Journal of Machine Engineering, 2024, Vol. 24, No. 2, 18-33
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received*: 12 April 2024* / Accepted: *13 May 2024* / Published online: *19 May 2024*

Industry 4.0,
digital manufacturing,
data spaces,
cognitive production

Christian FRIEDRICH[1*], Stefan VOGT[2],
Franziska RUDOLPH[3], Paul PATOLLA[2],
Jossy Milagros GRÜTZMANN[3], Orlando HOHMEIER[4],
Martin RICHTER[1], Ken WENZEL[3],
Dirk REICHELT[2], Steffen IHLENFELDT[1]

ENABLING FEDERATED LEARNING SERVICES USING OPC UA, LINKED DATA AND GAIA-X IN COGNITIVE PRODUCTION

Value creation in production is based on collaboration of different stakeholders and requires the secure and sovereign exchange of knowledge. Today, knowledge has mostly been built up individually and is only exchanged in a proprietary manner. This paper presents an exemplary pipeline for federated services in cross-domain and cross-company value creation networks for cognitive production. On the example of collaboratively training of a federated machine learning model, machine tool lifetime is predicted in industrial manufacturing for high-end operating resources (high-quality cutting tools). From the shop floor to the cloud, all service relevant information is structured using existing digital twin standards and a linked data approach. In particular, the Industry 4.0 Asset Administration Shell (AAS) and OPC UA are used for collecting and referencing operational and engineering data. GAIA-X connectors transfer the service relevant data through a shared data space. The solution enables intelligent analysis and decision-making under the prioritization of data sovereignty and transparency and, therefore, acts as an enabler for future collaborative, data-driven manufacturing applications.

1. INTRODUCTION AND STATE OF THE ART

In production, enormous amounts of data are generated every second and processed internally in manufacturing machines. This data provides valuable insights regarding the condition of machines, tools, and products that can unlock unprecedented opportunities for value creation. Although machine manufacturers and service providers possess the necessary expertise to develop data-driven applications, they often face a disadvantage due to limited access to this crucial data. Despite advancements in interconnecting plant components with

[1] Department of IIoT Controls and Technical Cybernetics, Fraunhofer Institute for Machine Tools and Forming Technology, Dresden, Germany
[2] Chair of Information Management, HTWD University of Applied Sciences, Dresden, Germany
[3] Department Digitalization in Production, Fraunhofer Institute for Machine Tools and Forming Technology, Chemnitz, Germany
[4] Katulu GmbH, Hamburg, Germany
[*] E-mail: christian.friedrich@iwu.fraunhofer.de
https://doi.org/10.36897/jme/188618

higher-level control systems or integrating entire plants with production control systems like MES, the management, storage, and generation of data predominantly occur in isolated, proprietary systems [1]. Against this background, cognitive production refers to artificial intelligence and machine learning in production processes; technologies that enable a smarter production by automating decision-making processes and optimizing operational procedures.

The adoption of cloud infrastructures, which facilitate collaborative work and leverage machine learning for new insights, has been modest within the production sector for several reasons. A significant barrier is the lack of assured technical solutions for secure data storage and controlled, purpose-specific data exchange that respects data sovereignty, especially with the current leading cloud service providers [2]. Investments in digital infrastructure are often associated with uncertain returns and risks. Nevertheless, machine manufacturers, service providers, and users are adopting Industry 4.0 initiatives and cloud data services to improve their competitive edge and increase sales through the utilization of data. Further, they engage in data exchange and utilization, particularly with the promise of reduced costs, improved quality, and enhanced productivity, provided that the balance of risks and rewards is maintained. However, the fear of a possible loss of intellectual property (IP) and business secrets remains a major problem in highly competitive markets and the reason why many projects and products have not been successful to date.

The journey towards integrating advanced digital solutions in manufacturing has encountered several hurdles, starting with the underutilization of standards that hinder seamless connection to data sources. A significant challenge is the high level of machine-related hardware and software requirements for data collection and provisioning, coupled with a notable lack of consistency in interoperable description forms across different levels. Despite the potential of semantic industry standards such as OPC UA Companion Specifications (CSs) [3] and the Asset Administration Shell (AAS) [4], their adoption as data sources for service applications remains limited. To bridge these gaps, the European cloud initiative GAIA-X, leveraging the International Data Spaces (IDS) Standard [5, 6], proposes a framework aiming to facilitate data-driven applications in alignment with European principles of data protection and sovereignty, emphasizing openness, transparency, and trust [7]. Despite this, the ecosystem still lacks sophisticated connectors, such as the Eclipse Dataspace Components (EDC) Connector [8], essential for linking domain-specific data spaces and enabling the practical realization of new business domains. While a pioneering architectural approach integrating Asset Administration Shells, OPC UA, and the EDC connector within industrial production settings has been explored [9], the systematic definition of cross-domain architecture patterns remains an uncharted area, especially for incorporating new use–cases into the GAIA-X Federation Services. A critical concern that remains inadequately addressed by GAIA-X is the safeguarding of IP after the data transfer and navigating regulatory hurdles like export control, which poses significant barriers to data sharing, especially for data constrained by geographical boundaries.

Moreover, the challenge extends beyond mere data collection from machinery. For the enablement of advanced machine learning applications, data must be enriched with additional information – such as quality measurements or order specifics – to support intelligent analysis and decision-making. This enrichment process is pivotal in generating actionable insights and developing predictive models to improve operational efficiency and product quality.

However, current digital twin standards like AAS fall short in addressing the need for data enrichment. The shift towards Semantic Web and linked data technologies highlights their crucial role in representing, capturing, and integrating industrial data. These technologies, embodying a range of methodologies for articulating and disseminating knowledge across the World Wide Web in standardized formats, promise to enhance data integration and support reasoning processes [10, 11]. Nevertheless, challenges persist, such as data retention issues in Resource Description Framework (RDF) stores, when dealing with high sample rates or extended durations. While specialized time series databases excel in storing vast amounts of measurements, they are limited by their metadata support capabilities. This predicament calls for a flexible approach that merges Linked Data principles with straightforward data formats and APIs for time series data, laying the groundwork for creating semantic digital twins within the manufacturing domain, as outlined in the LinkedFactory (LF) proposal [12], which advocates for portable SPARQL queries over contextualized time series data in industrial settings.

Federated Learning (FL) was successfully used to train and offer services like Google GBoard [13] and many services in Apple iOS [14] all utilizing highly sensitive data like user locations and inputs. FL involves a multi-step process where algorithms are dispatched to and trained directly at the data source. Together with other techniques, such as differential privacy, FL provides a mechanism for training machine learning models while maintaining privacy, addressing intellectual property concerns and circumventing regulatory barriers. In difference to small and simple models that can be trained on-device, i.e. on a charging Wi-Fi connected phone, processing and training machine learning models on large amounts of machine data requires significant computation power and memory. Small to medium sized enterprises usually do not have the IT infrastructure nor personal to supply the required resources. FL as a Federated GAIA-X Service addresses data minimization and anonymization requirements while providing privacy preserving and standardized on-demand infrastructure essential for training complex models. By linking sources using the LinkedFactory, data suitable for machine learning (ML) and cognitive production can be provided.

One significant use-case in production industry is tool lifetime estimation: Tool breakage is one of the main causes of unscheduled downtime in machining and is usually the result of a combination of tool damage and wear characteristics that occur during the machining time and have a negative impact on added value. Accordingly, tool failures are responsible for 7–20% of the total downtime of milling machines [15]. The costs for tools and tool changes account for 3–12% of the total machining costs [16]. Currently, on average, only 50-80 % of the effective life of milling tools is utilized [17]. Research has shown that an accurate and reliable tool condition monitoring (TCM) system can reduce the cost of machining by 10–40% by reducing downtime and maximizing tool life [18, 19]. Lack of access to the required data is the main impedance for developing a well working solution for predicting tool wear across varying machines, tools, workpiece materials and, thus, is an ideal use-case to validate the proposed architecture.

In detail, the following motivations and market requirements arise in the machining production of the exemplary actors described: Tool manufacturers aim to offer new services for their high-end equipment and to open up new value chains based on real tool and load data collected in practice. This can consist of an application-specific tool life prediction, precise parameter recommendations for individual applications, machines and materials or the efficient estimation of the cost of a tool overhaul. On the other hand, tool users want precise parameterization for their application to improve product quality, a reliable tool life prediction to increase the tool life with low risk and a quick statement on the effort and costs of tool reworking. Successful development of a suitable ML model failed due to lack of data. The central challenge within the manufacturing sector is clear: access to and analysis of data is crucial yet impeded by various technical, proprietary, and regulatory barriers. Current deficits in transferring data in accordance with industry standards into the GAIA-X data infrastructure, the systematic integration of machine learning, and ensuring semantic interoperability of various services in GAIA-X emphasizes the complexity of realizing cross-company GAIA-X services based on this real data.

Table 1 summarizes the aspects of already existing solutions presented in this chapter, which exhibit deficiencies and the requirements for the approach presented in the Chapter 2 that improves upon these solutions and is beneficial to meet those shortcomings.

Table 1. Comparison of existing secure data exchange solutions and demanded features

Criteria	Federated Learning	GAIA-X	Market demands
Shopfloor Connectivity	Not Applicable (N/A)	Low, not inherently supported	High, using Digital Twin based data collection
Combining Data Sources	Not Applicable (N/A)	Not Applicable (N/A	Linked Data, seamlessly combining multiple sources ensuring data alignment
Data Sovereignty and Privacy	High, as data remains local (e.g, on-premise)	Moderate, sovereign, but requires strict adherence to GAIA-X policies, as transferred data could be misused	Combining the benefits of Federated Learning and GAIA-X
IT Infrastructure Requirements	Significant, demands robust local computational and storage capabilities	Minimal, cloud-based services minimize local infrastructure dependency	Minimal, optimally uses cloud resources to lower the necessity for extensive local infrastructure

Regarding automated model-update by the use of federated learning, the presented approach can be transferred to other use-cases, such as machine-integrated force measurement (kinetic model) [20], geometric error calibration (kinematic model) [21], or thermal error correction for machine tools (thermal model) [22].

2. APPROACH AND ARCHITECTURE

The aim of this contribution is to realize the basic architecture and data models for an innovative product service application in industrial manufacturing using the example of high-end operating resources, such as high-quality cutting machine tools in a value-added network using Federated Leaning and the secure, sovereign data infrastructure of GAIA-X. This effort is contextualized by the tool lifetime estimation use-case presented earlier, serving as a pivotal experiment and proof of concept that both challenges and substantiates the proposed methodologies. Data about the tool itself, its manufacture, use, and logistics processes are captured and established machine/plant level communication protocols are explicitly considered and semantically integrated based on their information models. This strategic collection of user-generated data provides a foundation for innovative business services based on data analytics. As today, numerous digital service technologies are being developed at the same time, small and medium-sized enterprises (SMEs) have difficulties to apply them to their applications. The novelty of this contribution is the application of technologies developed in the context of Industry 4.0 and GAIA-X to the specific and relevant use-case tool wear prediction, while, at the same time, including federated learning as a GAIA-X service as well the Fraunhofer LinkedFactory as an interconnected database. The presented architecture fosters a collaborative space for manufacturers (of machines and tools alike) and users, facilitating the training of ML models with authentic application data in a privacy-preserving manner.

The utilization of FL, in particular, emphasizes the commitment to data sovereignty, enabling model training at the GAIA-X service node and the synthesis of data local insights into a cohesive, higher-level model. The results of this collaborative learning not only enhance the accuracy of tool life predictions and enable more tailored parameterizations for specific applications (considering machine, tool, material, and process variables) but also empower manufacturers to refine their products based on actionable data insights. Conversely, users stand to gain improved product quality and productivity, alongside reduced operational risks.

The consistent use of information models from OPC UA (especially from the umati community) and AAS enables broad applicability beyond the consortium. The AI models generated in the process significantly reduce the entry barrier to machine learning in production while at the same time securing data sovereignty. Figure 1 illustrates the relationship between the basic components explained below. In industry, the standardized *OPC UA interface* is becoming increasingly popular for connecting machines and control systems [23]. This standardization is useful for connecting various control systems from different manufacturers. For the use-case presented in this paper, the generic VDMA Companion Specification for machine tools 40501 [24] was taken as a base and applied to the specific machine tool. The result shown in Fig. 2 depicts an exemplary milling machine to illustrate how the structures of the machines are described in the OPC UA information model. For example, the Maho machine consists of four axes and one spindle as components. While OPC UA is the standard for productive use and maintenance, it is not recommended for communication across the entire life cycle and the associated information along the value chain. Instead, the Asset Administration Shell (AAS) is used for this purpose [25].

The *Asset Administration Shell* (AAS) serves as a standardized digital representation of an asset, such as a machine, tool or good, and thus plays a fundamental role for the digital twin-based data provisioning in the presented use-case by facilitating interoperability between applications responsible for managing manufacturing systems. It holds digital models of different aspects of an asset, referred to as sub models, and provides descriptions of technical functionalities exposed by the assets or the corresponding AAS [4].

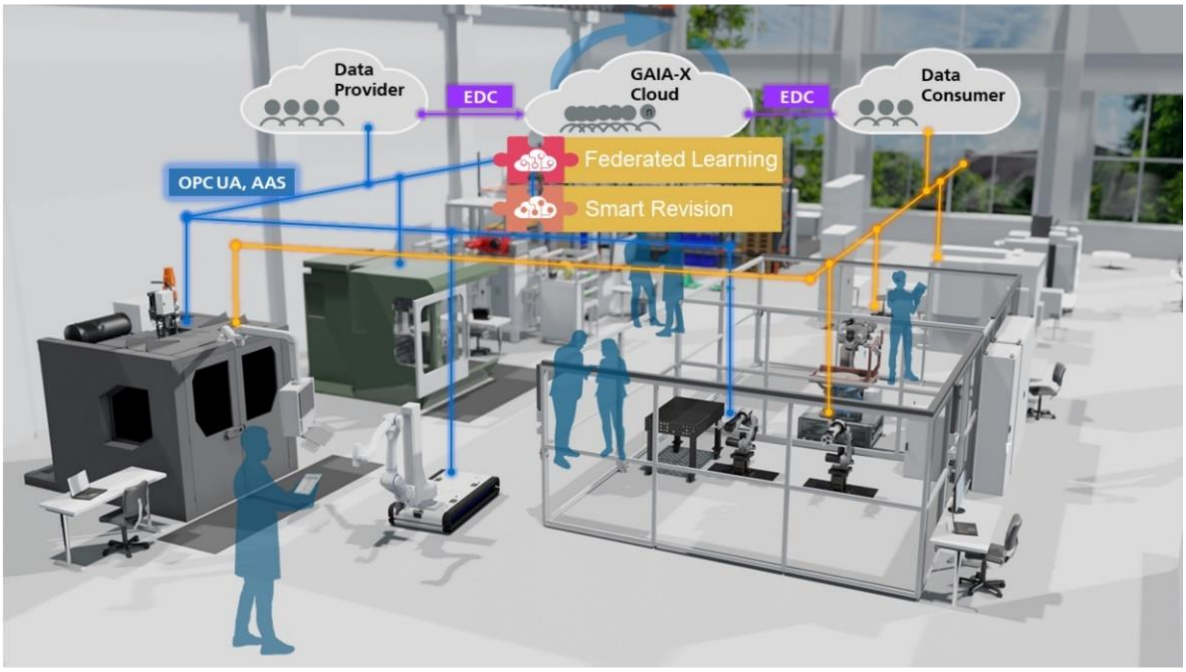

Fig. 1. Approach of Federated Services using OPC UA and the GAIA-X Cloud

The overall structure of the AAS is described as a metamodel defined by the Industrial Digital Twin Association (IDTA). During its life cycle or the assets life cycle, respectively, the AAS can be enriched with additional information. An important part of the AAS infrastructure is the AAS Registry, a collection to store and look up information on multiple AAS instances, their endpoints, and sub models. Currently there are several AAS frameworks under development. The most common are FA³ST AAS from Fraunhofer IOSB, BaSyx AAS Environment from Fraunhofer IESE and the AASX server from the IDTA itself. As the FA³ST AAS currently does not offer a dedicated MQTT interface for external system communication [26] and the BaSyx and IDTA frameworks are equivalent regarding the supported functionalities, the Java-based BaSyx framework was chosen for the implementation.

The *Eclipse Dataspace Connector* is the main component of the open-source Eclipse Dataspace Components (EDC) Framework [8] for trusted, self-sovereign, inter-organizational data exchange. It allows to build, access, and connect multi-cloud, policy-based dataspaces while retaining full control over the own data offered. The approach utilizes the connector to catalogue available data assets on the data provider side, such as time series data for smart revisioning or federated learning services and publish them as contracts with associated policies to the participants of the data space. Data sources can be any references to local filesystems, cloud storage systems, or APIs. For policy definition, the Open Digital Rights Language (ODRL) is used to determine permissions, prohibitions, and obligations that regulate access and usage of an asset. For example, the usage of an asset can be restricted to the purpose of training of a machine learning model. The framework includes a set of official service extensions that enhance the connector's core functionality with features like persistence, authentication, or connectivity to provider-specific cloud services. On the consumer side, another connector instance can be used to fetch the provider's catalogue and request access to the desired data assets.

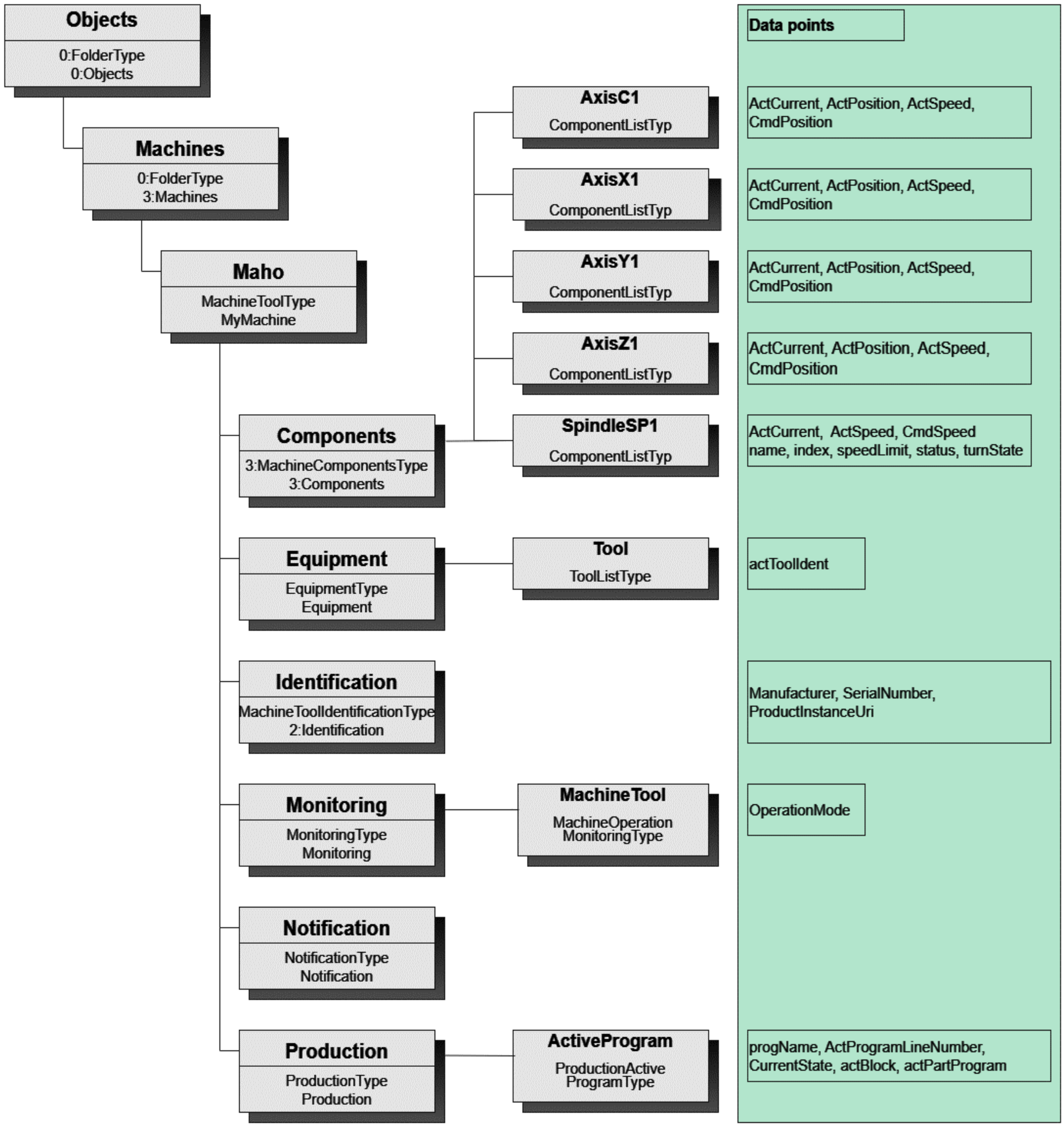

Fig. 2. OPC UA CS Machine Tool

The proposed architecture describes a robust software pipeline from the Shopfloor to the cloud based on open-source software. At its core, the architecture encapsulates a multi-tiered approach, orchestrating the collection and linking of operational (OPC UA) and engineering data (ERP/MES) originating from different sources of the Shopfloor and Topfloor level. Next, digital twin-based data aggregation and referencing as well as secure and sovereign data transmission via Gaia-X dataspaces is implemented, as shown in Fig. 3. To this end, it enhances the base architecture described in [9] with Linked Data concepts to support intelligent analysis and decision-making for federated services.

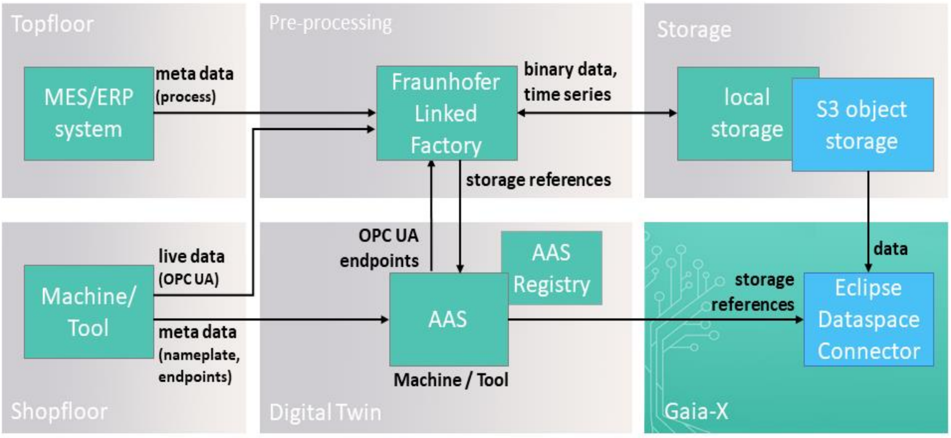

Fig. 3. System architecture within a company

3. IMPLEMENTATION

Based on the proposed architecture, an exemplary implementation for the use-case of tool wear predicting for milling machines is described next, including the data description, the data pipeline as well as the integration of federated learning algorithms.

3.1. SEMANTIC MODELS FOR PRODUCT AND PROCESS DATA

To draw conclusions on the factors influencing product quality and production performance, the combination of data on processes, products, and resources is necessary. A flexible data architecture should allow for integration of observed objects and their attributes at any time while semantic models help to organize the data by capturing structural as well as provenance information. The proposed approach is based on the LinkedFactory (LF) architecture that defines a simple data format and protocol to capture, store, and retrieve production data. This web-oriented approach allows to combine the semantic descriptions of production systems (stored as RDF) and large amounts of production data (stored in a time series database), i.e. from MES or ERP systems and OPC UA servers [12]. The LF uses a JSON-LD inspired data format - a lightweight linked data format that employs a simple hierarchical structure to represent items and their attributes - and an HTTP-based APIs to exchange the data [27] The attribute values are further qualified by a context (organizational unit, simulation run, etc.) and a timestamp. Therefore, all LF-JSON data represents time series, which can be accessed by item, attribute, and context, as shown in Fig. 4.

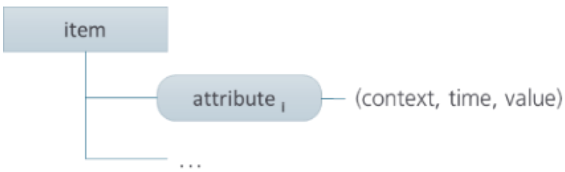

Fig. 4. Simple hierarchical data format

This approach enables the representation of data for multiple use-cases in different views, for example part-, process-, and machine-centric views, exemplary depicted in Fig. 5.

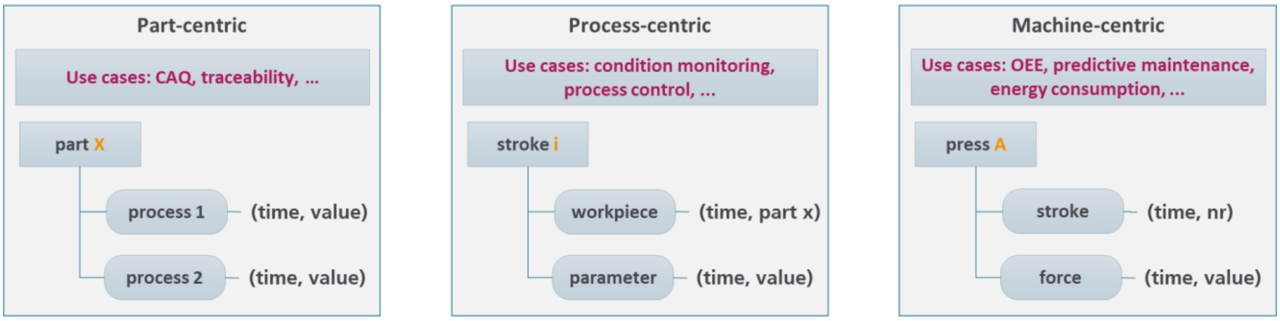

Fig. 5. Different views of production data

The retrieval of time series data was facilitated by machine data collection from an OPC UA communication. In this way, the time series data provided by an OPC UA server is serialized into the above-mentioned JSON format to be stored into the instance of LinkedFactory, which is accomplished through a specific OPC UA connector. For the presented use-case, the information model, shown in Fig. 2, serves as a base for building the serialized JSON data. In addition to the data collected as time series data, metadata of the produced workpiece and the tool used needs to be stored to add the required semantic to the data points. Therefore, a metamodel is created according to the Semantic Aspect Meta Model (SAMM), in which Digital Twins are seen as a collection of aspects that supply a domain-specific view on an asset using RDF vocabulary [28]. SAMM describes in a formal and therefore machine-readable format the structure of an aspect, i.e. the physical unit and the possible value range. For the metamodel of the underlying use-case, the properties of the workpieces and the machine tools (parameters, their unit of measurement and their limit/threshold) were described (see excerpt in Fig. 6) since they are currently not included in the existing OPC UA information models (Fig. 2). The depiction in Fig. 7 shows exemplary how the three different views (machine, workpiece, and tool) are linked among each other. Based on the machine runs, the start and end times of the production of a workpiece are tracked. During the presented run on the machine “Maho” with the given time the workpiece “wp-3” was manufactured with the machine tool “X1_Fed_X_Pro”. In the milling process, the variable “spindleCurrent” includes the time series data of the spindle current, shown here only for one datapoint. The produced workpiece “wp-3” originates from a specific raw material or previous production step with a given hardness and tensile strength. Furthermore, the production process by which the workpiece was machined in the given production time is shown, e.g. “Drilling”. Moreover, the designated machine tool “X1_Fed_X_Pro” can be identified with a specific Data Matrix Code (DMC) and had the following process values (feed and rotational speed). The previous explanation points out different representations of the data stored in the LinkedFactory instance for the machine, workpiece, and machine tool data. These are instance data and time series data as JSON-based format as well as semantic description as RDF models. With this regard, it is possible to use the query language SPARQL to merge the data sources into one result.

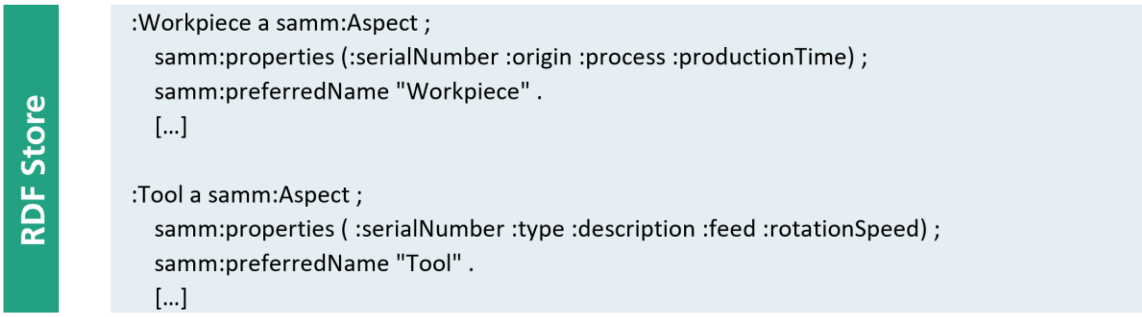

Fig. 6. Semantic Description in RDF

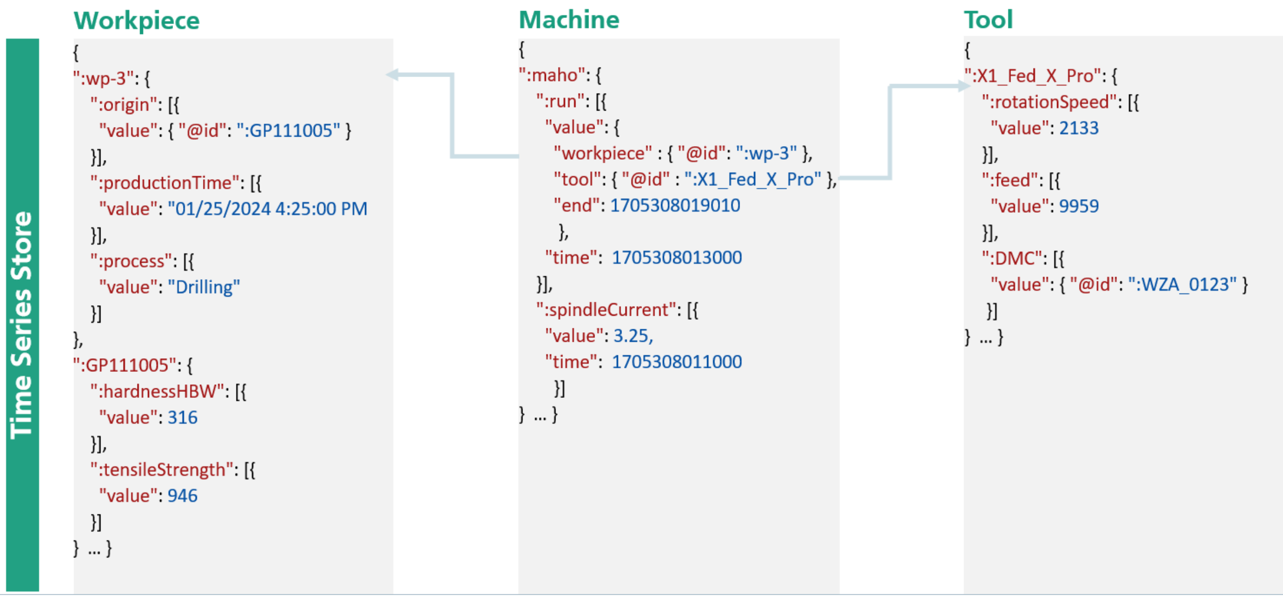

Fig. 7. Connection of different time series items for machine, workpiece and tool (data linking)

An extension leveraging so-called federated queries allows to connect with LF HTTP-APIs [29]. For fetching time series data related to a particular item with SPARQL, federated queries are used with a custom service extension to retrieve the data. In conclusion, it can be stated that the LF approach provides a standardized format for managing time series data emerging from diverse sources and enriches it with the according semantic behind. With the LF SPARQL endpoint interface data frames can be retrieved, filtered, and combined for building and training machine learning models or to perform other data analysis..

3.2. DATA PIPELINE

For the pipeline it is assumed that the AAS of a machine or tool is created and registered once the asset exists in the company, as it is supposed to represent the entire life cycle of an asset. Fig. 8 illustrates the data pipeline from asset registration to the transfer of digital twin information via a shared data space. The LinkedFactory uses the digital twin information stored in the AAS, i.e. the OPC UA endpoint references, to configure its connectors and retrieve the required information directly from the different source systems. This is particularly useful for time series information, as a lot of machine data is generated at a high cycle frequency on the production floor and OPC UA only propagates variables after a change. The LinkedFactory APIs for managing the meta model and the time series data are implemented within a platform called LinkedFactory POD. This is a Java-based software that provides the LF web APIs and several backends for storing RDF models and LF-JSON-based time series data. The linked data output is written to a storage location, that is then referenced in the AAS as "External Segment" [30] and catalogued via its data address by the EDC connector for data transfer. On the consumer side another EDC connector instance in the shared data space can be used to fetch the provider's catalogue and request access to the desired AAS properties. It is essential to highlight that the storage references are only used by the provider's EDC connector to retrieve the actual data. References are not shared with the consumer and the consumer does not access the storage locations.

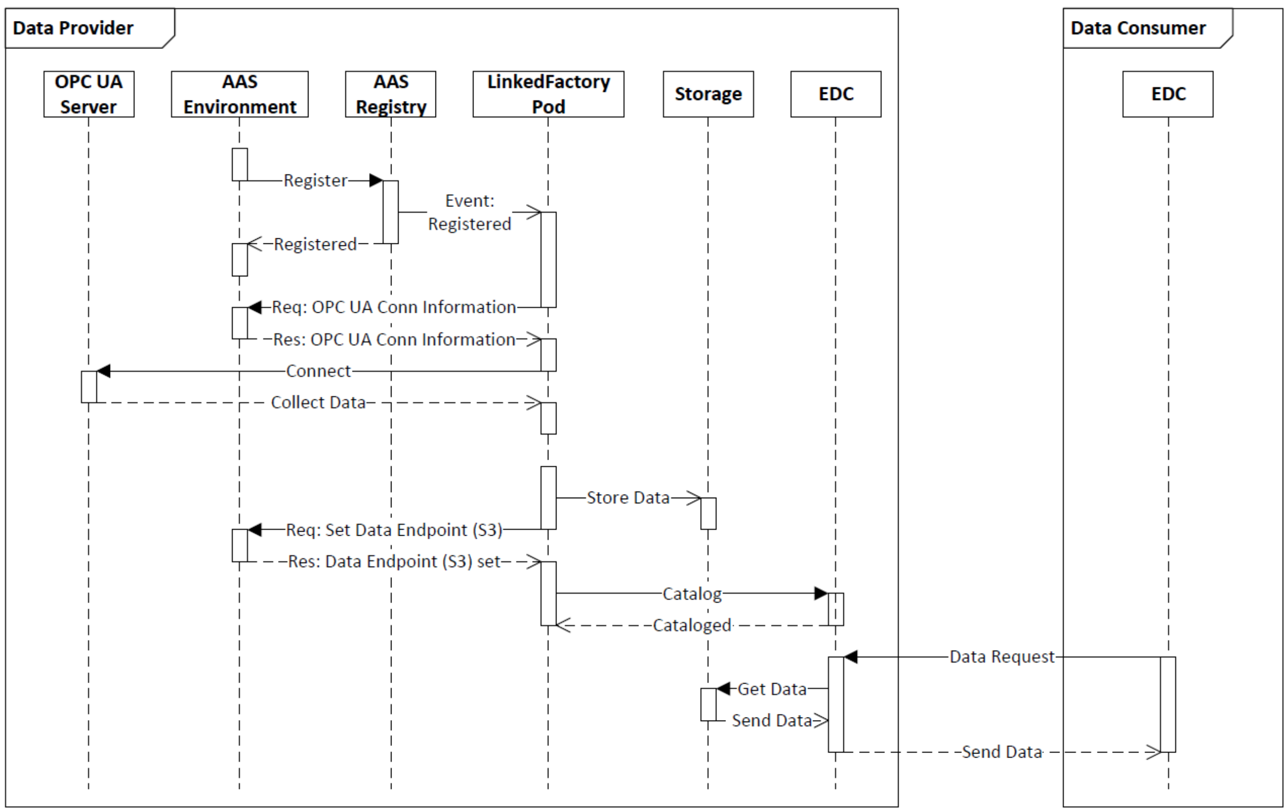

Fig. 8. Sequence diagram of the proposed data pipeline

3.3. EDC AAS EXTENSION

Unlike *FA³ST* [31], the Eclipse BaSyx project does not yet feature an AAS extension for the EDC connector that automatically catalogues entries from Asset Administration Shells. Therefore, a custom service extension had to be implemented to periodically check the AAS registry for any changes to the registered administration shells and update the catalogue if necessary. The update process can also be triggered using MQTT events, for example from the LinkedFactory. To avoid the cataloguing of all AAS entries, which could unintentionally give away information about the overall structure of the Asset Administration Shells once the catalogue gets fetched by a consumer [9], an entity of the AAS metamodel attribute “HasExtension” [32] is used as an indicator to whether the entry should be catalogued or not. The flag of type Boolean can be set for each level of the metamodel hierarchy: AAS, SubModel, SubModelCollection or property. If the flag is negated at runtime, the corresponding catalogue entry and all subordinated AAS elements are removed recursively. Additionally, a second extension entity of type String was added to enable the automatic policy assignment from an expandable set of pre-defined policy definitions, for example “confidential-policy” or “ML-training-only-policy”. After cataloguing an asset, a policy that matches the entity name is assigned. This allows for automatic linking of policies and assets beyond a single default policy for all assets.

3.4. FEDERATED LEARNING

Federated Learning represents a paradigm shift in machine learning, where the training of algorithms is decentralized across multiple devices or systems, known as nodes, without exchanging the data they hold. This approach is particularly beneficial in scenarios where data privacy, security, and sovereignty are of paramount importance. The integration of the Eclipse Dataspace Connector with the GAIA-X Federated Service, operating a Federated Learning agent, marks a pivotal step in the data flow from shop floor to machine learning in the cloud.

The Federated Learning Agent, utilizing data facilitated through the EDC connector, enables the aggregation of data from diverse participants into a unified machine learning model. This process builds on the GAIA-X guarantee of data sovereignty and extends it for the development of machine learning models addressing core concerns about participating in machine learning efforts with an emphasis on maintaining privacy. The Federated Learning Agent retrieves sample data from the Data Provider using the EDC connector. It then extracts the data schema, which becomes the basis for registering the data source within the Federated Learning Platform. This procedure allows the Federated Learning Platform to employ its patent-pending federated pipelines for data alignment and preprocessing, ensuring dataset consistency across different parameters such as frequency and field availability. These pipelines play an essential role in the preparation of data for machine learning, performing critical operations, such as data normalization [33]. As federated learning sessions are initiated, the Federated Learning Platform dispatches tasks to the Federated Learning Agent, as depicted in Fig. 9. The Agent then processes these tasks using the *local* data while adhering to the data privacy and security standards set by GAIA-X. This deployment offers a scalable and flexible solution providing the required memory and computational resources when needed while maintaining data privacy and security. A combination of feature extraction on mechanical signals like forces, torques, and vibrations, along with a deep learning architecture is used to provide a robust modelling approach. The Federated Learning model architecture is designed to forecast tool wear with high precision. It is built around stacked recurrent neural network (RNN) layers and employs nonlinear regression to effectively model the temporal dependencies inherent in the tool wear process. The rich set of dynamically computed features—encompassing statistical, Fourier, and Wavelet transformations—play a critical role in the model's success. These features, computed on a rolling basis, introduce an additional computational burden but are indispensable for capturing the nuanced behaviours indicative of tool condition.

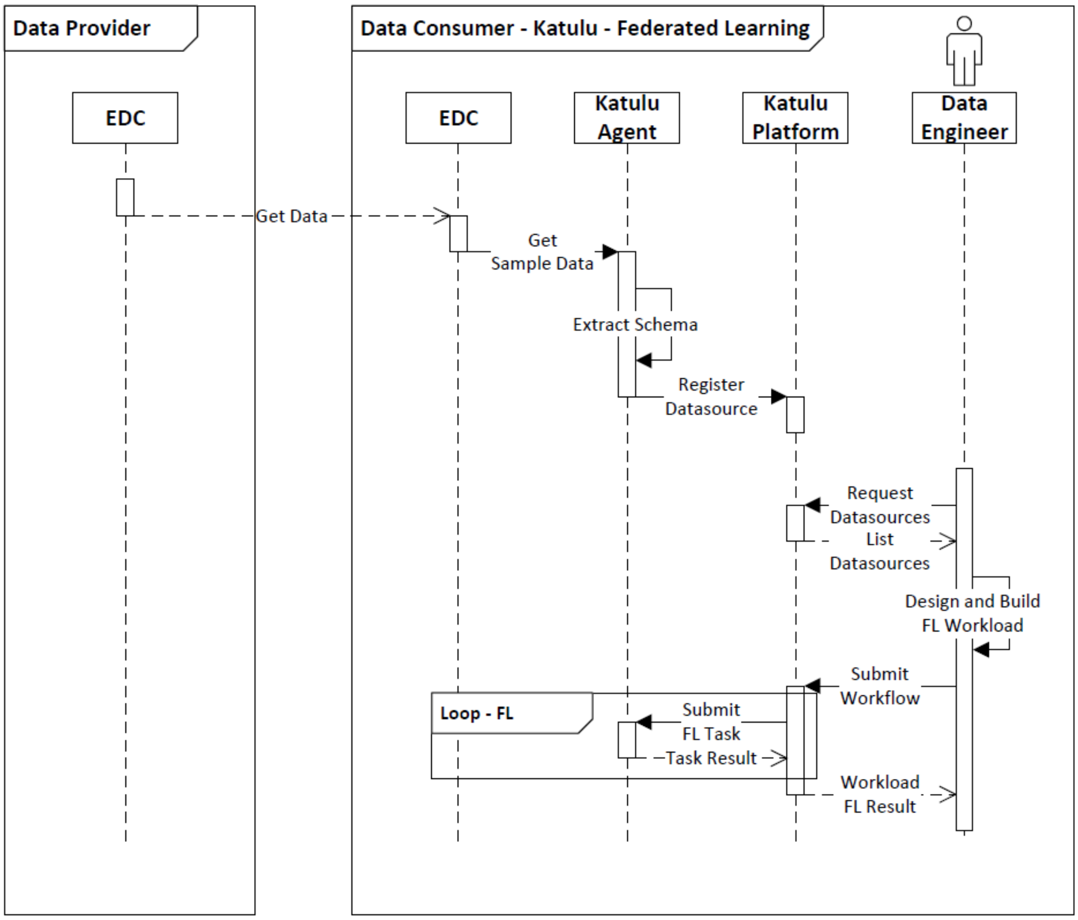

Fig. 9. FL data flow with EDC and Federated Learning Agent

This additional computational burden highlights the need for a flexible IT infrastructure to efficiently provide the required compute and memory. In the FL implementation, the FedAvg algorithm applies as the primary federated learning method, leveraging its efficacy and simplicity in model update aggregation [34]. In addition, for environments where data sources are markedly heterogeneous, the concept of Clustered FL is incorporated as an advanced strategy to enhance performance by grouping similar datasets for training [35], thus, optimizing the learning process for data diversity.

The implementation of the proposed architecture and validation experiments are in progress. Detailed results will be presented in a subsequent contribution that focus on the practical aspects and data analysis.

3.5. LIMITATIONS

The presented implementation primarily describes an architecture for data provisioning and has some limitations regarding the digital twin-based data exchange on the consumer side. The custom EDC connector extension does not yet allow to automatically synchronize Asset Administration Shells across company boundaries. Therefore, the consumer needs to know or identify which asset to request from the catalogue listing. One general limitation is the lack of a timestamp for AAS entries that could serve as an update indicator for synchronization purposes. The utilization of the AAS "HasExtension" attribute enables the automatic assignment of a suitable policy to an asset in the EDC connector. However, this also means that the policy templates need to be managed and maintained. Finally, the pipeline does not yet provide an easy-to-use Plug&Play solution for SMEs and has no graphical user interface to review the EDC connectors catalogue.

4. CONCLUSION AND SUMMARY

In conclusion, this paper presents an approach to address the complex challenges inherent in collaborative value creation networks within the manufacturing sector using the example for tool wearing prediction in cognitive production. By leveraging existing AAS digital twin standards, Linked Data concepts and GAIA-X federated services, it proposes a robust software pipeline for the secure and sovereign exchange of knowledge. Thereby the paper offers a proof-of-concept architecture, to orchestrate a seamless integration of disparate systems, focused on interoperability and standardization within industrial settings. The integration of the LinkedFactory approach in the solution enables the usage and querying of annotated time series and semantics data in order to have complete production data sets that allows for the potential outsourcing of complex data analysis tasks. The extension of the EDC connector for BaSyx, aimed at mitigating the inherent limitations of existing systems [9]. Using federated learning and the cohesive collection, aggregation, and transmission of data, the solution sets a foundation not only for the specific use-case of machine tool lifetime prediction but also for extending to broader manufacturing challenges. By prioritizing data sovereignty and transparency, while also facilitating intelligent analysis and decision-making, the proposed architecture emerges as an enabler for future collaborative, data-driven manufacturing applications. In summary, the solution outlined in this paper represents a step forward in unlocking the potential of distributed data for value creation in manufacturing. It emphasizes the importance of collaboration among stakeholders and the necessity of secure and sovereign knowledge exchange in driving innovation and efficiency within the industry.

ACKNOWLEDGEMENTS

The project was funded by the German Federal Ministry of Education and Research (BMBF) with the project "Federated Services based on OPC UA and GAIA-X in Production (Fed-X-Pro)", 02J21D160. The authors want to thank the editors and reviewers for their helpful comments and constructive suggestions regarding the revision of this paper.

REFERENCES

[1] VOLKWEIN M., SCHMITT J., HEIDELBACH J., SCHÖLLHAMMER O., EVCENKO D., KETT H., 2022, Blinde Flecken in der Umsetzung von Industrie 4.0 – identifizieren und verstehen, Acatech, https://publica. fraunhofer. de/handle/publica/434370.
[2] HILLERMEIER O., PUNTER M., SCHWEICHHART K., USLÄNDER T., 2021, Data Sovereignty – Critical Success Factor for the Manufacturing Industry, https://doi.org/10.5281/zenodo.5675873.
[3] OPC FOUNDATION, UA Companion Specifications, 2024, https://opcfoundation.org/about/opc-technologies/opc-ua/ua-companion-specifications.
[4] Plattform Industrie 4.0, 2024, Details of the Administration Shell – Part 1: The Exchange of Information Between Partners in the Value Chain of Industrie 4.0, available from: https://www.platt-form i40.de/ip/redaktion/de/downloads/publikation/details_of_the_asset_administration_shell_part1_v3.pdf.
[5] IDS International Data Spaces Association, 2024, IDS Reference Architecture Model (Version 3.0).
[6] OTTO B., RUBINA A., EITEL A., TEUSCHER A., SCHLEIMER A.M., LANGE C., et al, 2021, GAIA-X and IDS, International Data Spaces Association.
[7] GAIA-X Technical Architecture, 2024, GAIA-X- Architecture Documnt, https://gaia-x.eu/wp-content/uploads/2022/06/gaia-x-architecture-document-22.04-release.pdf.
[8] Eclipse Foundation, 2024, Eclipse Dataspace Components Connector, https://github.com/eclipse-edc.
[9] NEUBAUER M., STEINLE L., REIFF C., AJDINOVIC S., KLINGEL L., LECHLER A., et al, 2023, Architecture for Manufacturing-X: Bringing Asset Administration Shell, Eclipse Dataspace Connector and OPC UA Together, Manufacturing Letters, 37, 1–6.
[10] ZHENG Z., ZHOU B., ZHOU D., SOYLU A., KHARLAMOV E., 2022, ExeKG: Executable Knowledge Graph System for User-Friendly Data Analytics, Proceedings of the 31st ACM International Conference on Information & Knowledge Management, https://doi.org/10.1145/3511808.3557195.
[11] ZHENG Z., ZHOU B., ZHOU D., ZHENG X., CHENG G., SOYLU A., et al, 2022, Executable Knowledge Graphs for Machine Learning: A Bosch Case of Welding Monitoring, The Semantic Web – ISWC 2022, 21st International Semantic Web Conference, Virtual Event, Proceedings, Berlin, Heidelberg, Springer-Verlag, 791–809.
[12] FRIEDEMANN M., WENZEL K., SINGER A., 2019, Linked Data Architecture for Assistance and Traceability in Smart Manufacturing, MATEC Web. Conf., 304, 4006.
[13] XU Z., ZHANG Y., Andrew G., Choquette-Choo C.A., KAIROUZ P., MCMAHAN H.B., et al, 2023, Federated Learning of Gboard Language Models with Differential Privacy, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics.
[14] Apple Machine Learning Research. Learning Iconic Scenes with Differential Privacy, 2024, https://machine learning.apple.com/research/scenes-differential-privacy.
[15] VETRICHELVAN G., SUNDARAM S., KUMARAN S.S., VELMURUGAN P.,2015, An Investigation of Tool Wear Using Acoustic Emission and Genetic Algorithm, Journal of Vibration and Control, 21, 3061–3066.
[16] LIU C., WANG G.F., LI Z.M., 2015, Incremental Learning for Online Tool Condition Monitoring Using Ellipsoid Artmap Network Model, Applied Soft Computing, 35, 186–198.
[17] SALONITIS K., KOLIOS A., 2014, Reliability Assessment of Cutting Tool Life Based on Surrogate Approximation Methods, The International Journal of Advanced Manufacturing Technology, 71/5, 1197–1208.
[18] REHORN A.G., JIANG J., ORBAN P.E., 2005, State-of-the-Art Methods and Results in Tool Condition Monitoring: A Review. The International Journal Of Advanced Manufacturing Technology, 26/7, 693—710.
[19] KARANDIKAR J., MCLEAY T., TURNER S., SCHMITZ T., 2015, Tool Wear Monitoring Using Naïve Bayes Classifiers, The International Journal of Advanced Manufacturing Technology, 77/9, 1613–1626.
[20] FRIEDRICH C., IHLENFELDT S., 2021, Model Calibration for a Rigid Hexapod-Based End-Effector with Integrated Force Sensors, Sensors, 21/10, 3537; https://doi.org/10.3390/s21103537.
[21] KAUSCHINGER B., FRIEDRICH C., ZHOU R., IHLENFELDT S., 2020, Fast Evaluation of the Volumetric Motion Accuracy of Multi-Axis Machine Tools Using a Double-Ballbar, Journal of Machine Engineering, 20/3, 44–62.
[22] FRIEDRICH C., GEIST A., YAQOOB Mf., HELLMICH A., IHLENFELDT S., 2024, Correction of Thermal Errors In Machine Tools by a Hybrid Model Approach, Applied Sciences, 14/2, 671, https://doi.org/10.3390/ app14020671.
[23] WEISS A., IHLENFELDT S., 2022, Integration of Opc Ua Information Models Into Enterprise Knowledge Graphs, Journal of Machine Engineering, 22/2, 138—147.
[24] UMATI, 2024, OPC 40501 UA for Machine Tools, https://umati.org/industries_machine-tools/.
[25] DRATH R., MOSCH C., HOPPE S., 2023, Diskussionspapier – Interoperabilität mit der Verwaltungsschale, OPC UA und Automationm, Zielbild und Handlungsempfehlungen für Industrielle Interoperabilität.

[26] FRAUNHOFER IOSB, 2023, FA³ST Service: Getting Started, https://faaast-service.readthedocs.io/en/latest/basics/getting-started.html.
[27] FRAUNHOFER, 2024, Linked Factory Industrial Digital Twins: Documentation of The Linkedfactory Data Formats and APIs, https://github.com/linkedfactory/specification?tab=readme-ov-file.
[28] ECLIPSE FOUNDATION, 2024, Eclipse Semantic Modelling Framwork (ESMF) Project Documentation - Semantic Aspect Meta Model (SAMM) Introduction, https://eclipse-esmf.github.io/samm-specification/snapshot/index.html.
[29] W3C, 2024, Sparql 1.1 Federated Query, https://www.w3.org/tr/sparql11-federated-query/.
[30] ITDA, 2024, Idta 02008-1-1 Time Series Data, https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/03/idta-02008-1-1_submodel_timeseriesdata.pdf.
[31] FRAUNHOFER IOSB, 2024, Edc-Extension-for-AAS, https://github.com/fraunhoferiosb/edc-extension-for-aas.
[32] ITDA, 2024, Specification of the Asset Administration Shell Part 1 Metamodel, https://industrialdigitaltwin.org/wp-content/uploads/2023/06/idta-01001-3-0_specificationassetadministrationshell_part1_metamodel.pdf.
[33] KAIROUZ P., MCMAHAN H.B., AVENT B., BELLET A., BENNIS M., BHAGOJI A.N, et al, 2021, Advances and Open Problems in Federated Learning, https://doi.org/10.48550/arXiv.1912.04977.
[34] MCMAHAN H.B., MOORE E., RAMAGE D., HAMPSON S., ARCAS B.A., 2023, Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.
[35] SATTLER F., MÜLLER K-R., SAMEK W., 2019, Clustered Federated Learning: Model-Agnostic Distributed Multi-Task Optimization under Privacy Constraints, https://doi.org/10.48550/arXiv.1910.01991.