Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 130 (2024) 1644–1652

www.elsevier.com/locate/procedia

57th CIRP Conference on Manufacturing Systems 2024 (CMS 2024)

Data Acquisition Framework for spatio-temporal analysis of path-based welding applications

Georgij Safronov[a,b,*], Heiko Theisinger[a,b], Vasco Sahlbach[c], Christoph Braun[d], Andreas Molzer[d], Anabelle Thies[a,b], Christian Schuba[a], Majid Shirazi[a], Thomas Reindl[b], Albrecht Hänel[c], Philipp Engelhardt[a], Steffen Ihlenfeldt[c], Peter Mayr[b]

[a]*BMW AG, Petuelring 130, 80788, München, Germany*
[b]*Technical University of Munich, Chair of Materials Engineering of Additive Manufacturing, Freisinger Landstraße 52, 85748, Garching bei München, Germany*
[c]*Fraunhofer Institute for Machine Tools and Forming Technology IWU, Nöthnitzer Straße 44, 01187, Dresden, Germany*
[d]*nebumind GmbH, Willy-Messerschmitt-Straße 1, 82024, Taufkirchen, Germany*

* Corresponding author. Tel.: +49-89-382-76743; *E-mail address:* georgij.safronov@bmw.de

Abstract

The use of digital technologies in industrial manufacturing reduces operational costs and improves production quality. The "Framework for Spatiotemporal Production Data Acquisition" (PathSense) aims to improve access and usability of production data by applying a common data ecosystem. Through integrating operational technology (OT) and information technology (IT), PathSense supports decision-making and process optimization. The framework uses the concept of digital shadows to collect critical information from sensors within the continuous manufacturing processes, e.g., welding or gluing lines. Spatial and temporal digital shadows are developed in accordance with human spatial cognition to support sophisticated–yet intuitive–human-computer interactions without the need for advanced data science skills. The backbone of the framework consists of robust data pipelines with a setup incorporating modern IT protocols, such as OPC UA and MQTT, to support the efficient acquisition and management of data. The paper addresses the challenges associated with the combination of IT and OT in cyber-physical systems, stressing modern complex data-intensive manufacturing as one exemplary domain to be tackled by scalable and secure data architectures. The paper identifies two major future directions: refining data integration processes and embedding advanced machine learning algorithms to enable automated data analysis and improve process quality monitoring. In summary, PathSense proposes a data-driven approach for quality inspection in manufacturing, which may eventually enhance industry practices and move towards data-driven decisions and increased operational flexibility.

Keywords: PathSense, digital twin; welding process; data acquisition; spatiotemporal analysis; defect detection; OPC UA; MQTT; AI

1. Introduction and challenges

Digitization and the integration of smart factory concepts are fundamentally restructuring the production landscape [7,17], striving to make every step agile, efficient, and environmentally sustainable [34]. Focusing on the physical production from the forming of coils in the press shop, the gluing and welding of parts in the body shop to the painting of the entire body, two main challenges become apparent:

1. Production processes largely depend on extensive expert knowledge accumulated over decades to be effectively configured and optimized [15,19].
2. There is an increased necessity for standardization of data availability and continuity [12,16].

The combination of expert knowledge and user-friendly data access is essential to overcome these challenges. Organizations can better leverage decades of expertise within

2212-8271 © 2024 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2024.10.295

digital systems by developing a unified and explainable data framework to facilitate decision-making and process optimization [12,13]. Such a framework makes data a practical, easily accessible tool for all employees in production. Using a self-service model that represents the data in a standardized format, no extensive data science knowledge is required to use the data [26]. Specifically, this paper seeks to develop and illustrate a “Framework for Spatiotemporal Production Data Acquisition” (from here on referred to as “PathSense”) that could enable easier data usability and enhance operational efficiency in manufacturing processes.

2. State of the art in managing production data

Implementing a robust data pipeline is the key to PathSense's success. This pipeline must address overarching architectural and technological requirements while naturally accumulating human expertise [28]. Simplified, a data pipeline in manufacturing processes involves four essential elements, as shown in Figure 1. The first element is the process of physical interaction and industrial value generation. The second element are the sensors that measure relevant process data and process parameters during production. These two elements are highly interdependent and require extensive knowledge on the part of the production expert to define meaningful sensors and process parameters [14]. The third element, evaluation, collects the data, consolidates it, and analyses it to gain valuable insights. Again, Expert knowledge is necessary to label complex data sets with insights into complex process understanding. Lastly, the results of the evaluation support the fourth element, utilization. This could be for documentation or decision-making purposes, which may be manual or automated [10,27]. For instance, determining whether a part should proceed to post-processing. Though not covered in this paper, this last element could also play a role in process regulation.

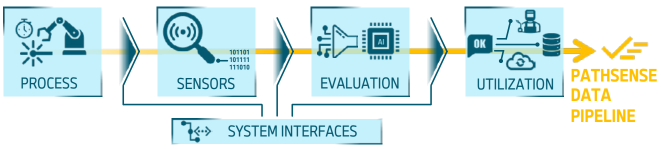

Fig. 1: General concept of the PathSense data pipeline

Ensuring that these steps are linked via interchangeable system interfaces is equally essential to maximize flexibility, as these four elements often crosscut all architectural levels, starting at the shop floor and extending to digital applications hosted in the cloud [3,22,23].

2.1. Challenges and opportunities integrating IT and OT in cyber-physical production systems

The transition towards PathSense requires integrating various data sources in the production layer, contributing distinct complexities to the digital representation of processes [8], as illustrated in Figure 2. These sources encompass meta-information about product design and material properties, operational data from industrial controllers, such as details about motion systems and programmable logic controller (PLC) variables, and diverse sensor readings. Sensor data varies widely, from low-rate metrics like material flow rates to high-rate data captured by advanced technologies such as high-resolution cameras, microphones, or contour scanners. Environmental data are also collected through sensors networked across each station. The value of this data is judged based on its impact on product quality, process stability, and defect detection capabilities.

The Reference Architecture Model Industry 4.0 (RAMI 4.0) is designed to manage these complexities. Although RAMI 4.0 offers a valuable framework, tailoring it to meet specific manufacturing needs remains the responsibility of the manufacturers. Figure 2 depicts a simplified application of RAMI 4.0 across a cyber-physical factory. [35]

A significant challenge in realizing the vision of PathSense is that operational technologies (OT) often trail behind modern information technologies (IT) in terms of data availability, flexibility, and industry-wide standardization. Production facilities typically consist of various technologies from multiple manufacturers, operating on different, sometimes outdated protocols and processing data in isolated silos [29].

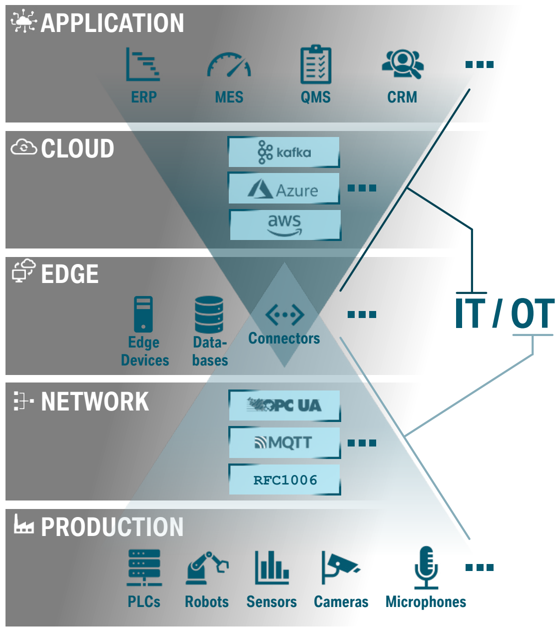

Fig. 2: Architectural visualization of the layers of a cyber-physical production system with converging IT/OT depiction

Quality monitoring systems often come bundled with proprietary analysis tools from sensor manufacturers, which, while convenient for straightforward applications, limit deeper analytical capabilities as they typically deliver simplistic results such as "OK/NOK" (all correct / not all correct). This constrains the ability to integrate data from various sensor systems, underscoring the need for robust interfaces between sensors and evaluation systems [33], as shown in Figure 1. Such interfaces allow sourcing these components from diverse suppliers, thus reducing vendor lock-ins and enhancing flexibility by decoupling sensor procurement from analytic services.

2.2. Network technologies on the production shop floor

Contrasting with the deeply integrated nature of modern IT systems, in which data is ubiquitous, services are abundant, and easy integration is promised through 'Application Programming Interfaces' (APIs), OT systems must continue to evolve while retaining their original robustness and reliability. At the same time, they must become more flexible [29]. The existing data control infrastructure in production, which utilizes PLCs and industrial sensors, employs a range of protocols and bus systems, from traditional technologies like Modbus to modern industrial Ethernet-based protocols like PROFINET [9].

Incorporating modern IT principles, newer OT network protocols such as OPC UA (Open Platform Communications Unified Architecture) and MQTT (Message Queuing Telemetry Transport) are gaining traction. OPC UA performs as an industrial M2M (machine-to-machine) communication protocol. OPC UA offers secure and reliable data exchange in industrial automation, supporting real-time data transfer, alarms, and conditions monitoring, among other features [36]. OPC UA is currently limited for high-frequency data acquisition, which restricts its use for some highly dynamic processes. Meanwhile, MQTT facilitates efficient and reliable data transfer in constrained environments, making it ideal for broadcasting data from industrial sensors to a central broker for real-time monitoring and integration. Given the durability expected of cost-efficient production lines, older standards such as RFC1006 are unlikely to be phased out soon. However, employing hybrid approaches that use multiple protocols is becoming typical in the IT/OT convergence phase as industries strive to incorporate modern data management principles [11].

2.3. Connecting industrial components to the cloud

Modern data applications are used increasingly in data-driven manufacturing, leveraging cloud infrastructure for big data analytics and collecting and storing data from multiple machines in cloud-based data centers [31]. Connectors are the essential mechanisms for linking production equipment with modern data applications. Connectors transform OT data into IT-compatible formats, enhancing data integrity and enabling advanced analytics and decision-making. Additionally, connectors enhance security by managing access controls, securing data transmissions, and protecting sensitive data.

Another aspect to consider is managing the data throughput from the edge to the cloud. High-frequency sensors provide critical insights into process stability and quality, and augmenting this data with further analytics can be highly beneficial. Despite these advantages, there is a pressing need to manage the volume of the data uploaded to the cloud, as data-intensive sensors pose significant infrastructural challenges due to the high network bandwidth they require. Edge devices or locally positioned data centers near the shop floor are integrated physically into the plant, especially those equipped with graphics processing units (GPUs) that enhance parallel processing capabilities, offering a practical solution for machine learning (ML) analytics. These devices process substantial data volumes locally, significantly reducing the amount of data transmitted to the cloud and minimizing data transmission and storage costs. This local processing also reduces latencies, addressing the limitations associated with cloud-only strategies [25]. These systems are remotely and centrally managed through container-based deployments using cloud services, creating a strategic balance between local processing capabilities and cloud-based oversight proper for large-scale implementations [1,18].

For example, Video data can be processed on-site using Artificial Intelligence (AI) models deployed and managed from a central AI platform. Key features extracted from these videos can then be integrated with data from other sensor types within the data pipeline. It is critical, however, to carefully consider this initial processing step. Reducing the volume of data handled can affect the accuracy and relevance of the analyses, so it must be precisely aligned with the specific analytical requirements.

Connecting manufacturing data globally through streaming services like Kafka to cloud architectures and utilizing modern private cloud solutions from providers like Azure or AWS enables the deployment of multiple containerized services on a single data stream. This stateless configuration supports demand-based scaling of computational resources and centralized orchestration, efficiently accommodating process changes and software updates [24]. By orchestrating well-structured cloud architecture, applications in the application layer are enabled, creating value for the production system.

However, this global connectivity introduces significant cybersecurity risks. Implementing stringent security protocols and continuous monitoring is essential to protect data and prevent unauthorized access, ensuring the secure adoption of these technologies [30].

2.4. Enhancing industrial production processes through advanced spatial visualization techniques

As process data from various sensor systems becomes increasingly complex, mainly when representing different data domains, its logical interpretation typically relies on the expertise of production process specialists. Thus, integrating human experts into the data management pipeline is imperative. Complex continuous processes, such as welding or painting, exemplify where this expert integration is crucial. Drawing on cognitive psychology and neuroscience insights, visualizing data in ways that align with human spatial cognition significantly enhances user understanding, particularly for those without a data science background [6]. Research has shown that spatial cognition is crucial in everyday functions and scientific performance, influenced by genetic and experiential factors [32].

Effective spatial data visualization has enhanced cognitive processing, facilitating a better understanding of complex information and supporting informed decision-making in challenging environments. In the realm of educational sciences, the significance of visual/spatial thinking in scientific education is emphasized as essential for problem-solving and grasping complex concepts [20,21]. Despite these advantages, today's production data landscapes, originating from various systems and sensors, often present inconsistent data representations and visualizations, typically in abstract graphs and tables. This necessitates considerable mental effort from users to logically integrate the data, either mentally or through third-party tools and manual data handling, requiring advanced data management skills and substantial time.

2.5. Conclusion and research gap

The core principle of PathSense is to automatically combine data points from different systems and sensors and visually present them to the relevant production experts, leveraging humans' natural spatial cognition in three-dimensional space. This approach equips production experts with tools for self-service, enhancing their ability to directly interact with the data. A reliable and standardized data pipeline and a capable architectural framework are imperative for a successful PathSense implementation.

Despite the existing capabilities, a research gap exists in developing a framework that fully aligns with cutting-edge IT and OT standards for widespread standardization while focusing on spatiotemporal data matching. While PathSense effectively aggregates data for individual use cases, the challenge lies in creating a universally adaptable framework that maintains high standards of adaptability across a wide range of technologies, from handling minimal data points to managing large-scale data streams like video or audio. Moreover, the necessity for interchangeable preprocessing methods to optimize cloud and edge computing resources underscores the need for an innovative approach. This framework must be built from the ground up, tailored explicitly for OT environments while encapsulating modern IT design principles and operational excellence across diverse manufacturing processes.

3. Methodology of a spatiotemporal data framework

By providing production experts with the capabilities to annotate the data with their extensive knowledge, such as marking the exact locations of detected defects in a manufactured part, PathSense opens opportunities for advanced analytics, including AI, to move beyond the conventional data silo analytics currently prevalent in production. This approach aims to boost productivity and deepen engagement with data, thus enhancing operational efficiency and reducing errors.

PathSense's methodology is designed around multiple functional functions, like tracing industry robots' Tool Center Point (TCP) for spatial correlation and synchronizing all data through time mapping, as illustrated in Figure 3. In the visualized concept, the data from a path-based process is recorded with the respective position in the cartesian space G, mapping through time stamp correlation to other sensor data like the current I and voltage U. Also visualized is using more complex sensor systems, like the recorded sound S and videos V. Both are mapped equivalently by the time t. In theory, by such a data representation, the related process expert can better understand correlations of the different sensors by attaching them correlating to the respective positions of the process path $\vec{p}(t)$.

The methodology maps all collected data from the production process, such as currents, voltages, acoustics, and video feeds, according to their specific spatial positions and temporal moments. The framework simplifies data management and enhances alignment with human spatial cognition by attaching data along the process spatial path with the appropriate timestamps.

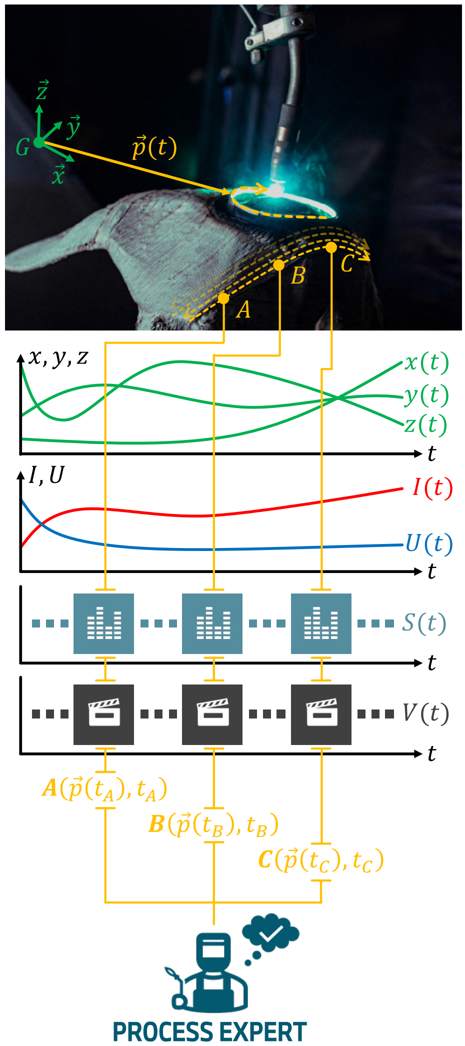

Fig. 3: Core principle of spatiotemporal data mapping

Unlike systems requiring real-time data interactions, PathSense focuses on a digital shadow model where time delays are negligible, and data integration occurs through later spatiotemporal matching. This approach allows for comprehensive data analysis without the immediate pressures of real-time processing [2].

3.1. Flexible data acquisition

Data acquisition within PathSense utilizes OPC UA and MQTT, the chosen protocols for gathering data from PLCs, industry sensors, and "Internet of Things" (IoT) devices. While possible, integration of legacy systems through RFC1006 is often not cost-effective due to the intensive manual labor required for implementation and maintenance. OPC UA file transfer could replace MQTT for specific applications.

OPC UA's support for the Publish-Subscribe (Pub-Sub) model distinguishes it from traditional Client/Server frameworks. This model enables efficient, event-driven data distribution, allowing direct connections to the server or employing middleware that facilitates publishing and subscribing across various applications [37]. This system reduces unnecessary network traffic and system load by minimizing continuous queries for controller or sensor variable changes.

PathSense incorporates a middle-layer data agent that acts as a dynamic intermediary between data sources and the time series database for analytics, providing flexibility to the system by eliminating custom connectors for each data stream from various industrial sensors. This agent enhances system adaptability, enabling seamless integration of new sensors, modification of data streams, or database swaps without extensive overhauls.

Time series databases are designed to efficiently manage time-stamped data, which devices typically generate at regular intervals. This design enhances data ingestion speeds and optimizes data compression, both essential for managing the large data volumes produced in industrial environments. Additionally, these databases offer specialized functions for time-based queries, aggregation, and down-sampling, significantly improving real-time monitoring and historical analysis of production processes [4].

PathSense, while incorporating principles of digital twins, functions more similarly to a digital shadow. It selectively gathers essential data for quality monitoring, using a one-way flow of information from physical systems to digital services. This focus on necessary metrics for quality control aligns more with a digital shadow's characteristics than a digital twin's bi-directional capabilities.

3.2. Localized feature extraction

The framework utilizes a localized, sensor-specific data feature extraction pipeline to manage data-heavy sensors effectively while minimizing network strain. This pipeline can be implemented near the data source through encapsulated services or dedicated hardware, ensuring that cloud management capabilities and access to raw data are maintained. This approach is crucial for avoiding a black-box implementation like in traditional quality monitoring systems, where visibility into data processing and analysis is limited. The pipeline extracts key signal features at set intervals, focusing on attributes like the relative brightness of a video frame or time-domain audio coefficients, such as root mean square (RMS) analysis. Unlike frequency-domain coefficients, RMS values retain time-related information, making them more practical for tagging with recording timestamps.

These features are then tagged with their recording timestamps, allowing for precise tracking and analysis. Separating data preprocessing from storage enhances the framework's agility and scalability, ensuring that data handling is efficient and transparent. For instance, video data can be processed locally using AI models deployed and managed from a central AI platform in the application layer. Important features extracted from these videos can then be integrated with data from other types of sensors in the data pipeline. However, this initial processing must be methodically reviewed to ensure alignment with the specific analytical requirements. Reducing the volume of data managed can affect the accuracy and relevance of the analysis, underscoring the need for a carefully balanced approach to data reduction and processing needs.

3.3. Extracting and matching positional and temporal data

To harness spatial relationships effectively, all data points are assigned specific positions in three-dimensional space. Positional data is either acquired through external tracking or directly from the movement controllers of industrial robots, the latter offering a more cost-effective and direct method. Such data includes timestamps that facilitate the integration of temporally independent sensor data with specific robot positions, enhancing the accuracy of the data fusion process.

Time synchronization, crucial for aligning data across devices, is managed through Network Time Protocol (NTP), ensuring all system clocks are aligned by dictating a mandatory time server application to every system participant as a synchronization target. This setup supports the integration of sensors with diverse sampling rates by allowing for data interpolation or duplication as needed for consistent data analysis.

The framework provides various methods for data fusion to accommodate different sensor outputs and analytical needs. For instance, data from low-frequency sensors can be aligned with high-accuracy positional data through interpolation or carrying forward the last valid values. This flexibility is crucial for detailed and accurate data analysis, supporting sophisticated manufacturing processes.

3.4. Spatial discretization of data into volumetric database

After data is acquired, combined, and synchronized in the time series database, it is transferred and organized into four-dimensional point clouds (x, y, z, t) in a spatiotemporal volumetric database. Additional metadata and unique tags associated with each dataset and sensor differentiate and identify each point cloud. As an essential element of PathSense, the additional spatiotemporal volumetric database enables efficient data fusion and query operations, maintaining the integrity of the original time series data structures.

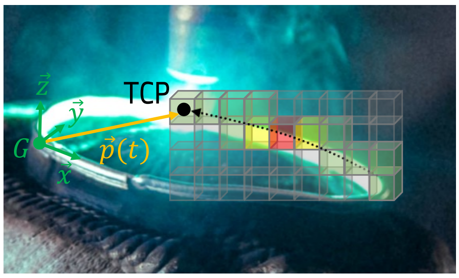

Fig. 4: Simplified voxel matching of the tool path with data aggregation

A specialized query language tailored for this data structure facilitates effective spatial querying and transforms point clouds into voxel data (volumetric pixels representing values within a three-dimensional space) [5], as seen in Figure 4. An indexing structure enhances the speed of voxel value queries, which is particularly beneficial for streaming query results into machine learning applications where transfer efficiency is essential. This querying capability allows the aggregation of sensor values within a voxel, streamlining data evaluation by providing only the necessary data for further analysis. Possible data aggregations include calculating minimum, maximum, average values, and standard deviations across all sensor samples within a spatial voxel. The system also supports other mathematical operations on the data, enhancing analytical flexibility.

Dataset tags play a dual role in querying. They identify the data sets being analyzed and influence the query results. Specific requirements can dictate that only data points from the same dataset be aggregated, even when different datasets are queried. Additionally, comparing datasets that share some tags but differ in others is possible, such as comparing identical components with different serial numbers to identify deviations. Strategically selected tags, combined with timestamps at each spatial point, enable the establishment of physical connections between data points. This capability is crucial for understanding dynamic movements within the manufacturing process, such as robot trajectories.

3.5. Adaptation and abstraction

The deployment of PathSense services is ideally managed through Docker and Kubernetes systems that support the containerization and mass deployments of services. The system's architecture is outlined in a Docker Compose file, which allows for quick and scalable setup of framework instances.

In the forthcoming scientific investigation detailed in the next chapter, the methodology's versatility is showcased through the implementation of Telegraf and InfluxDB as primary instruments for data management within a laboratory framework. These technologies and a proprietary spatiotemporal volumetric database with enhanced visualization capabilities demonstrate and validate the data acquisition and analysis principles in a simplified controlled environment. This experiment serves as a proof of concept for the foundational capabilities of PathSense's approach.

It is important to note, however, that the system's architecture is fundamentally designed to accommodate a variety of tools, extending beyond the confines of the current setup. The modularity enables the integration of alternate solutions such as Kafka, which could replace or complement elements like Telegraf, depending on the specific requirements of different systems.

This flexibility in tool selection affirms the abstract nature of the framework, ensuring that regardless of the specific technologies employed, the essential middle agent's role in a time series-based data collection for maintaining the spatiotemporal integrity of data remains unaltered. Such adaptability accentuates the framework's applicability across various operational scenarios, reinforcing its potential for widespread adoption in diverse data management applications.

3.6. Encapsulation of the PathSense framework

PathSense encapsulates a comprehensive approach to managing and analyzing production data within a spatiotemporal context. By leveraging advanced data acquisition techniques, localized processing, and flexible data integration strategies, the framework is well-suited to meet the evolving demands of modern industrial environments, driving efficiency and innovation in manufacturing operations.

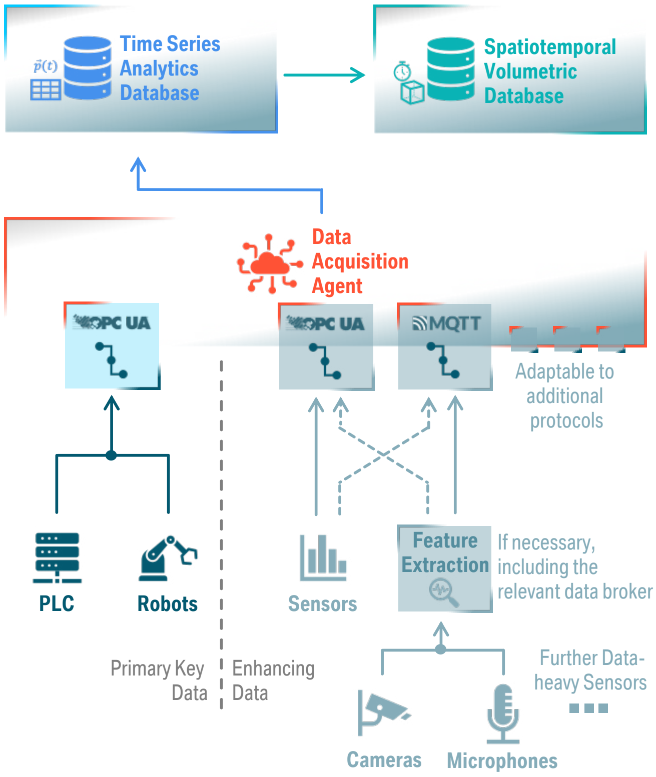

Fig. 5: Methodology of the PathSense framework

As depicted in Figure 5, the overall system implementation offers flexible deployment options tailored based on specific process requirements, accommodating configurations on the edge, in the cloud, or a hybrid model. This mixed deployment model combines the benefits of localized and cloud-based services, enhancing data processing and system responsiveness for diverse operational demands.

4. Experimental implementation of the framework

To validate the functionality of the developed framework within a laboratory environment, an experimental cell was established to emulate a typical path-based manufacturing process. This cell can also be utilized for Wire Arc Additive Manufacturing (WAAM). This setup provides a controlled environment to test the framework's capability to monitor and analyze production processes and possible occurring defects.

4.1. Hardware configuration

The cell consists of a six-axis KUKA KR 70 R2100 industrial robot coupled with a Fronius TPS 400i welding source, integrated with a WF 60i Robacta Drive CMT torch and enabled by the underlying EtherCAT fieldbus (Ethernet for Control Automation Technology). PROFINET could also achieve the same result. The robot's controller is a KR C5 unit with the KUKA.DeviceConnector Advanced 2.1 software package which offers an OPC UA interface for controlling and monitoring user-defined variables. The advanced welding source governs the welding parameters and possesses an OPC UA Server for subscribing to its internal status variables, such as welding current, voltage, and wire feed speed.

Attached to the welding torch, a Xiris WeldMic microphone is installed to capture acoustic data from the welding process. This microphone interfaces with the workstation through a USB connection. The workstation hosts the containerized software stack for the microphone, simulating the localized feature extraction. The workstation also serves as a shared NTP time server to synchronize all connected devices mandatory, ensuring unified data timestamping according to the PathSense Framework.

4.2. Software and data flow

Telegraf, functioning as the central data acquisition agent, establishes connections to data streams from the cell's equipment, as shown in Figure 6. Utilizing OPC UA and MQTT listener plugins, Telegraf subscribes to variable changes under the Pub-Sub model. As data flows from the cell's machinery through Telegraf, it is systematically relayed to the InfluxDB container, which acts as the intermediate time series database. InfluxDB is optimized explicitly for handling time-stamped data, a common feature in industrial settings, and supports rapid data ingestion and efficient data compression. These features are critical for managing the large data volumes produced by the welding process. The database facilitates powerful time-based querying, aggregation, and historical data analysis functionalities.

In addition to storing data in InfluxDB, the framework integrates a proprietary spatiotemporal volumetric database developed by nebumind GmbH that enhances data analysis capabilities by enabling complex queries and storage of volumetric data. This combination of data handling technologies and visualization illustrates the flexibility of the framework to adapt and integrate various data management tools, ensuring robust data processing and storage solutions that can be tailored to specific industrial applications.

4.3. Testing the data pipeline and discussing the results

For the welding experiments, the setup involved fabricating a cylinder using the WAAM process. The base material used was an unalloyed structural steel (type S235) and a G3Si1 steel wire with a 1.2 mm diameter, which served as the welding filler. Shielding was provided by Argon gas mixed with 10% CO_2, and before welding, the base plate was preheated. The cylinder was constructed to have a diameter of 100 mm and a layer height of 1 mm, with the welding robot moving upwards by 1 mm after completing each layer, following a spirally ascending path.

The TCP position and welding parameters, such as wire feed speed, gas flow, welding current, and voltage, were monitored and recorded during the welding process. These parameters were acquired from the welding power source via the OPC UA interface. To ensure accurate time synchronization NTP was used across all data sources. A welding microphone, triggered by a robot-defined variable, captured acoustic data during the process. The microphone was activated and deactivated via a Python script on the workstation, which also subscribed to the OPC UA signal. This script ensured that each audio recording was timestamped to correspond precisely with other welding data, facilitating a comprehensive and synchronized dataset for analysis.

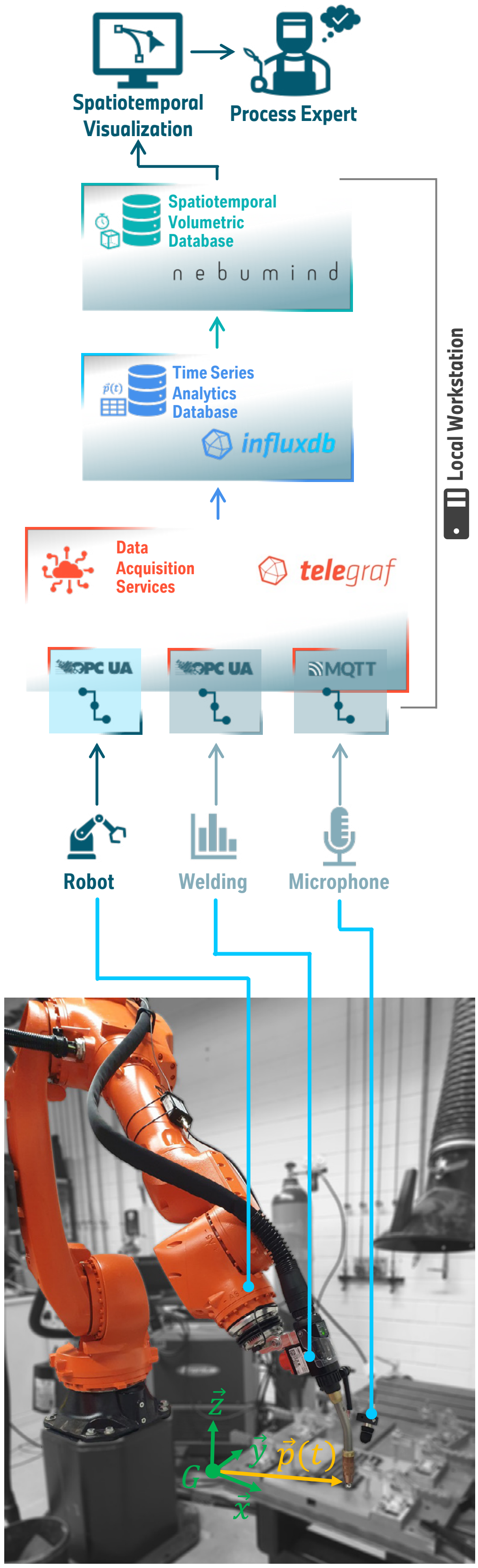

Fig. 6: Experimental PathSence demonstration cell

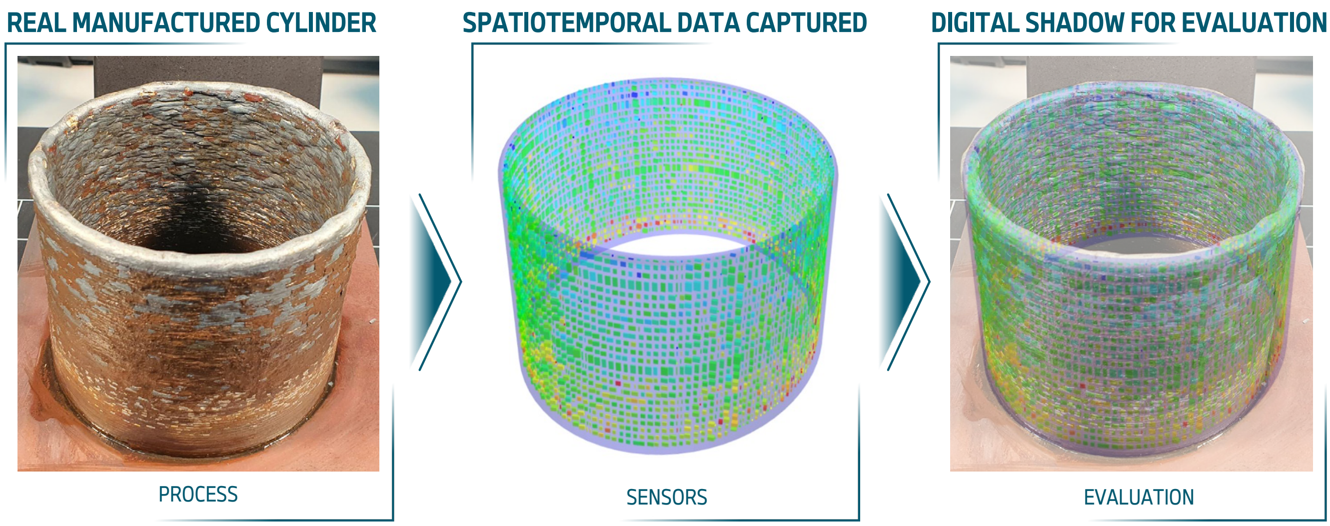

Fig. 7: Experimental results of the PathSense data pipeline visualizing RMS

The welding cell's activities, including the robot's motion and welding parameters, are monitored to implement and test the developed framework. The data captured by the robot, welding source, and the WeldMic microphone are ingested into the framework via Telegraf, processed and matched, and stored in a temporal database. Here, InfluxDB was used. Then, the results are handed over to the space-time volumetric database for spatiotemporal visualization. The principle of this laboratory test is to demonstrate the practicability and expedience of the framework in real manufacturing.

As depicted in Figure 7, the data pipeline of PathSense, as presented in the State of the Art, effectively captures, aggregates, and processes sensor data to create a digital shadow of the manufacturing process. It visualizes the RMS of the audio signal recorded during the WAAM process, which is one of the better ways to measure the average power or amplitude in the audio data. Additionally, the visualization shows a decreasing trend in RMS values correlating with the height of the cylinder during the welding process, which was also audibly perceived as a reduction in sound intensity. This streamlined visualization facilitates intuitive analysis by employees and can be integrated into AI systems for advanced evaluations. The three elements outlined are sensor data collection from the process, data aggregation, and visualization. This highlights the transformation of raw sensor inputs into actionable insights, underpinning the critical role of a digital shadow in enhancing decision-making in manufacturing environments.

This structured data acquisition and analysis approach underpins the framework's capacity to handle complex industrial processes, providing detailed insights into the operational dynamics of the welding cell and demonstrating the framework's applicability in real-world manufacturing scenarios.

5. Conclusion and outlook

The research presented in this paper establishes a foundational approach for the "Framework for Spatiotemporal Production Data Acquisition," referred to as PathSense. Through the strategic integration of operational technologies and information technologies, this framework enables a dynamic, responsive production environment that efficiently harnesses digital shadows. PathSense effectively facilitates the transition from traditional data silos to a unified data ecosystem, where quality monitoring and operational efficiency are significantly enhanced.

The experimental implementation of PathSense has demonstrated the potential to transform manufacturing operations by providing a robust, scalable data architecture that integrates distinct data sources across the production landscape of a specific production process. By leveraging spatial and temporal data, the framework offers enhanced insights that drive informed decision-making processes, thereby fostering a deeper engagement with digital tools among shop floor personnel.

The PathSense team is fully committed to further refining the integration of diverse data streams, expanding the system's capabilities to handle increasingly complex data types, and successfully implementing the framework in real-world production on the shop floor. Adopting advanced AI and machine learning algorithms will be crucial in automating data analysis. The sensor fusion of different data modalities for complex production data will be the main research topic for industrializing the framework's capabilities. Utilizing the visualization capabilities will also be a significant research area in cooperation with users from different plants. The goal is to enhance quality monitoring analytics and proactive data interaction while preserving the scalability inside the production system.

In conclusion, while the results mark a significant milestone, they represent just the initial steps toward realizing the full potential of digital integration in manufacturing. The PathSense framework enables innovations in production data management, promising a future where digital fluency and operational efficiency are linked.

Acknowledgments

The results presented in this paper were achieved under the strategic leadership and orchestration of BMW Group. Individual commissions were directed towards the Technical University of Munich (TUM) and Fraunhofer IWU. We sincerely thank both our longtime partners for their invaluable contributions.

We appreciate the input from nebumind GmbH, who was brought into this project as a distinguished third collaborator through TUM, significantly enriching our research efforts.

Our gratitude extends to Fronius for providing the essential welding power source for our experiments.

We also thank sincerely IWE Erich Kerscher, M.Sc. Attila Goekdemir, M.Sc. Alexander Ruhri, MBA Martin Kleiner, Dr. Anika Stephan and Dr.-Ing. Daniel Wagner from BMW, whose expertise and support have been invaluable in enabling and guiding this research.

References

[1] A.B.A. Alaasam, G. Radchenko, A. Tchernykh, J.L. González Compeán, Analytic Study of Containerizing Stateful Stream Processing as Microservice to Support Digital Twins in Fog Computing, Program. Comput. Softw. 46 (2020) 511–525.
[2] I.V. Anokhov, Digital shadow as a tool for industry exploring, E-Manag. 5 (2022) 80–92.
[3] R. Beregi, G. Pedone, B. Háy, J. Váncza, Manufacturing Execution System Integration through the Standardization of a Common Service Model for Cyber-Physical Production Systems, Appl. Sci. 11 (2021) 7581.
[4] A. Biem, H. Feng, A.V. Riabov, D.S. Turaga, Real-time analysis and management of big time-series data, IBM J. Res. Dev. 57 (2013) 8:1-8:12.
[5] A. Borrmann, S. Schraufstetter, E. Rank, Implementing Metric Operators of a Spatial Query Language for 3D Building Models: Octree and B-Rep Approaches, J. Comput. Civ. Eng. 23 (2009) 34–46.
[6] M. Bugajska, Framework for Spatial Visual Design of Abstract Information, in: Ninth Int. Conf. Inf. Vis. IV05, IEEE, London, England, 2005: pp. 713–723.
[7] R. Chaudhuri, S. Chatterjee, D. Vrontis, A. Thrassou, Adoption of robust business analytics for product innovation and organizational performance: the mediating role of organizational data-driven culture, Ann. Oper. Res. (2021).
[8] D. Flick, S. Gellrich, M.-A. Filz, L. Ji, S. Thiede, C. Herrmann, Conceptual Framework for manufacturing data preprocessing of diverse input sources, in: 2019 IEEE 17th Int. Conf. Ind. Inform. INDIN, IEEE, Helsinki, Finland, 2019: pp. 1041–1046.
[9] B. Galloway, G.P. Hancke, Introduction to Industrial Control Networks, IEEE Commun. Surv. Tutor. 15 (2013) 860–880.
[10] M. Ghahramani, Y. Qiao, M.C. Zhou, A. O'Hagan, J. Sweeney, AI-based modeling and data-driven evaluation for smart manufacturing processes, IEEECAA J. Autom. Sin. 7 (2020) 1026–1037.
[11] C. Giannelli, M. Picone, Editorial "Industrial IoT as IT and OT Convergence: Challenges and Opportunities," IoT 3 (2022) 259–261.
[12] L. Gleim, J. Pennekamp, M. Liebenberg, M. Buchsbaum, P. Niemietz, S. Knape, A. Epple, S. Storms, D. Trauth, T. Bergs, C. Brecher, S. Decker, G. Lakemeyer, K. Wehrle, FactDAG: Formalizing Data Interoperability in an Internet of Production, IEEE Internet Things J. 7 (2020) 3243–3253.
[13] Z. Kaoudi, J.-A. Quiané-Ruiz, Unified data analytics: state-of-the-art and open problems, Proc. VLDB Endow. 15 (2022) 3778–3781.
[14] D. Kibira, Q. Hatim, S. Kumara, G. Shao, Integrating data analytics and simulation methods to support manufacturing decision making, in: 2015 Winter Simul. Conf. WSC, IEEE, Huntington Beach, CA, USA, 2015: pp. 2100–2111.
[15] S. Kondoh, H. Komoto, H. Takeda, Y. Umeda, Acquisition and validation of expert knowledge for high-mix and low-volume production scheduling problems, J. Adv. Mech. Des. Syst. Manuf. 17 (2023) JAMDSM0008–JAMDSM0008.
[16] M. Kuhn, J. Franke, Data continuity and traceability in complex manufacturing systems: a graph-based modeling approach, Int. J. Comput. Integr. Manuf. 34 (2021) 549–566.
[17] J. Lee, E. Lapira, B. Bagheri, H. Kao, Recent advances and trends in predictive manufacturing systems in big data environment, Manuf. Lett. 1 (2013) 38–41.
[18] Z. Li, F. Fei, G. Zhang, Edge-to-Cloud IIoT for Condition Monitoring in Manufacturing Systems with Ubiquitous Smart Sensors, Sensors 22 (2022) 5901.
[19] P. Link, M. Poursanidis, J. Schmid, R. Zache, M. von Kurnatowski, U. Teicher, S. Ihlenfeldt, Capturing and incorporating expert knowledge into machine learning models for quality prediction in manufacturing, (2022).
[20] J.H. Mathewson, Visual-spatial thinking: An aspect of science overlooked by educators, Sci. Educ. 83 (1999) 33–54.
[21] A.J. Mccormack, Developing Visual/Spatial Thinking in Science Education, in: K.S. Taber, B. Akpan (Eds.), Sci. Educ., SensePublishers, Rotterdam, 2017: pp. 143–156.
[22] N.G. Nayak, F. Durr, K. Rothermel, Software-defined environment for reconfigurable manufacturing systems, in: 2015 5th Int. Conf. Internet Things IOT, IEEE, Seoul, South Korea, 2015: pp. 122–129.
[23] G. Neugschwandtner, M. Reekmans, D. Van Der Linden, An open automation architecture for flexible manufacturing, in: 2013 IEEE 18th Conf. Emerg. Technol. Fact. Autom. ETFA, IEEE, Cagliari, Italy, 2013: pp. 1–5.
[24] C. Pahl, Containerization and the PaaS Cloud, IEEE Cloud Comput. 2 (2015) 24–31.
[25] J. Pizoń, J. Lipski, Perspectives for Fog Computing in Manufacturing, Appl. Comput. Sci. (2016) 37–46.
[26] N. Sadati, R.B. Chinnam, M.Z. Nezhad, Observational data-driven modeling and optimization of manufacturing processes, Expert Syst. Appl. 93 (2018) 456–464.
[27] G. Schuh, C. Reuter, J.-P. Prote, F. Brambring, J. Ays, Increasing data integrity for improving decision making in production planning and control, CIRP Ann. 66 (2017) 425–428.
[28] S. Shankar, A.G. Parameswaran, Towards Observability for Production Machine Learning Pipelines, Proc. VLDB Endow. 15 (2022) 4015–4022.
[29] M.C. Shilenge, A. Telukdarie, Optimization of Operational and Information Technology Integration Towards Industry 4.0, in: 2022 IEEE 31st Int. Symp. Ind. Electron. ISIE, IEEE, Anchorage, AK, USA, 2022: pp. 1076–1081.
[30] S. Sriram, G. Rajeshkumar, S. Sadesh, E. Saranya, K. Saranya, K. Venu, Cyber Security Control Systems for Operational Technology, in: 2023 Second Int. Conf. Electron. Renew. Syst. ICEARS, IEEE, Tuticorin, India, 2023: pp. 1–8.
[31] F. Tao, Q. Qi, A. Liu, A. Kusiak, Data-driven smart manufacturing, J. Manuf. Syst. 48 (2018) 157–169.
[32] M. Vasilyeva, S.F. Lourenco, Development of spatial cognition, WIREs Cogn. Sci. 3 (2012) 349–362.
[33] K. Wang, P. Dave, A. Hanchate, D. Sagapuram, G. Natarajan, S.T.S. Bukkapatnam, Implementing an open-source sensor data ingestion, fusion, and analysis capabilities for smart manufacturing, Manuf. Lett. 33 (2022) 893–901.
[34] Die Produktion von Morgen: BMW iFACTORY., URL: https://www.bmwgroup.com/de/news/allgemein/2022/bmw-ifactory.html., Accessed 15 April 2024.
[35] RAMI 4.0 – Ein Orientierungsrahmen für die Digitalisierung., URL: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/rami40-eine-einfuehrung.html., Accessed 20 April 2024.
[36] What is OPC?, Available at: OPC Foundation, URL: https://opcfoundation.org/about/what-is-opc/., Accessed 4 April 2024.
[37] OPC UA PubSub communication model., Available at: GitHub, URL: https://github.com/OPCFoundation/UA-.NETStandard/blob/master/Docs/PubSub.md., Accessed 1 April 2024.