MPS 403-1 (Distribution Group) [30]

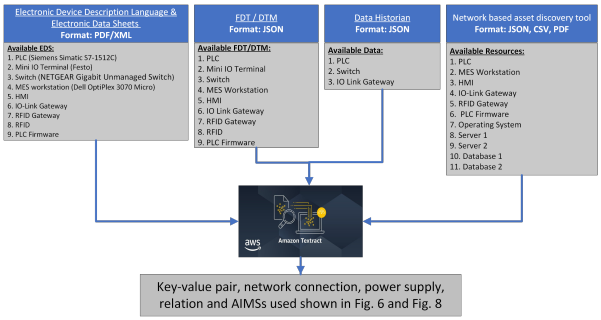


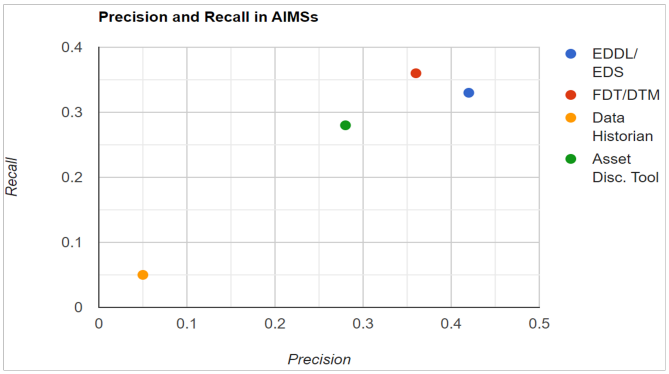
Fig. 4. Selection and Preprocessing of data

The retrieved data is represented in the knowledge representation shown in Fig. 6 to Fig. 8. The Roman numerals correspond to the AIMSs utilized, as follows: I. EDDL and EDS, II. FDT/DTM, III. Data historian, IV. Network-based asset discovery tool (Forescout). In total, **63** relevant parameters were extracted from I, **58** relevant parameters from II, **14** relevant parameters from III, and **48** relevant parameters using Forescout, which had the highest number of parameters among the other tools in IV. The decision to define the relevant parameters was made manually by the experts. According to Subsection III-E, Fig. 5 presents the precision and recall results of the AIMSs for the distribution group use case. The ranking for obtaining the most relevant asset information from the AIMSs is as follows: I, II, IV and III, respectively.

In Fig. 6, the blue block represents the main/subclass in the ontology representation. Hardware and Software are *subclasses* of Component, which is an *abstract subclass* of Group, and Group is a *subclass* of Plant. The orange blocks represent the instantiations of the main/subclasses. Plant is instantiated as a `Production Plant`, Group is instantiated as a `Distribution Group`, and Hardware is instantiated as `PLC`, `Mini`

TABLE II
FEW KEY PERFORMANCE INDICATOR METHODOLOGIES FOR OT ENVIRONMENT BASED ON [28]

Measurement Methodologies	Key Performance Indicator (KPI)	Description
Manufacturing Process Performance	Throughput Rate	Measures the amount of end products produced over a specific amount of time.
	Input Feed Rate	Measures the amount of input materials consumed over a specific amount of time.
	Output Yield	Measures the percentage of the end product produced at the output of the process.
	Equipment Availability	Measures the availability of the critical equipment in the Process Control System.
	Duration of Continuous Operation	Measures the time when the production starts to the next unplanned stop in production.
	Number of Unplanned Stops	Measures the number of times the production stops unexpectedly, including emergency shutdown.
	Capacity Utilization	Measures the percentage of production capacity utilized over a specific amount of time.
	Cycle Time Ratio (CTR)	Measures standard cycle time over real cycle time.
	Rejection Rate	The percentage is obtained by dividing the waste material cost by the total cost of production or by dividing the total quantity of waste materials by the total number of aluminum castings produced.
	Overall Equipment Effectiveness (OEE)	OEE measures how well a manufacturing operation is utilized compared to its full potential during the periods when it is scheduled to run.
	Unit Cost	Measures the production cost of the end products.
Network Performance	Packet Path Delay	Measures the time delay along the path from transmitter to receiver.
	TCP Packet Round Trip Time	Measures the percentage of the quantity of process information packets over all packets transmitted.
	Information Ratio	Measures the amount of time for the source node to receive the acknowledgement of receipt (ACK) from the destination node.
	Packet Error Rate	Measures the rate of packets received with errors over a specific amount of time.
	Proportion of Protocol Type	Measures the percentage of a unique packet protocol type observed over a specific amount of time.
	Network Utilization	Measures the percentage of network capacity utilized over a specific amount of time.
	Packet Loss Rate	Measures the percentage of packets that failed to reach the destination node over a specific amount of time.
Computing Resources Performance	CPU Utilization	Measures the percentage of central processing unit (CPU) utilized over a specific amount of time.
	Memory Utilization	Measures the percentage of memory utilized over a specific amount of time.
	Disk I/O	Measures the mean rate and standard deviation of read and write operations to the server hard drive per software application.
	Network Throughput	Measures the mean rate and standard deviation of packets transmitted and received per software application.



	AIMS Used	Know Asset Information Parameters [V1]	Retrieved Asset Information Parameters (via Amazon Textract) [V2]	Relevant Asset Information Parameters [V3]	Precision [V3/V2]	Recall [V3/V1]
	EDDL/ EDS	187	149	63	0,422818792	0,336898396
	FDT/DTM	157	157	58	0,369426752	0,369426752
	Data Historian	248	248	14	0,056451613	0,056451613
	Network Based Asset Discovery Tool	167	167	48	0,28742515	0,28742515

Fig. 5. Precision and Recall of AIMSs for Distribution Group

`IO terminal`, `Switch`, `MES`, `HMI`, `IO-Link Gateway`, `RFID Gateway`, `RFID`. SOFTWARE is instantiated as `PLC Firmware`, `Operating System`, `Server 1`, `Server 2`, `Database 1`, `Database 2`.

In Fig. 6, the green line indicates the network connection, the red line represents the power supply, and the black line shows the relations between classes or instantiations. In Fig. 7 and Fig. 8, the light green blocks represent the data entities or KPI calculations of the instantiations, and the grey blocks represent the AIMSs used.

VII. CONCLUSION AND FUTURE WORK

In this work, we have demonstrated the knowledge representation methodology using AIMSs in an OT environment. The utilization of AIMSs can play a significant role in both legacy and modern industrial infrastructures. This work highlights the importance and applications of the proposed methodology in gaining knowledge to develop a deeper understanding of systems or subsystems and evaluating their performance using KPIs. Additionally, we have presented the knowledge acquired through an ontology-based model and demonstrated its appli-

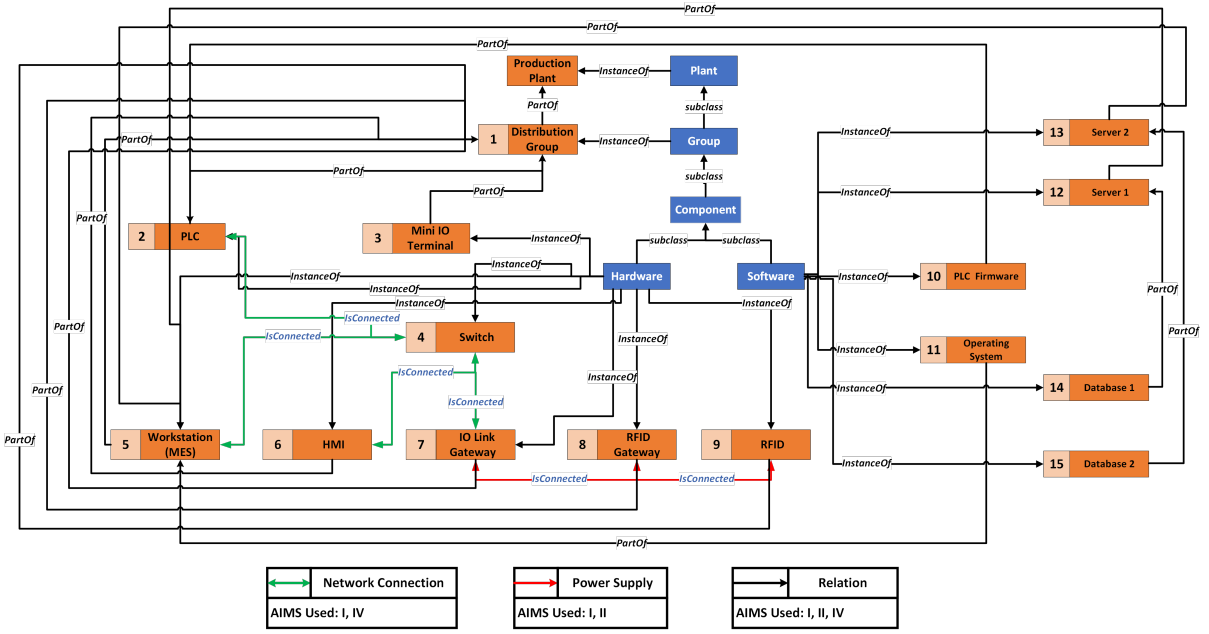


Fig. 6. Knowledge Representation of Distribution Group Based on YAGO [32]

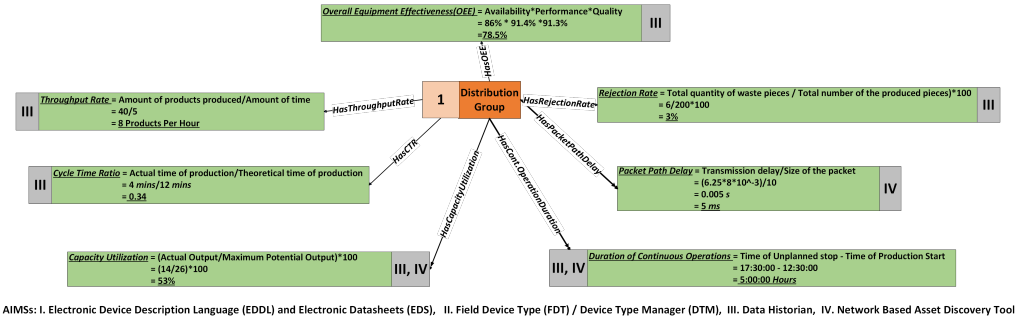


Fig. 7. Knowledge Representation Distribution Group KPIs Based on YAGO [32]

cation to the distribution group of an MPS [30]. The acquired knowledge is crucial for identifying and implementing standard best practices in information architecture, governance, and asset management. Obtaining system information at the OT level for the distribution group also benefits brownfield systems by identifying weak points that can be addressed by different management teams within the industry.

In the future, we expect to further apply the proposed methodology to track constantly evolving system architectures with changing requirements. This includes automating integrated safety and security assessments of the system and experimenting with system configurations to maximize production while minimizing costs. The automation of KPI calculations can accelerate production by providing accurate and up-to-date information, resulting in significant savings of resources and manual work. These applications can be deployed in multi-domain industries, considering the specific needs of each domain.

ACKNOWLEDGMENT

This paper was supported by TÜV AUSTRIA #SafeSecLab Research Lab for Safety and Security in Industry, a research cooperation between TU Wien and TÜV AUSTRIA.

REFERENCES

[1] A. M. Hosseini, T. Sauter, and W. Kastner, "A safety and security reference architecture for asset administration shell design," in *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*, 2022, pp. 1–8.

[2] "CDA MAGAZINE - Connecting the old with the new," https://tinyurl.com/722ftz77.

[3] S. Mansfield-Devine, "The state of operational technology security," *Network Security*, vol. 2019, no. 10, pp. 9–13, 2019.

[4] "A patient has died after ransomware hackers hit a German hospital," https://tinyurl.com/5n9ab624.

[5] P. J. Ballard, "Measuring Performance Excellence: Key Performance Indicators for Institutions accepted into the Academic Quality Improvement Program (AQIP)," 2013.

[6] D. A. Bishop, "Key Performance Indicators: Ideation to Creation," *IEEE Engineering Management Review*, vol. 46, pp. 13–15, 2018.

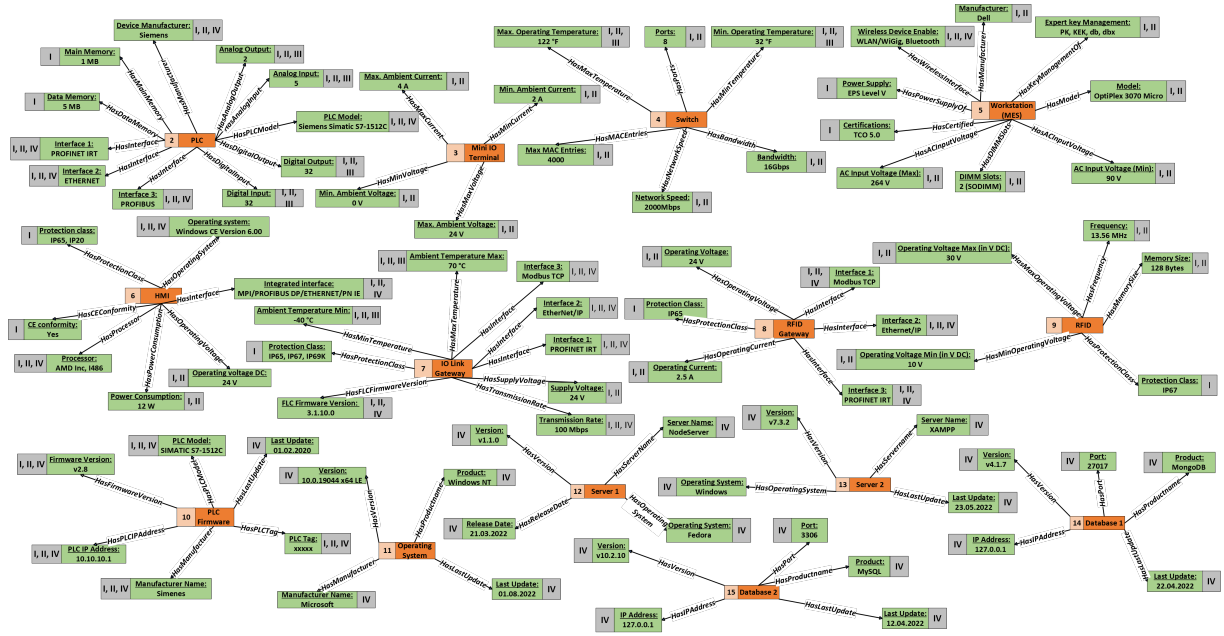


Fig. 8. Knowledge Representation Distribution Group Data Entities Based on YAGO [32]

[7] R. F. Borja, M. Usman, M. M. Wael, and J. L. M. Lastra, "Implementing and Visualizing ISO 22400 Key Performance Indicators for Monitoring Discrete Manufacturing Systems," *Machines*, 2018.

[8] R. Raul Rodriguez, A. S. Juan José, and B. Angel Ortiz, "Quantitative relationships between key performance indicators for supporting decision-making processes," *Computers in Industry*, vol. 60, no. 2, pp. 104–113, 2009.

[9] X. Li, C. Zhou, Y.-C. Tian, N. Xiong, and Y. Qin, "Asset-Based Dynamic Impact Assessment of Cyberattacks for Risk Analysis in Industrial Control Systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 608–618, 2018.

[10] A. Bunte, A. Diedrich, and O. Niggemann, "Integrating semantics for diagnosis of manufacturing systems," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2016, pp. 1–8.

[11] C. Hildebrandt, A. Scholz, A. Fay, T. Schröder, T. Hadlich, C. Diedrich, M. Dubovy, C. Eck, and R. Wiegand, "Semantic modeling for collaboration and cooperation of systems in the production domain," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017, pp. 1–8.

[12] H. Walzel, M. Vathoopan, A. Zoitl, and A. Knoll, "An Approach for an Automated Adaption of KPI Ontologies by Reusing Systems Engineering Data," *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1693–1696, 2019.

[13] M. Taheriyan, C. A. Knoblock, P. Szekely, and J. L. Ambite, "Learning the semantics of structured data sources," *Journal of Web Semantics*, vol. 37-38, pp. 152–169, 2016.

[14] A. Paulus., A. Pomp., L. Poth., J. Lipp., and T. Meisen., "Gathering and Combining Semantic Concepts from Multiple Knowledge Bases," in *Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, INSTICC. SciTePress, 2018, pp. 69–80.

[15] M. Bhole, W. Kastner, and T. Sauter, "A Model Based Framework for Testing Safety and Security in Operational Technology Environments," in *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2022, pp. 1–4.

[16] B. Mehta and Y. Reddy, "Chapter 16 - Asset management systems," in *Industrial Process Automation Systems*. Oxford: Butterworth-Heinemann, 2015, pp. 479–506.

[17] "Smart Manufacturing Starts Data-driven FDT 3.0 Device Type Managers (FDT/DTMS)," https://tinyurl.com/ybkaj6m2.

[18] E. Staff, "NAMUR NE107 Standard," https://tinyurl.com/4hpukz62.

[19] J. Crompton, "5 - Data management from the DCS to the historian and HMI," in *Machine Learning and Data Science in the Power Generation Industry*, P. Bangert, Ed. Elsevier, 2021, pp. 93–122.

[20] E. Wilfried and M. Andrea, *3 - Configuration and Management of Networked Embedded Devices*, ser. Industrial Information Technology, R. Zurawski, Ed. CRC Press, 2017.

[21] B. Mehta and Y. Reddy, "Chapter 22 - Database systems," in *Industrial Process Automation Systems*. Oxford: Butterworth-Heinemann, 2015, pp. 583–592.

[22] "Discovery, Tactic TA0102 - ICS — MITRE ATT&CK®," https://tinyurl.com/mpbzdd89.

[23] B. Carterette and E. M. Voorhees, *Overview of Information Retrieval Evaluation*. Springer Berlin Heidelberg, 2011, pp. 69–85.

[24] J. Zobel, "How Reliable Are the Results of Large-Scale Information Retrieval Experiments?" in *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '98. New York, NY, USA: Association for Computing Machinery, 1998, p. 307–314.

[25] J.-M. Bremer and M. Gertz, "Integrating document and data retrieval based on XML," *The VLDB Journal*, vol. 15, no. 1, pp. 53–83, Jan 2006.

[26] M. Kaszkiel, J. Zobel, and R. Sacks-Davis, "Efficient Passage Ranking for Document Databases," *ACM Trans. Inf. Syst.*, vol. 17, no. 4, p. 406–439, oct 1999.

[27] "Intelligently Extract Text & Data with OCR - Amazon Textract - Amazon Web Services," https://aws.amazon.com/textract/.

[28] C. Tang, "Key Performance Indicators for Process Control System Cybersecurity Performance Analysis," 2017.

[29] S. Šarotar Žižek, Z. Nedelko, M. Mulej, and Živa Veingerl Čič, "Key Performance Indicators and Industry 4.0 – A Socially Responsible Perspective," *Naše gospodarstvo/Our economy*, vol. 66, no. 3, pp. 22–35, 3920.

[30] "Festo Didactic InfoPortal," https://tinyurl.com/f9mn7ay7.

[31] S. K. Anand and S. Kumar, "Uncertainty Analysis in Ontology-Based Knowledge Representation," *New Gen. Comput.*, vol. 40, no. 1, p. 339–376, 2022.

[32] T. Pellissier Tanon, G. Weikum, and F. Suchanek, "YAGO 4: A Reason-Able Knowledge Base," in *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 583–596.