Check for updates

# Enterprise Integration and Interoperability for Big Data-Driven Processes in the Frame of Industry 4.0

*Alexandros Bousdekis\* and Gregoris Mentzas*

Information Management Unit (IMU), School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Athens, Greece

Traditional manufacturing businesses lack the standards, skills, processes, and technologies to meet today's challenges of Industry 4.0 driven by an interconnected world. Enterprise Integration and Interoperability can ensure efficient communication among various services driven by big data. However, the data management challenges affect not only the technical implementation of software solutions but the function of the whole organization. In this paper, we bring together Enterprise Integration and Interoperability, Big Data Processing, and Industry 4.0 in order to identify synergies that have the potential to enable the so-called "Fourth Industrial Revolution." On this basis, we propose an architectural framework for designing and modeling Industry 4.0 solutions for big data-driven manufacturing operations. We demonstrate the applicability of the proposed framework through its instantiation to predictive maintenance, a manufacturing function that increasingly concerns manufacturers due to the high costs, safety issues, and complexity of its application.

Keywords: conceptual modeling, data analytics, enterprise architecture, data management, smart manufacturing, predictive maintenance

## INTRODUCTION

Enterprise integration and interoperability has been established as a scientific challenge of outmost importance (Panetto et al., 2016), especially in the frame of the emergent technologies of the Internet of Things (IoT), big data, and Artificial Intelligence (AI). This trend has inevitably affected the manufacturing domain, which is currently feeling the approach of the fourth industrial revolution and new approaches on data management and interoperability (Romero and Vernadat, 2016; Fraile et al., 2019; Zeid et al., 2019). However, the technical challenges for the transition to Industry 4.0 are strongly related to those of the whole business environment. To this end, the Enterprise Integration and Interoperability research domain should be adapted to smart manufacturing requirements in order to facilitate the design and implementation of Industry 4.0 solutions and enable the stakeholders to meet their expectations.

In this context, the increasing size of big data poses additional challenges and asks for novel techniques of software engineering for their design, analysis, and development. The existing literature has been contributing to these challenges, mainly focusing on the development of architectures for addressing lower level interoperability challenges of Industry 4.0, such as distributed storage, data aggregation, and service orchestration, as well as big data infrastructures.

On the other hand, the literature is rich on software frameworks, often overlapping or controversial to each other, something which has resulted in the design of *ad-hoc* and complex architectural big data solutions (Davoudian and Liu, 2020).

Arguably, the most widespread and influential architectural framework in the manufacturing domain is RAMI 4.0. Overall, in the literature, there are only a few case studies that follow the RAMI 4.0 model, and even fewer not requiring much effort to reach the level of practical implementation (Pisching et al., 2018). However, the key issue of any design and system development in the context of Industry 4.0 is the proper implementation of RAMI 4.0 in various manufacturing operations and the definition of appropriate sub-models for individual aspects and processes according to the technical background of Industry 4.0 (Zezulka et al., 2016; Moghaddam et al., 2018; Bousdekis et al., 2019a). To this end, there is the need for architectural frameworks that will enable the systematic design and development of Industry 4.0 solutions so that they tackle the big data-rich, complex, and uncertain manufacturing environment in a holistic way.

In this paper, we propose an architectural framework for the design and development of software solutions for big data-driven processes in Industry 4.0 in order to deal with the integration, interoperability, and data management challenges of the manufacturing environment. The proposed framework is based upon three pillars: Enterprise Integration and Interoperability, big data processing, and Industry 4.0. The synergies among them derive from the requirements that guide the design of the proposed framework. Then, the framework is instantiated to predictive maintenance and serves as the basis for the development of a predictive maintenance platform. The platform was applied to three business cases according to their requirements. In this paper, we describe its deployment and present the evaluation results of a case study from the steel industry.

The rest of the paper is organized as follows. Section Literature Review provides a literature review on the three pillars of the proposed framework, i.e., Enterprise Integration and Interoperability, big data processing, and Industry 4.0. Section The Proposed Architectural Framework for Big Data-driven Processes in Industry 4.0 describes the proposed architectural framework for big data processing in Industry 4.0. Section Application to Predictive Maintenance explains the application of the proposed framework in the context of predictive maintenance in the steel industry and presents the evaluation results. Section Conclusions and Future Work concludes the paper and presents our plans for future work.

## LITERATURE REVIEW

In this section, we present the literature review on the three pillars of the proposed architectural framework: Enterprise Integration and Interoperability (Section Enterprise Integration and Interoperability), big data processing (Section Big Data Processing), and Industry 4.0 (Section Industry 4.0). The literature review identifies the state-of-the-art in these research areas and enables the extraction of requirements for the design of the architectural framework for big data-driven processes in Industry 4.0.

## Enterprise Integration and Interoperability

Enterprise integration is the process of ensuring the interaction between enterprise entities necessary to achieve domain objectives (EN/ISO I9439, 2003), while enterprise interoperability refers to the ability of interactions (exchange of information and services) between enterprise systems (Chen et al., 2008). In this context, enterprise architecture facilitates enterprise modeling from various viewpoints and guides its implementation by providing a formal description of a system at a component level (ISO 15704, 2000; Open Group TOGAF, 2000; Bernus et al., 2003).

Since the 1980's, a lot of research has been conducted to develop enterprise architecture frameworks for enterprise integration, such as the Computer Integrated Manufacturing Open System Architecture (CIMOSA) (AMICE, 1993), the Purdue Enterprise- Reference Architecture (PERA) (Williams, 1994), the GIM architecture (Girard and Doumeingts, 2004), ARIS (Scheer, 1994), and Zachman Framework (Zachman, 1996). On top of them, the Generalized Enterprise-Reference Architecture and Methodology (GERAM) was developed (IFAC–IFIP Task Force, 1999), while additional frameworks, such as TOGAF (developed by Open Group on Architecture Framework) and C4ISR (or DoDAF) (DoDAF, 2007), as well as software engineering standards, were developed (e.g., ISO 15704, EN/ISO I9439, ISO 42010). Despite their differences, these architectures converge in three main levels of integration (Romero and Vernadat, 2016): (i) Physical Integration, which deals with systems interconnections and data exchange; (ii) Application Integration, which deals with interoperability of software applications and database systems; and (iii) Business Integration, which deals with co-ordination of functions, processes, and people.

On the other hand, several enterprise interoperability frameworks have been proposed in the literature (Chen et al., 2008), such as LISI (Levels of Information Systems Interoperability) (C4ISR, 1998), IDEAS interoperability framework (IDEAS, 2002), ATHENA interoperability framework (AIF) (ATHENA, 2003), Framework for Enterprise Interoperability (FEI), Big Data Value (BDV) Reference Model (BDVA, 2017), National Institute of Standards and Technology (NIST), and Big Data Interoperability Framework (NIST, 2019). In addition, during the last years, new domain-specific interoperability frameworks have been proposed, such as the European Interoperability Framework (EIF) (EIF, 2017), the Internet of Things-based interoperability framework for fleet management (Backman et al., 2016), the Smart City Interoperability Framework (Ahn et al., 2016), the interoperability framework for software as service systems in cloud (Rezaei et al., 2014), the International Image Interoperability Framework (Snydman et al., 2015), and the conceptual interoperability framework for large-scale systems (Selway et al., 2017). Overall, the enterprise interoperability frameworks can be seen in the frame of three main layers (Romero and Vernadat, 2016; Leal et al., 2020; Technical, Semantic, and Organizational).

## Big Data Processing

Big data is characterized by the 4Vs: Volume, Velocity, Variety, and Veracity (De Mauro et al., 2016). Volume is related to how much data is generated, velocity is related to how fast data is generated, variety is related to how many different types of data are generated, and veracity is related to how accurate data are (Chen et al., 2014; Xu and Duan, 2019). Big data analytics is classified in three main stages (Lepenioti et al., 2020): (i) descriptive analytics, identifying what has happened, examining why it happened, as well as providing real-time information about what is happening; (ii) predictive analytics, predicting what will happen and why; and (iii) prescriptive analytics, supporting decisions about what should be done and why.

The increasing size of big data poses challenges related to the complexity of big data-driven information systems and asks for novel techniques of software engineering for their design, analysis, and development (Varghese and Buyya, 2018; Xu and Duan, 2019; Davoudian and Liu, 2020). To this end, the literature has proposed asynchronous communications protocols, such as the Advanced Message Queuing Protocol (AMQP) (Vinoski, 2006) and the Message Queuing Telemetry Transport (MQTT) (MQTT, 2019), as well as scalable architectures, such as Lambda and Kappa Architecture (Davoudian and Liu, 2020). Moreover, the edge computing paradigm has emerged, aiming at addressing networking and computing challenges that cannot be met by existing cloud computing infrastructure (Trinks and Felden, 2018; Varghese and Buyya, 2018). Further, the fog computing aims at leveraging the existing computing resources on edge nodes or integrating additional computing capability between user devices and cloud data centers (Varghese and Buyya, 2018; Papageorgiou et al., 2019).

Apart from these, the recent emergence of a wide range of overlapping software frameworks in the literature, each one having a different focus, has resulted in the design of *ad-hoc* and complex architectural big data solutions (Davoudian and Liu, 2020). According to their focus and contribution, these research works can be classified into four categories: (i) empirically-grounded architectural design (Galster and Avgeriou, 2011; Angelov et al., 2012, Maier et al., 2013; Pääkkönen and Pakkala, 2015); (ii) implementation and deployment of big data systems (Schmidt and Möhring, 2013; Zimmermann et al., 2013; Salma et al., 2017); (iii) database management (Doshi et al., 2013; Zhong et al., 2013); and (iv) analytics integration (Westerlund et al., 2014; Sang et al., 2017). It should be noted that the literature is rich on domain-specific big data architectures, developed in order to address particular problems for specific application domains. For more details on the literature about software architectures, the reader may refer to Marjani et al. (2017) and to Davoudian and Liu (2020).

## Industry 4.0
### RAMI 4.0

The German Federal Ministry of Education and Research defines Industry 4.0 as "the flexibility that exists in value-creating networks by the application of Cyber Physical Systems (CPS)" (Platform Industrie 4.0, 2019). In this context, Reference Architectural Model Industrie 4.0 (RAMI 4.0) is based on a three-dimensional model consisting of the Architecture Layers, Life Cycle and Value Stream, and Hierarchy Levels dimensions, as shown in **Figure 1**. RAMI 4.0 considers any technical asset of the factory as an entity that can be represented in the digital world to conform an I4.0 component. Industry 4.0-related core topics are on the way to being standardized with a strong focus on interoperability in order to ensure networking across company and industry boundaries (Deutsches Institut für Normung, 2019; Standardization Council Industrie 4.0, 2020).

The main scope of each dimension is described below. For a more detailed introduction to RAMI 4.0, the reader may refer to Hankel and Rexroth (2015), Adolphs et al. (2015), and Deutsches Institut für Normung (2019).

**Architecture Layers**: The Architecture Layers enable the development of Industry 4.0 software solutions in a consistent way so that different and mutually dependent manufacturing operations are interconnected, taking into account the physical and the digital world. RAMI 4.0 consists of six layers: Asset layer, Integration layer, Communication layer, Information layer, Functional layer, and Business layer.

**Life Cycle and Value Stream**: The second axis in RAMI 4.0 represents the lifecycle of products and systems and is taken from the IEC 62890 standard (International Electrotechnical Commission, 2017). The product lifecycle model introduces a differentiation between product type and product instance.

**Hierarchy Levels:** The third axis of RAMI 4.0 is the hierarchical representation of the different functional levels of the factory, based on the IEC 62264 (International Electrotechnical Commission, 2016) and IEC 61512 standards. These hierarchy levels are: Connected World, Enterprise, Site, Area, Work Centers, Work Units or Station, Control Device, Field Device, and Product.

In this context, a digital twin is the container for integrating information, executing operations, and producing data describing its activity which can be in different formats, from different software tools, and not necessarily deployed in one central repository (Ganz, 2018; Catarci et al., 2019). Both the physical and the digital twins are equipped with networking devices to guarantee a seamless connection and a continuous data exchange between a generic physical system (or process) and its respective Digital Twin (Platform Industrie 4.0, 2019), while they facilitate predictions about future situations and prescriptions about production optimization (Grieves and Vickers, 2017; Zillner et al., 2018; Barricelli et al., 2019). The digital twin is implemented by the Asset Administration Shell (AAS). The AAS consists of a number of sub-models in which all the information and functionalities of a given asset—including its features, characteristics, properties, status, parameters, measurement data, and capabilities—are described (Bedenbender et al., 2017). The German Federal Ministry of Economic Affairs and Energy provides specifications for the exchange of information with the AAS (German Federal Ministry of Affairs and Energy, 2018; German Federal Ministry of Economic Affairs and Energy, 2018).
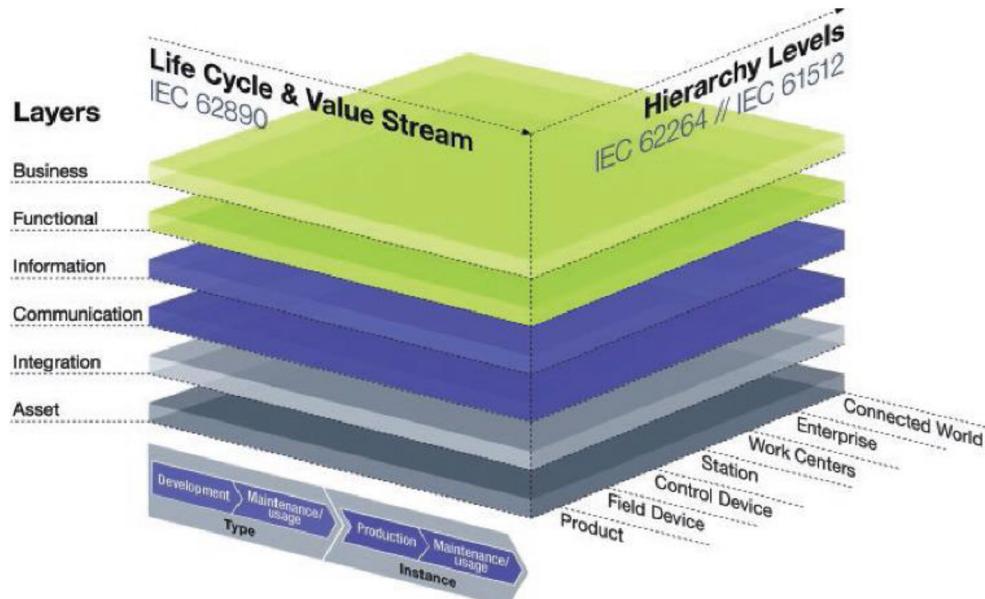
**FIGURE 1 |** The RAMI 4.0 (Source: Hankel and Rexroth, 2015).

## Other Architectural Frameworks

Although, arguably, the most widespread and influential architectural framework in the manufacturing domain is RAMI 4.0, several other collaborative paradigms have emerged. The Industrial Internet Reference Architecture (IIRA), developed by the Industrial Internet Consortium (IIC) Task group, deals with different Industrial Internet of Things (IIoT) application domains such as Energy, Healthcare, Manufacturing, Public Domain, and Transportation (Industrial Internet Consortium, 2017a). An alignment of IIRA and RAMI 4.0 has been recently developed in order to identify the complementary, contradictory, and similar aspects of these two architectural paradigms (Industrial Internet Consortium, 2017b). On top of this, BDVA presented the big data challenges in smart manufacturing and designed the BDVA SRIA 4.0, a Reference Model for data-driven mapping of BDVA Reference Model to manufacturing scenarios, taking into account RAMI 4.0 and IIRA (BDVA, 2017). Other initiatives include NIST smart manufacturing (American National Institute of Standards Technology, 2017), China's National Intelligent Manufacturing System Architecture (IMSA) (Wei et al., 2017), Made in China 2025 vision for intelligent manufacturing, and (Ministry of Industry Information Technology of China Standardization Administration of China, 2015).

In parallel, the scientific literature has been contributing to the challenges of enterprise integration and interoperability in the smart manufacturing era (Zeid et al., 2019). Existing literature mainly focuses on the development of architectures for addressing lower level interoperability challenges of Industry 4.0, such as distributed storage, data aggregation, and service orchestration (Pisching et al., 2018; Bicocchi et al., 2019; Fraile et al., 2019) as well as big data infrastructures (Pedone and Mezgár, 2018; Calabrese et al., 2020). A considerable amount of research has also focused on architectures for CPS, digital twins, and AAS (Lee et al., 2015; Bader and Maleshkova, 2019; Bousdekis et al., 2020a; Cavalieri and Salafia, 2020). For more details, the reader may refer to Moghaddam et al. (2018), Cheng et al. (2018), Fraile et al. (2019), and Zeid et al. (2019).

## THE PROPOSED ARCHITECTURAL FRAMEWORK FOR BIG DATA-DRIVEN PROCESSES IN INDUSTRY 4.0

In this section, we present the proposed architectural framework for big data-driven processes in Industry 4.0. First, we present the requirements to be fulfilled by the proposed framework (Section Requirements for the Architectural Framework). Second, we place the big data technologies and functions in the frame of RAMI 4.0 in order to assure enterprise integration and interoperability (Section Big Data Technologies and Functions in RAMI 4.0). On this basis, we provide a technical view of the 3-tier architecture in accordance with the Industry 4.0 principles and the existing big data technologies and architectures (Section Technical View of the Architecture).

## Requirements for the Architectural Framework

The literature review of Section Literature Review presented the background, as well as the main trends and challenges for each pillar of the proposed framework, i.e., Enterprise Integration and Interoperability, Big Data Processing, and Industry 4.0. Based on this analysis, we synthesized the requirements for each pillar. We concluded with 11 requirements, presented in **Table 1**. These

**TABLE 1 |** The requirements for the architectural framework.

| Enterprise integration and interoperability | |
|---|---|
| R1 | The architecture shall support physical, application, and business integration. |
| R2 | The architecture shall support technical, semantic, and organizational interoperability. |
| **Big data processing** | |
| R3 | The architecture shall support the data analytics lifecycle, i.e., descriptive, predictive, and prescriptive analytics. |
| R4 | The architecture shall support both stream processing and batch processing. |
| R5 | The architecture shall support the whole cloud continuum, i.e., edge, fog, and cloud computing. |
| R6 | The architecture shall integrate heterogeneous data sources. |
| **Industry 4.0** | |
| R7 | The architecture shall conform to RAMI 4.0. |
| R8 | The architecture shall tackle the uncertain manufacturing environment. |
| R9 | The architecture shall be compatible with digital twins. |
| R10 | The architecture shall provide various levels of insights. |
| R11 | The architecture shall be generic and scalable. |

requirements guide the design of the proposed architectural framework for enterprise integration and interoperability for big data-driven processes in the frame of Industry 4.0.

# Big Data Technologies and Functions in RAMI 4.0

## Architecture Layers

The proposed framework places the key components of a solution for big data-driven processes in the context of the RAMI 4.0 Architecture Layers, as depicted in **Figure 2**.

**Asset Layer:** Since this layer represents the reality, *Production Equipment* and *Users* are part of it. The Assets have their digital twin which is implemented with the AAS. The AAS can be applied on the level of a field device, a control device, and a station device. Moreover, the user, i.e., the Operator 4.0, can also have their "digital replica" according to the concept of "Human Digital Twin" (Bousdekis et al., 2020a).

**Integration Layer**: This layer provides information related to the assets in the appropriate format by connecting elements and people with information systems. This layer involves the *Sensors* and *Actuators* assigned to the machines as well as the *Enterprise Systems* (MES, ERP, etc.). It also includes the *Human Machine Interfaces* through which the users interact with the platform and the enterprise systems.

**Communication Layer**: Since this layer provides standardization of communication by means of uniform data format and deals with the physical support of information processing, it includes the *IoT Gateway*, the *Enterprise Data Uplifting*, and the *Message Broker*. The latter follows the AMQP or the MQTT protocol, while sensory and enterprise data are gathered with domain-specific data adapters. The machine-to-machine communication is implemented with the OPC UA protocol. In this way, data from various sources are collected for further processing in the Information Layer.

**Information Layer**: This layer provides pre-processing of events and execution of event-related rules by enabling their formal description for the interpretation of the information (Bousdekis et al., 2019a). It also manages data persistence and ensures consistent data integrity and transformation for feeding into the Functional Layer. Therefore, it includes sensor and enterprise data pre-processing in the Edge, Fog, and Cloud computational environments. *Edge Computing* contributes to this data pre-processing on the edge layer by performing edge analytics. *Fog Computing* leverages the resources on edge nodes and integrates additional computing capability along the entire data path between user devices and the cloud. *Cloud Computing* provides the infrastructure for the implementation of a big data reference software architecture. The types of data that need to be processed include real-time sensor data, historical sensor data, enterprise data, and human knowledge. In this case, the Lambda architecture is selected because it is able to incorporate stream processing and batch processing. Although the single stream processing engine of the Kappa architecture simplifies the implementation in completely data-driven computational environments, Industry 4.0 dictates that the human is an integral part of the process. Therefore, a batch processing engine is needed in order to facilitate the human-machine collaboration by embedding the expert knowledge at configuration or as soon as it becomes available. Apart from the real-time sensor data which are processed in the stream processing engine, historical sensor data and enterprise data can be processed in both engines. The information layer also includes the *Models DB*, for the storage of the simulation, context, knowledge and analytics models, the *Enterprise DB*, for the storage of the enterprise data, and the *Time-Series DB*, for the real-time data generated by equipment-installed sensors. In this way, the required data is extracted and combined accordingly in order to be available to the functions of the Functional Layer.

**Functional Layer**: This layer enables the formal description of functions and creates the platform for horizontal integration of various functions (Bousdekis et al., 2019a). It includes the run time and modeling environment for services supporting the business processes and a run time environment for applications and technical functionalities. In order to assure scalability and efficiency, the functionalities are developed as web services, following the background of SOA or EDA. Based upon the data integrity of the previous layer, the following functions are included:

- *Context Model*: It includes the definition of the manufacturing system's elements, including the assets, causes, and effects along with appropriate reactive and proactive actions. It also aims at performing knowledge representation and reasoning and, thus, it forms the basis for the enrichment of data analytics algorithms. In the current proposal, the context
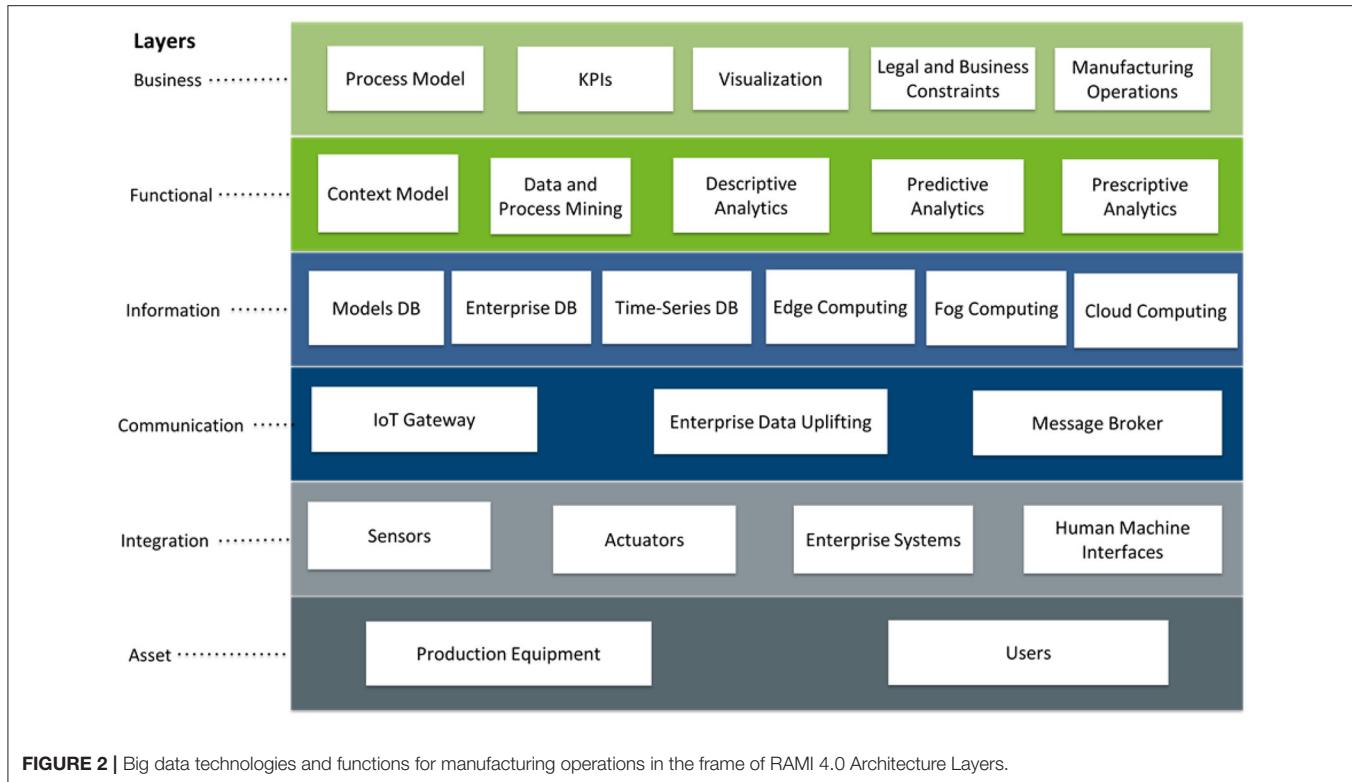
**FIGURE 2 |** Big data technologies and functions for manufacturing operations in the frame of RAMI 4.0 Architecture Layers.

model is foreseen to be implemented with Probabilistic Web Ontology Language (PR-OWL) (Carvalho et al., 2017), in order to model the Bayesian relationships among the various entities of the context model (Setiawan et al., 2019) and to achieve uncertainty representation and reasoning. The queries can be performed with RDF and SPARQL.

● *Data and Process Mining*: Data mining deals with information extraction and discovery of structures and patterns in large and complex data sets (Hand and Adams, 2014). Process mining extracts knowledge from event logs stored in the information systems in order to discover, monitor, and improve processes (Van Der Aalst et al., 2011). In the proposed approach, they facilitate the structuring and analysis of enterprise data and event logs in order to extract meaningful insights about the past performance of manufacturing business' processes and KPIs. They also enable longer-term decisions, while their results contribute to the enrichment and labeling of the descriptive, predictive, and prescriptive analytics outcomes.

● *Descriptive Analytics*: In a stream processing context, descriptive analytics deals with (deep) machine learning methods and algorithms for anomaly detection (Yue et al., 2019). In the proposed architecture, descriptive analytics is implemented as real-time algorithms on the basis of sensor-generated data streams in order to detect abnormal behaviors.

● *Predictive Analytics*: Predictive analytics are executed both offline and in real-time. It includes the development of offline predictive analytics models based on historical sensor data until a decision horizon enriched by the Data and

Process Mining and the Context Model functions. Moreover, at runtime, when the descriptive analytics function detects an abnormal behavior, predictive analytics retrieves the appropriate offline model and predicts the future situation and/or the time of its occurrence.

● *Prescriptive Analytics*: Prescriptive analytics is triggered by real-time predictions about future situations in order to generate recommendations about proactive actions and their optimal time and formulate the action plan. The prescriptive analytics models have been developed offline by taking advantage of the Data and Process Mining function as well as the Context Model.

**Business Layer**: This layer ensures the integrity of functions in the value stream and enables mapping business models and the outcomes of the overall process (Bousdekis et al., 2019a). In other words, it assures the enterprise integration and interoperability by providing the interfaces with the rest of the business functions. It takes into account the constraints, rules, and policies that affect the system operation and facilitates the interaction of manufacturing operations with the overall business environment. To do this, it provides the *Process Model* in accordance with the business models and *KPIs* defined at the strategic level, also taking into account the *Legal and Business Constraints*. In this way, it provisions both the organizational and legal interoperability. Finally, this layer provides the business interface with other *Manufacturing Operations* of the business environment as well as with the associated users through *Visualization* of results in order to achieve explainability of

the AI algorithms that were implemented in the Functional Layer and, thus, to be compatible with the "Ethics Guidelines for Trustworthy Artificial Intelligence" for the development of ethical and human-centric AI, as has been reported by the High-Level Expert Group on AI (European Commission, 2019).

## Life Cycle and Value Stream
The proposed approach facilitates the interchange of data and data analytics outcomes to all the stakeholders of the supply chain and at all stages of the manufacturing operations. This fact enables timely and reliable cooperation among them throughout the whole lifecycle of operations.

## Hierarchy Levels
The proposed approach is also compatible with the Hierarchy Levels of RAMI 4.0. The I4.0 AAS concept enables its application to the various Hierarchy Levels on the basis of a defined Asset or as a synthetics AAS, combining the AAS of various Assets at various levels. To achieve this, the AAS, which implements the Digital Twin and ranges from the Asset Layer of RAMI 4.0 to the Functional Layer, is formulated as shown in **Figure 3**. It is based on the AAS template proposed by Ye and Hong (2019) and provides a generalized and data-driven way of constructing the associated sub-models. The AAS can be extended in order to include additional asset-specific sub-models (e.g., quality, maintenance, supply chain, etc.).

## Technical View of the Architecture
Based upon the aforementioned structuring of big data technologies and functions in the frame of RAMI 4.0, we design the technical view of the proposed architecture for big data-driven processes in Industry 4.0, depicted in **Figure 4**. It shows the main interactions and the data flow among the components through the definition of end-to-end integration and communication processes and consists of three tiers: Presentation Tier, Logic Tier, and Data Tier. The technical view of the architecture drives the development procedure of associated information systems and platforms in accordance with the Industry 4.0 principles.

Below, we explain the interactions and the data flow among the components of the aforementioned architecture based upon the descriptions provided in Section Big Data Technologies and Functions in RAMI 4.0. The Logic Tier is put at the core of the description, while its functions are presented in two steps according to whether the data flow takes place at design time or at runtime:

**Design time**: The entities of the *Context Model* are defined based on the expert knowledge that is inserted through the *User Interaction* of the Presentation Tier, while their Bayesian relationships are extracted by the *Data and Process Mining* function of the Logic Tier and are communicated through RESTful APIs. The *Data and Process Mining* function retrieves data in order to produce process models, to analyze past performance, and to estimate KPIs. This data are derived from *Enterprise Systems* and are stored in the *Enterprise DB* (a NoSQL DB) of the Data Tier through *Enterprise Data Uplifting*. Its outcomes are exposed to the user through *Visualization* of the

Presentation Tier. The Context Model and the outcomes of Data and Process Mining are stored in the Models DB.

**Runtime**: The *Descriptive Analytics* function processes sensor-generated data streams in real-time in order to detect the actual performance of the *Production Equipment*. To do this, it mainly implements (deep) machine learning algorithms. This data are stored to the *Time-Series DB* and are communicated with the AMQP and the MQTT protocols. An initial pre-processing of the data may have taken place through edge analytics techniques in *Edge Computing*. These outcomes have been extracted through an *IoT Gateway*, while they have been further processed in *Fog Computing*. The *Predictive Analytics* function relies on both *Batch Processing* and *Stream Processing*. At Stream Processing, it receives streams of the Descriptive Analytics outcomes through the *Message Broker* and generates predictions about the future states of the *Production Equipment*. The predictions are generated according to the predictive analytics models that have been developed at Batch Processing based on historical sensor data enriched by the *Data and Process Mining* and the *Context Model* functions. The *Prescriptive Analytics* function receives streams of predictions through the *Message Broker* and generates prescriptions about proactive actions, i.e., actions that mitigate the impact of a future undesired event or exploit future opportunities. The Prescriptive Analytics models are enhanced by the *Data & Process Mining* and the *Context Model* functions. The outcomes of Descriptive, Predictive, and Prescriptive Analytics are exposed to the user through *Real-time Monitoring* of the Presentation Tier, while they are stored to the *Models DB* of the Data Tier.

# APPLICATION TO PREDICTIVE MAINTENANCE

In this section, we present the instantiation of the proposed architectural framework to predictive maintenance and the deployment of an associated platform to a steel industry case study. More specifically, we describe the motivation for selecting the maintenance process (Section Predictive Maintenance in Industry 4.0), we instantiate the proposed framework to the predictive maintenance context (Section Instantiation of the Proposed Architectural Framework to Predictive Maintenance), and we present the deployment of a predictive maintenance platform, developed according to the proposed framework, to a case study from the steel industry (Section Case Study in the Steel Industry).

## Predictive Maintenance in Industry 4.0
Predictive maintenance is an indispensable aspect of Industry 4.0, since it aims at achieving availability of the production equipment while avoiding unplanned downtimes with the use of condition monitoring. Predictive Maintenance in the context of Industry 4.0 is the maintenance strategy that takes advantage of the huge amounts of real-time and historical data in the enterprise ecosystem in order to detect early anomalies in equipment behaviors, to predict the future health state of the equipment and potential future failure modes, and to formulate
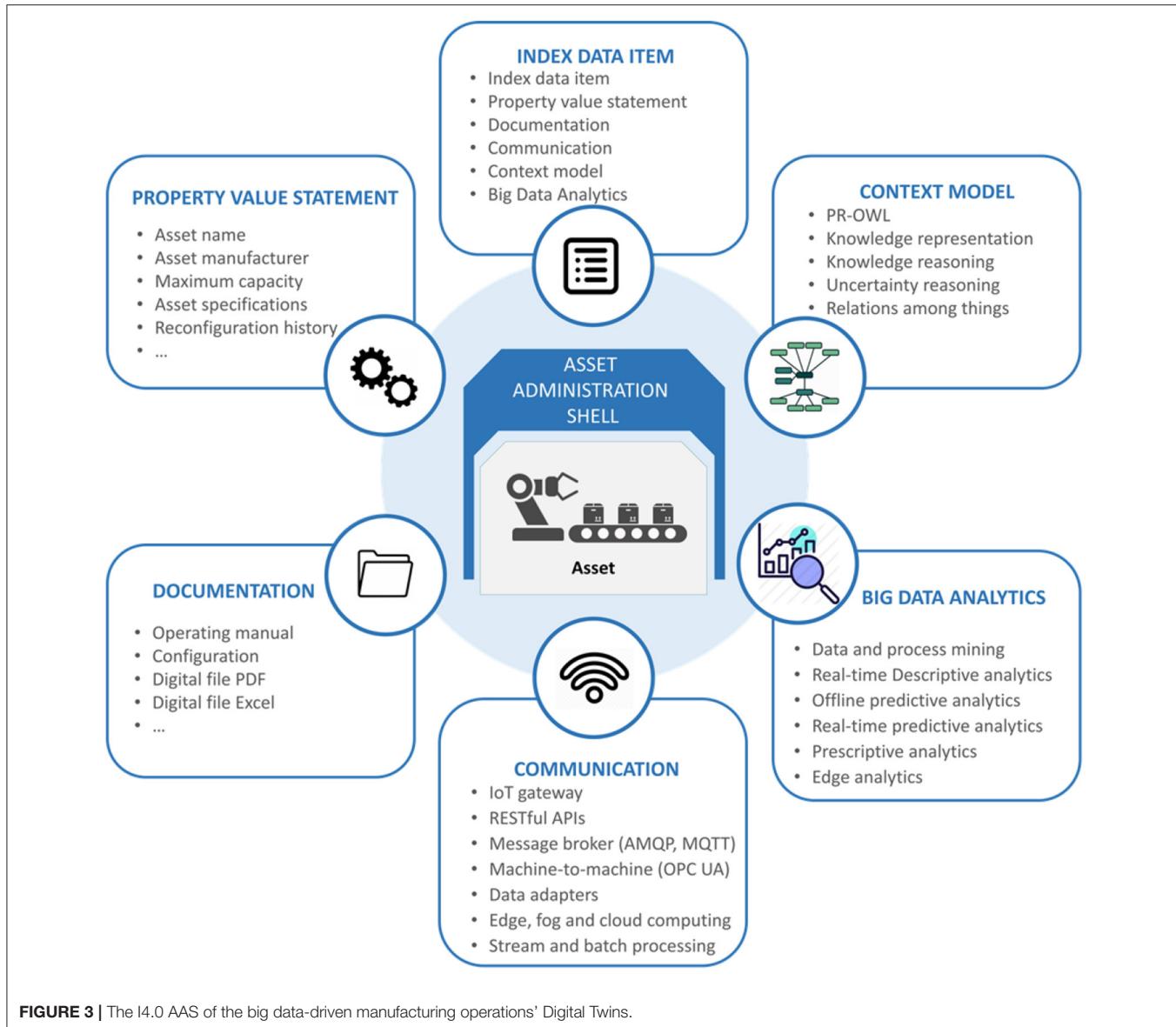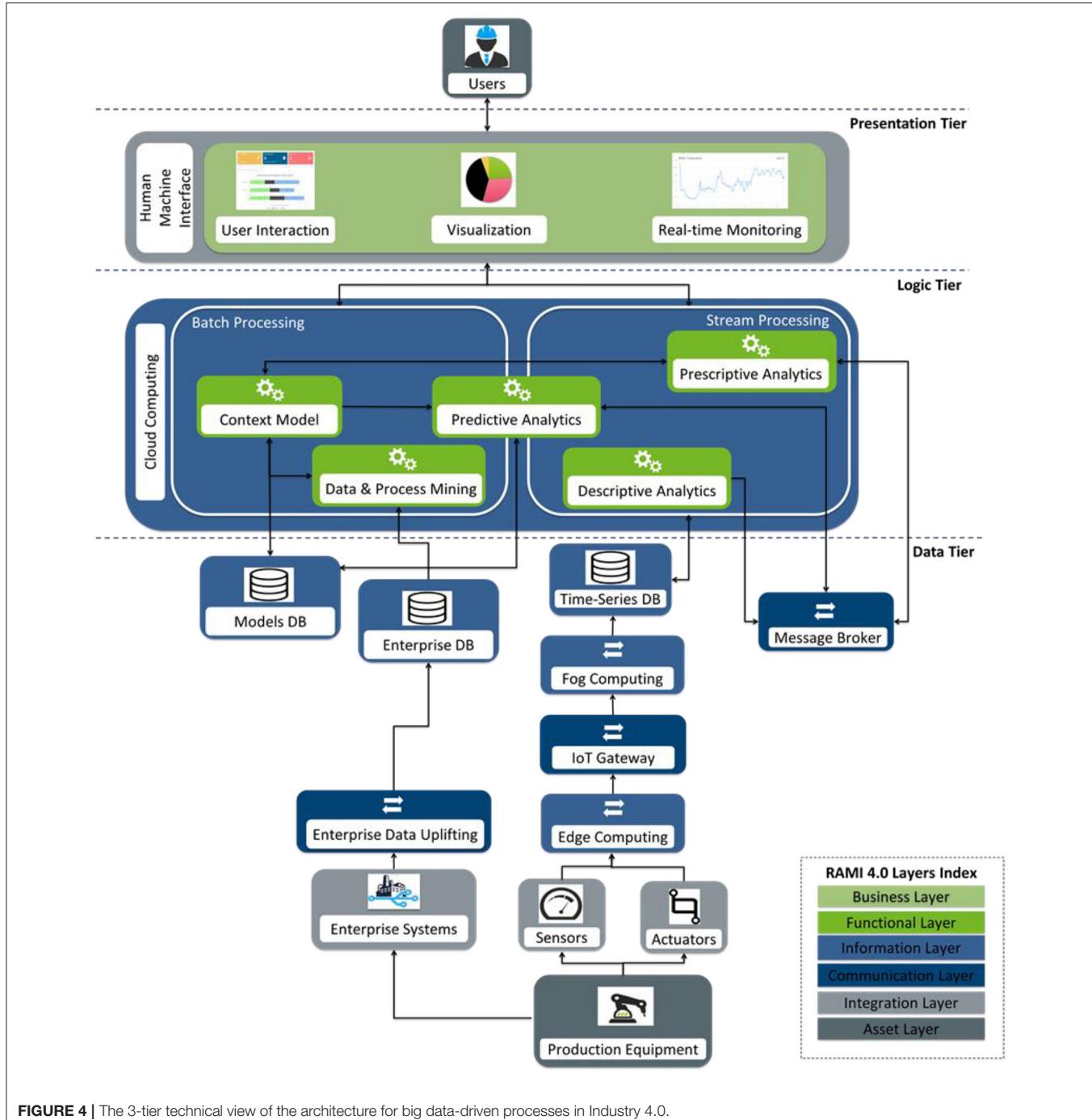
**FIGURE 3 |** The I4.0 AAS of the big data-driven manufacturing operations' Digital Twins.

proactive maintenance plans with the aim to eliminate or mitigate the impact of the predicted failures (Bousdekis et al., 2020b). Due to its importance, the German Federal Ministry of Affairs and Energy has published "The Standardization Roadmap of Predictive Maintenance for Sino-German Industrie 4.0/ Intelligent Manufacturing" (German Federal Ministry of Affairs and Energy, 2018), while the European Federation of National Maintenance Societies (EFNMS) has published the maintenance Body of Knowledge (BoK) (EFNMS, 2019).

Predictive maintenance has gathered increasing interest in both literature and practice. However, the lack of successful case studies and the development of *ad-hoc* approaches and platforms have led the manufacturers to be reluctant for its adoption (Guillén et al., 2016; Hribernik et al., 2018). The potential of predictive maintenance can be demonstrated through the concept of P-F curve, as shown in **Figure 5**. The P-F curve is a

well-established representation of asset's behavior. According to the P-F curve, the condition of an asset deteriorates over time, leading to functional failure. Therefore, the failure is considered as a process instead of an instant event. As shown in **Figure 6**, this approach provides an opportunity time window, i.e., the P-F interval, between the time of the potential failure (P), i.e., the point that it can be found out that the equipment is failing, and the functional failure (F), i.e., the point when the equipment actually fails, within which proactive decisions and actions can be taken. The point F is typically a distribution of the possible failure times for the failure mode under examination, derived from the historical data analysis. While Breakdown Maintenance is implemented after the point F and Time-Based Maintenance is scheduled at specific time intervals in order to avoid F, predictive maintenance can take advantage of the big data in order to maximize the P-F interval closer to the degradation start

**FIGURE 4 |** The 3-tier technical view of the architecture for big data-driven processes in Industry 4.0.
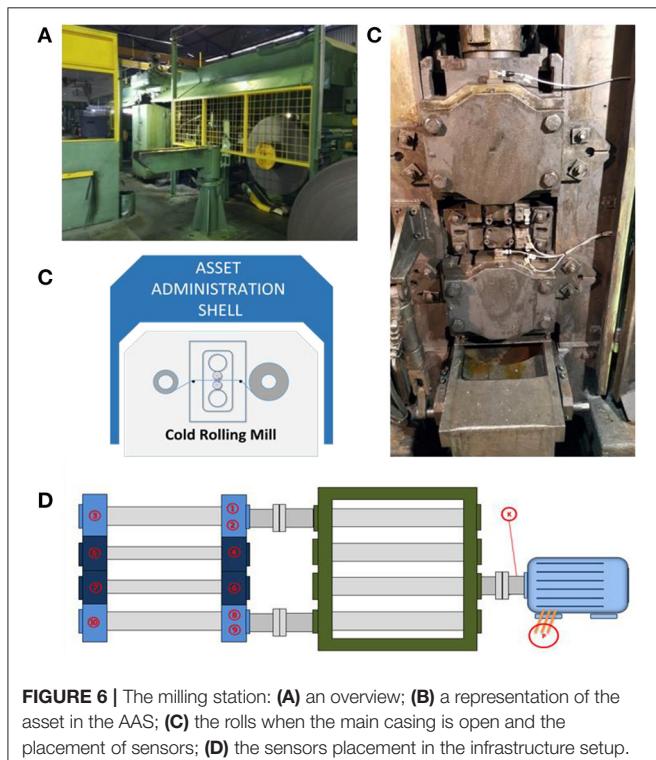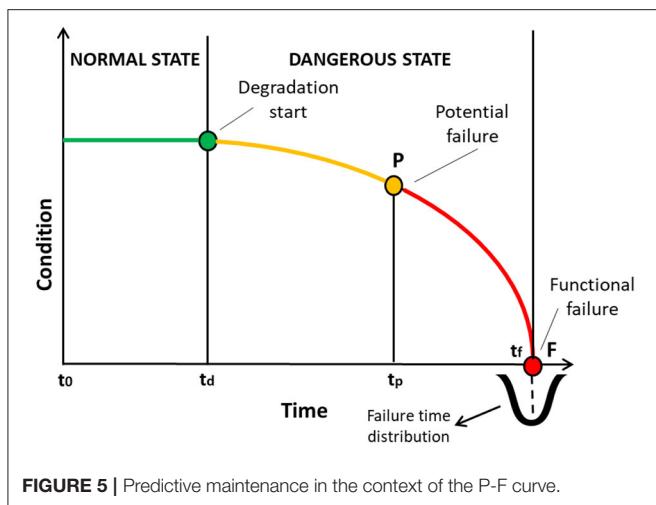
and support timely and cost-efficient decisions even before the symptoms are visible by humans.

## Instantiation of the Proposed Architectural Framework to Predictive Maintenance

The systematic representation of a predictive maintenance solution enables the reusability and knowledge transfer, an aspect of outmost importance in Industry 4.0 platforms

(Bousdekis et al., 2018a; Gröger, 2018). The proposed approach provides the capability of designing the "Predictive Maintenance Digital Twin" in order to facilitate the development of a predictive maintenance platform in the frame of Industry 4.0. Below, we illustrate how the functions of the Functional Layer are instantiated to predictive maintenance.

- *Context Model*: It includes the definition of the manufacturing system including the assets, failure causes, failure modes, and

**FIGURE 5 |** Predictive maintenance in the context of the P-F curve.



**FIGURE 6 |** The milling station: **(A)** an overview; **(B)** a representation of the asset in the AAS; **(C)** the rolls when the main casing is open and the placement of sensors; **(D)** the sensors placement in the infrastructure setup.

effects, along with appropriate reactive and proactive actions in order to create the maintenance data model. Existing related literature usually develops data models and ontologies based on the FMECA background (Zhou et al., 2015; Guillén et al., 2016; Nunez and Borsato, 2017; Ali and Hong, 2018; Hribernik et al., 2018). However, their deterministic and static nature creates obstacles to the full exploitation of big data in the frame of Industry 4.0 (Bader and Maleshkova, 2019). The PR-OWL can enable the representation of domain knowledge enhanced by data analytics in the form of uncertain relationships between the FMECA entities, e.g., failure causes, the failure

modes, and the mitigating actions, while the root causes are mapped to the available sensors that serve as indirect indicators of the failure modes.

- *Data and Process Mining*: It extracts useful insights about root causes of failures and maintenance-related business processes based upon the enterprise and operational data [e.g., Overall Equipment Effectiveness (OEE) data, Statistical Process Control (SPC), Enterprise Resource Planning (ERP), and Computerized Maintenance Management System (CMMS), maintenance event logs] in order to construct the maintenance process model. Data mining algorithms have been widely used in the related literature (Accorsi et al., 2017), while process mining is an emerging research area (dos Santos Garcia et al., 2019).
- *Descriptive Analytics*: In a predictive maintenance context, descriptive analytics is implemented as real-time diagnostic algorithms. Diagnosis aims at assessing the actual health state of the equipment and identifying abnormal behaviors based on sensor-monitored indicators of degradation, e.g., when anomalies are detected (Bousdekis et al., 2015). The literature is rich on real-time diagnostic algorithms; see e.g., Xu et al. (2017) and Li et al. (2020).
- *Predictive Analytics*: Predictive analytics is implemented as prognostic algorithms, triggered by real-time diagnostic outputs. Prognosis aims at predicting the future health state of the equipment, particularly when a failure mode is expected to occur, and estimating the Remaining Useful Life (RUL) (Bousdekis et al., 2015). This function includes the development of offline predictive analytics models based on historical sensor data that indicate the degradation process until the failure. In this sense, the models utilize information that exists in maintenance logs and in the context model, such as the time that a failure occurred, the type of the failure mode, the asset, etc. At runtime, when the descriptive analytics function detects an abnormal behavior which moves the equipment from a normal state to a deteriorating state, predictive analytics retrieves the appropriate offline model and predicts the failure mode and/or the time of the failure occurrence. Prognostic algorithms have been widely investigated in the last years; see e.g., Zonta et al. (2020).
- *Prescriptive Analytics*: Prescriptive analytics models are triggered by real-time predictions about future failure modes in order to generate recommendations about proactive actions and formulate the maintenance plan. The prescriptive analytics models have been developed offline by taking advantage of the Data and Process Mining function as well as the Context Model. Although the development of maintenance decision making algorithms is a well-established area (Ruschel et al., 2017), the sensor-driven computational environment and the need for proactive actions ahead of time (instead of reacting to incoming events) ask for novel proactive decision making methods (Bousdekis et al., 2019b).

**Table 2** provides an overview of the instantiation of the proposed architecture to predictive maintenance by addressing the Functional Layer of RAMI 4.0. More specifically, it defines the five big data functions of the proposed architecture in the

**TABLE 2 |** Instantiation of RAMI 4.0 functional layer to predictive maintenance.

| Function | Context model | Data and process mining | Descriptive analytics | Predictive analytics | Prescriptive analytics |
|---|---|---|---|---|---|
| Predictive maintenance | Maintenance data model | Maintenance process model | Diagnosis | Prognosis | Maintenance planning |
| Methods | Probabilistic ontology, belief networks, uncertainty reasoning | Knowledge discovery, statistical analysis, descriptive, predictive | Signal processing, unsupervised machine learning, deep learning | Unsupervised and supervised machine learning, deep learning | Supervised and reinforcement learning, operational research |
| Inputs | Domain knowledge Enterprise data Analytics models Process models | Enterprise data Maintenance logs | Sensor data | Current health state Historical prognostic models Context | Estimated RUL Failure PDF Context Human feedback |
| Outputs | System definition Uncertain relationships Contextual elements | KPIs Business performance Process models | Time and frequency features Current health state Early warnings | Estimated RUL Breakdown prediction Failure PDF | Proactive actions Maintenance plan Optimal time of maintenance |
| Data processing | Batch processing | Batch processing | Stream processing | Stream processing (batch processing) | Stream processing |
| Data storage | OWL 2 RL profile Triplestore NoSQL | Time-series DB OWL 2 RL profile NoSQL | Time-series DB | Time-Series DB OWL 2 RL profile NoSQL | Time-series DB OWL 2 RL profile NoSQL |
| Communication protocol | RESTful APIs | RESTful APIs | AMQP | AMQP RESTful APIs | AMQP RESTful APIs |

frame of predictive maintenance along with related categories of methods and their inputs and outputs. Moreover, it defined the data processing approach, the data storage high-level specifications, and the appropriate communication protocols for each function.

## Case Study in the Steel Industry

In this section, we demonstrate a predictive maintenance platform that was developed in accordance with the proposed architectural framework for big data-driven processes in Industry 4.0 and based upon its instantiation to predictive maintenance, as presented in Section Instantiation of the Proposed Architectural Framework to Predictive Maintenance. It has been applied in three use cases from different manufacturing sectors: the steel industry, domestic appliances production, and the aerospace industry. The following demonstration deals with the steel industry.

### The Manufacturing Process and Equipment

The case under examination is the cold rolling process. Cold rolling is a manufacturing process of metal deformation involving a pair of rotating metal rolls aiming at reducing the cross-sectional area or shaping a metal piece below its recrystallization temperature. The main components of the milling station are:

- The work rolls: a pair of rolls responsible for the actual milling until the required width is achieved.
- The backup rolls: a pair of rolls transmitting motion to the working rolls.
- The motor unit: the component supporting the rotation of the backup rolls.

**TABLE 3 |** The installed accelerometers.

| Sensor ID | Measurement point | Sensor direction |
|---|---|---|
| 1 | Upper backup roll – DE side | Vertical |
| 2 | Upper backup roll – DE side | Axial |
| 3 | Upper backup roll – NDE side | Vertical |
| 4 | Upper working roll – DE side | Reverse horizontal |
| 5 | Upper working roll – NDE side | Horizontal |
| 6 | Down working roll – DE side | Reverse horizontal |
| 7 | Down working roll – NDE side | Horizontal |
| 8 | Down backup roll – DE side | Vertical |
| 9 | Down backup roll – DE side | Axial |
| 10 | Down backup roll – NDE side | Vertical |

**Figure 6A** shows the milling station on the shop floor; **Figure 6B** illustrates the cold rolling manufacturing process. **Figure 6C** depicts the work and the backup rolls. During the operation, the rolls are continuously being sprayed by soap oil in order to reduce heat and friction. **Figure 6D** depicts the infrastructure setup for sensor data acquisition and the placement of the accelerometers. Their description is presented in **Table 3**.

Before the installation of sensors, maintenance was performed on a time-based mode. The rolls were replaced every 8 h (i.e., when there was a shift change) regardless of their health state. According to their condition, identified with visual inspection, the removed rolls were subject either to repair or they were sent to waste. In this way, on the one hand, replacement took place even if it was not necessary and, on the other hand, unexpected failures occurred between successive replacements. This fact led to high maintenance costs, despite the improvements of the last years

**TABLE 4 |** Summary of the streaming dataset.

| | |
|---|---|
| Dataset title | Roller vibrations |
| Origin | Milling station |
| Sensor type | Accelerometer |
| Physical world measurements | Acceleration, velocity, shock, and overall bearing |
| Sensor reporting frequency | 10 readings per minute (configurable) |
| Data stream rate | ~8 kb per minute |
| Sensor input signal(s) | Mechanical |
| Data type | Acceleration, velocity, bearing: Float<br>Shock: Integer |
| Interfaces to obtain sensor readings | PLC TCP connection |

**TABLE 5 |** The CMMS dataset.

| | |
|---|---|
| Dataset title | Operational and legacy |
| Data type | **Roll data**<br>Roll position: upper roll/down roll<br>Roll ID<br>Roll past operation<br>Input characteristics: timestamp, roll diameter Output characteristics: timestamp, reason of replacement (type of failure mode/planned maintenance)<br>**Production data**<br>KPIs: number of failures, time of planned maintenance, OEE, MTBF, MTTR |
| Interface | Data uplifting or API |

**TABLE 6 |** Technology stack.

| | |
|---|---|
| **Presentation tier** | |
| User interaction | Thymeleaf |
| Visualization | Grafana, Kibana, Thymeleaf |
| Real-time monitoring | Grafana, Kibana, Thymeleaf |
| **Logic tier** | |
| Context model | Spring Boot framework, Apache Maven |
| Data and process mining | Django Web Framework, django-rest-framework, Elasticsearch, Jupyter, Pandas, scikit-learn, numPy, somPy, Keras, pm4Py |
| Descriptive analytics | Spring Boot framework, Apache Commons, Netlib, Common Math, MOA: Massive Online Analysis, MathParser Org MXparser, Elasticsearch |
| Predictive analytics | Spring Boot framework, Apache Commons, Weka, Common Math, MOA: Massive Online Analysis, Netlib |
| Prescriptive analytics | Apache Maven, Spring Boot framework, Drools, BURLAP, Common Math, Netlib |
| **Data tier** | |
| Models DB | MongoDB, MySQL, MariaDB |
| Enterprise DB | MongoDB, PostgreSQL |
| Time-series DB | InfluxDB |
| Message broker | Apache Kafka |

due to the adoption of a Total Productive Maintenance (TPM) management program.

## The Datasets

The data used in this case were derived from sensors and a CMMS, while the required expert knowledge was embedded through a Graphical User Interface (GUI). The sensor infrastructure consists of 10 Accelerometers collecting vibration data which are gathered in an MVX which are then transmitted *via* Modbus TCP to a Siemens S7-1500 PLC. The values are exposed from the PLC to the DB port and can thus be collected *via* external modules that have access to the PLC *via* the network. An adapter was developed in order to sample the DB Port every 5 ms – 5 s. The data are then processed *via* a Storm-Kafka pipeline and are stored and retrieved into the *Time-series D*. This pipeline is responsible for performing normalization procedures before the data are pushed to the Logic Tier. Normalization is also configurable and can be adjusted by attaching new Storm Bolts. A set of Bolts for rounding continuous variables has been deployed. The summary of the streaming dataset derived from the aforementioned accelerometers is presented in **Table 4**. In addition, there is a CMMS which stores operational and enterprise data and performs basic calculations about OEE, time-based maintenance plan, failure modes, downtime, etc. The summary of the CMMS dataset used in this case is presented in **Table 5**.

## Implementation of the Architecture

The core technology stack of the predictive maintenance platform which implements the 3-tier architecture for big data-driven processes in Industry 4.0 is shown in **Table 6**. The platform embeds various algorithms for each function, as shown in **Table 7**. Some of these algorithms have been developed within the platform while others are retrieved through APIs from open source data analytics tools. In this way, the platform is able to tackle a wide variety of cases, assets, and degradation behaviors, while it is extensible in order to embed more algorithms.

At design time, the *Context Model* represents and stores probabilistic relationships among failure causes, failure modes, and mitigating actions. It is in accordance with MIMOSA OSA-CBM data model, while it introduces the capability of uncertain relationships among the entities according to PR-OWL. The *Data and Process Mining* function creates the maintenance process model and enriches the offline predictive analytics models. Moreover, it feeds into the context model with the cost models of the failure modes and the mitigating actions. At runtime, the platform provides real-time monitoring of the vibration and ensures that the gathered data at the on-site PLC are transmitted

| Data and process mining | Linear Regression, Bayesian Networks, Self-Organizing Map (SOM), K-means clustering, Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), Inductive Miner, Fuzzy Miner |
|---|---|
| Descriptive analytics | Feature Extraction, k-Nearest Neighbor, association rules, online Bayesian changepoint detection |
| Predictive analytics | Logistic Regression, Exponential fitting, Weibull fitting, Hidden Markov Model (HMM) |
| Prescriptive analytics | Association rules, Bayesian Networks, Markov Decision Process, Reinforcement Learning |

through the communication channel. The acquired data feed into the stream processing functionalities, i.e., Descriptive, Predictive, and Prescriptive Analytics, which implement various data fusion, (deep) machine learning, and optimization algorithms, which are configurable according to the specific use case requirements and complexity.

## Illustrative Scenario

As already mentioned, the platform embeds various algorithms in order to deal with the complexity and the requirements of the complex manufacturing environment, e.g., various assets, different degradation behaviors, different specifications, etc. In order to demonstrate a typical information flow across the aforementioned functions, we herein present an instance of the platform in the case study as an illustrative scenario. The scenario in the case study under examination is depicted in **Figure 7**.

In this scenario, *Descriptive Analytics* performs feature extraction with rolling kurtosis and online Bayesian changepoint detection (Wen et al., 2018) in order to estimate in real-time the log likelihood of having a changepoint from the normal state to the dangerous state of equipment. In **Figure 7**, it identifies a changepoint 385 min after the setup of the machine. This outcome triggers the *Predictive Analytics* service in order to provide a prediction about the failure mode occurrence by retrieving the most similar cluster that corresponds to a failure mode and applying Weibull fitting (Zhang et al., 2018). Therefore, it predicts that the failure mode "Broken Roll" will occur in 115 min. This prediction feeds into *Prescriptive Analytics* which recommends the optimal proactive action, out of the alternative actions for this failure mode, along with the optimal time of its application. In the current scenario, the prescriptive analytics model has been developed as a Markov Decision Process (MDP) (Bousdekis et al., 2018b) and there are three alternative maintenance actions: lower the speed of the machine, replace the rolls, and perform full maintenance on the equipment. As shown in **Figure 7**, the optimal action is to replace the rolls in 51 min, because at that time, the expected loss is minimized.

## Evaluation Results

We evaluated the value of the proposed approach by performing a before/after analysis of KPIs retrieved by the CMMS. The comparison was performed on the basis of 2 complete years. In the 1st year of the evaluation period ("before"), the company performed time-based maintenance by replacing the rolls every 8 h, according to the supplier's specifications. As a result, either the rolls were replaced in their normal state or unexpected breakdowns took place in this time interval. In the 2nd year, the aforementioned platform had been deployed and the company started performing predictive maintenance of the rolls. In this way, not only the decisions about the roll replacement were taken dynamically, but also it adopted imperfect maintenance actions (e.g., lower the speed of the mill, increase the soap oil to eliminate friction, optimal utilization of repaired rolls, etc.) that extend the lifetime of the equipment when downtime is not acceptable (e.g., when customers' demands need to be met).

The evaluation focused on the aforementioned milling station and not on the whole production process in order to eliminate other factors that may affect the KPIs' values. It is expected that the effect of the proposed approach can be multiplied if it is applied to the whole factory and for a longer period of time. The results are shown in **Table 8**. We also interviewed key persons from the company, such as the General Manager, the Quality Manager, the Production Manager, and the operators of this milling station. According to the results, the number of failures was decreased by 47.69%, the time for performing planned (time-based) maintenance was decreased by 62.5%, the OEE was increased by 5.03%, the Mean Time Between Failures (MTBF) was increased by 22.90%, and the Mean Time To Repair (MTTR) was decreased by 21.88%.

## Lessons Learned

In this section, we summarize the lessons learned from all three use cases (steel industry, domestic appliances production, and aerospace industry), mainly related to the enterprise integration and interoperability challenges. In this sense, we emphasize the challenges with respect to the complexity of applying the proposed framework and to the integration with the manufacturing environment, information systems, and measuring devices. Below, we discuss the main lessons learned.

**Combination of process knowledge and data analytics**: On a business level, collaboration and communication between domain experts and data analysts is a challenging task. On the one hand, the data analysts need to understand the manufacturing process, it potential and constraints, as well as the business requirements in order to decide on the functionalities, the algorithms, and the configuration of the platform to the specific business needs. On the other hand, the domain experts need to understand the system and technical requirements that may lead, not only to new investments on information systems and sensor infrastructure, but also to a disruptive way of thinking and practice.

**Project management**: The adoption of disruptive technologies in the frame of Industry 4.0 needs efficient project management. Such projects usually last for a long period of time in order to tackle the large variety of integration and data
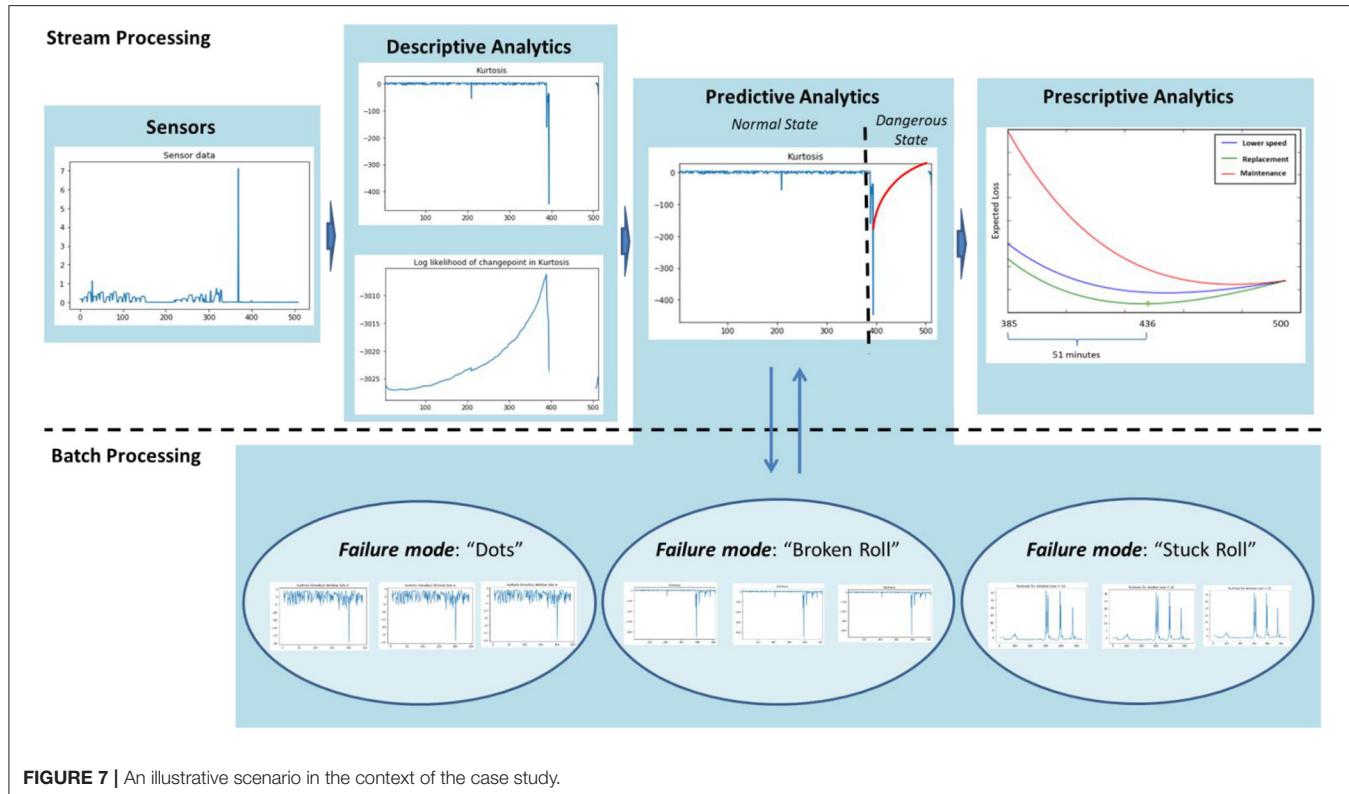
**FIGURE 7 |** An illustrative scenario in the context of the case study.

**TABLE 8 |** Evaluation results of KPIs in the steel industry.

| *KPI* | Number of failures | Planned maintenance duration | OEE | MTBF | MTTR |
|---|---|---|---|---|---|
| Before | 96 | 2,340 min | 59.43% | 2,135 min | 117 min |
| After | 65 | 1,440 min | 64.46% | 2,772 min | 96 min |
| Improvement | *47.69%* | *62.5%* | *5.03%* | *22.90%* | *21.88%* |

management challenges and to sufficiently assess the efficiency of the software solutions. Long-lasting software maintenance and support is also an important aspect.

**Data privacy and security:** The manufacturing data is highly confidential for the manufacturers, since they are critical to their processes. Moreover, the increasing use of sensors and actuators on the shop floor makes cybersecurity a topic of outmost importance for the robustness of operations and the safety of operators. For these reasons, manufacturing companies are usually reluctant to open-source platforms "intervening" with their legacy systems and sensor infrastructure and externally processing the data. Such platforms should use state-of-the-art technologies and mechanisms that ensure data security, while a close collaboration with the enterprise systems and sensors is essential.

**Utilization of heterogeneous data sources:** The manufacturing environment includes various and heterogeneous data sources that have the potential to provide insights on various aspects of the processes. This heterogeneity is caused by, among others, the co-existence of old-fashioned systems and disruptive

Industry 4.0 technologies. Therefore, an Industry 4.0 platform needs to take advantage of all the data sources which may include sensors, actuators, legacy and operational systems, enterprise systems, and Excel files, but also expert knowledge. The identification of all the available data sources, the development of appropriate interfaces, and the implementation of the right algorithms are among the main challenges in the deployment of the proposed solution.

**Integration to the legacy and operational systems**: Legacy and operational systems are usually proprietary solutions, something which poses additional challenges to the integration of open-source platforms. Apart from the close collaboration with the provider, it is important to define the required data formats, and process the data in order to enrich the context model and the data analytics algorithms. In this sense, the database management of the proposed solution is crucial for the storage and retrieval according to the consuming functions. When this is not possible, legacy data uplifting can be applied.

**Integration to the sensor infrastructure**: Integration to the sensor infrastructure also requires a close collaboration with the

provider in order to develop adapters capable of extracting the data at a pre-configured sampling time. The adapters need to be configured in order to sample the DB port at appropriate times according to the specific process (e.g., frequency of events, criticality of operations, time constraints, etc.). They should ensure data security.

## CONCLUSIONS AND FUTURE WORK

Traditional manufacturing businesses lack the standards, skills, processes, and technologies to meet today's challenges of Industry 4.0 driven by an interconnected world. Enterprise Integration and Interoperability can ensure efficient communication among various services in alignment with the business needs and requirements. However, the data management challenges affect not only the technical implementation of software solutions but the function of the whole organization. A key issue in Industry 4.0 is the effective application of the Reference Architecture Model Industrie (RAMI) 4.0 in various manufacturing operations. In this paper, we bring together Enterprise Integration and Interoperability, Big Data Processing, and Industry 4.0 in order to identify synergies that have the potential to enable the so-called "Fourth Industrial Revolution." On this basis, we propose an architectural framework for designing and modeling Industry 4.0 solutions for big data-driven manufacturing operations. We demonstrate the applicability of the proposed framework through its instantiation to predictive maintenance, a manufacturing function that increasingly concerns manufacturers due to the high costs, the safety issues, and the complexity of its application. The proposed approach achieved to exploit the full potential of predictive maintenance in a case study from the steel industry, since it provides a systematic way of designing the maintenance operations and developing a software platform. At the same time, the developed solution can be seen in the context of the whole enterprise architecture, according to the digital manufacturing strategy, in order to balance the wide-ranging—vertical and horizontal—effects within the organization. The effect of such a solution is strongly affected by the data availability (quality and quantity) and the algorithms suitability; however, the enterprise integration and interoperability in the frame of Industry 4.0 is an area usually underestimated.

Our future work will move toward four main directions. First, we will apply the proposed architectural framework for the development of software platforms for additional manufacturing operations. In particular, we will prioritize its application to quality processes in the frame of the predictive quality strategy. Second, we will focus on the user interaction aiming at achieving an optimized human-machine collaboration through explainable AI and digital intelligent assistants. Third, we will provide a taxonomy of appropriate technologies for each part of the architecture in order to facilitate the software implementation. Fourth, we will design a holistic business process management view of all the main manufacturing operations in order to further assure the organizational interoperability for the transition to Industry 4.0.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

The review of Enterprise Integration and Interoperability, Big Data Processing, and Industry 4.0 as well as the first version of the proposed architectural framework and its instantiation to predictive maintenance was performed by AB. GM supervised and guided this research work, refined the proposed architectural framework, and made improvements to the paper. All authors contributed to the article and approved the submitted version.

## FUNDING

## REFERENCES

Accorsi, R., Manzini, R., Pascarella, P., Patella, M., and Sassi, S. (2017). Data mining and machine learning for condition-based maintenance. *Proc. Manufacturing* 11, 1153–1161. doi: 10.1016/j.promfg.2017.07.239

Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., et al. (2015). *Reference Architecture Model Industrie 4.0 (rami4. 0). ZVEI and VDI, Status Report*. Frankfurt: ZVEI.

Ahn, J., Lee, J. S., Kim, H. J., and Hwang, D. J. (2016). "Smart City Interoperability Framework Based on City Infrastructure Model and Service Prioritization," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)* (Vienna: IEEE), 337–342.

Ali, N., and Hong, J. E. (2018). Failure detection and prevention for cyber-physical systems using ontology-based knowledge base. *Computers* 7:4. doi: 10.3390/computers7040068

American National Institute of Standards and Technology (2017). *Software Requirements Specification to Distribute Manufacturing Data*. Gaithersburg, MD: American National Institute of Standards and Technology.

AMICE (1993). *CIMOSA—Open System Architecture for CIM, 2nd Edn*. Verlag, Berlin: Springer.

Angelov, S., Grefen, P., and Greefhorst, D. (2012). A framework for analysis and design of software reference architectures. *Inf. Sof. Tech.* 54, 417–431. doi: 10.1016/j.infsof.2011.11.009

ATHENA (2003). *Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Applications, FP6-2002-IST1, Integrated Project*. Walldorf: FP6-IST.

Backman, J., Väre, J., Främling, K., Madhikermi, M., and Nykänen, O. (2016). "IoT-based interoperability framework for asset and fleet management," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)* (Berlin: IEEE), 1–4. doi: 10.1109/ETFA.2016.7733680

Bader, S. R., and Maleshkova, M. (2019). "The semantic asset administration shell," in *International Conference on Semantic Systems* (Cham: Springer), 159–174. doi: 10.1007/978-3-030-33220-4_12

Barricelli, B. R., Casiraghi, E., and Fogli, D. (2019). A survey on digital twin: definitions, characteristics, applications, and design implications. *IEEE Access* 7:2953499. doi: 10.1109/ACCESS.2019.2953499

BDVA (2017). *BDV SRIA: European Big Data Value Strategic Research and Innovation Agenda*. Available online at: https://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf (December 12, 2020).

Bedenbender, H., Billmann, M., Epple, U., Hadlich, T., Hankel, M., Heidel, R., et al. (2017). *Examples of the Asset Administration Shell for Industrie 4.0 Components–Basic Part. ZVEI White Paper*. Frankfurt: ZVEI.

Bernus, P., Nemes, L., and Schmidt, G. (2003). *Handbook on Enterprise Architecture*. Berlin: Springer-Verlag. doi: 10.1007/978-3-540-24744-9

Bicocchi, N., Cabri, G., Mandreoli, F., and Mecella, M. (2019). Dynamic digital factories for agile supply chains: an architectural approach. *J. Indus. Inform. Integr.* 15, 111–121. doi: 10.1016/j.jii.2019.02.001

Bousdekis, A., Apostolou, D., and Mentzas, G. (2020a). A human cyber physical system framework for operator 4.0–artificial intelligence symbiosis. *Manuf. Let.* 25, 10–15. doi: 10.1016/j.mfglet.2020.06.001

Bousdekis, A., Apostolou, D., and Mentzas, G. (2020b). Predictive maintenance in the 4th industrial revolution: benefits, business opportunities, and managerial implications. *IEEE Eng. Manag. Rev.* 48, 57–62. doi: 10.1109/EMR.2019.2958037

Bousdekis, A., Lepenioti, K., Apostolou, D., and Mentzas, G. (2019b). Decision making in predictive maintenance: literature review and research agenda for industry 4.0. *IFAC PapersOnLine* 52, 607–612. doi: 10.1016/j.ifacol.2019.11.226

Bousdekis, A., Lepenioti, K., Ntalaperas, D., Vergeti, D., Apostolou, D., and Boursinos, V. (2019a). "A RAMI 4.0 view of predictive maintenance: software architecture, platform and case study in steel industry," in *International Conference on Advanced Information Systems Engineering* (Cham: Springer), 95–106. doi: 10.1007/978-3-030-20948-3_9

Bousdekis, A., Magoutas, B., Apostolou, D., and Mentzas, G. (2015). A proactive decision making framework for condition-based maintenance. *Ind. Man. Data Syst.* 115, 1225–1250. doi: 10.1108/IMDS-03-2015-0071

Bousdekis, A., Magoutas, B., Apostolou, D., and Mentzas, G. (2018a). Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *J. Intell. Manufact.* 29, 1303–1316. doi: 10.1007/s10845-015-1179-5

Bousdekis, A., Papageorgiou, N., Magoutas, B., Apostolou, D., and Mentzas, G. (2018b). Enabling condition-based maintenance decisions with proactive event-driven computing. *Comput. Indus.* 100, 173–183. doi: 10.1016/j.compind.2018.04.019

C4ISR (1998). *Levels of Information Systems Interoperability (LISI)*. Washington, DC: Architecture Working Group.

Calabrese, M., Cimmino, M., Fiume, F., Manfrin, M., Romeo, L., Ceccacci, S., et al. (2020). SOPHIA: an event-based IoT and machine learning architecture for predictive maintenance in industry 4.0. *Information* 11:4. doi: 10.3390/info11040202

Carvalho, R. N., Laskey, K. B., and Costa, P. C. (2017). PR-OWL–a language for defining probabilistic ontologies. *Int. J. Approx. Reas.* 91, 56–79. doi: 10.1016/j.ijar.2017.08.011

Catarci, T., Firmani, D., Leotta, F., Mandreoli, F., Mecella, M., and Sapio, F. (2019). "A conceptual architecture and model for smart manufacturing relying on service-based digital twins," in *2019 IEEE International Conference on Web Services (ICWS)* (IEEE), 229–236. doi: 10.1109/ICWS.2019.00047

Cavalieri, S., and Salafia, M. G. (2020). A model for predictive maintenance based on asset administration shell. *Sensors* 20:21. doi: 10.3390/s20216028

Chen, D., Doumeingts, G., and Vernadat, F. (2008). Architectures for enterprise integration and interoperability: past, present and future. *Comp. Ind.* 59, 647–659. doi: 10.1016/j.compind.2007.12.016

Chen, M., Mao, S., and Liu, Y. (2014). Big data: a survey. *Mob. Netw. Appl.* 19, 171–209. doi: 10.1007/s11036-013-0489-0

Cheng, J., Chen, W., Tao, F., and Lin, C. L. (2018). Industrial IoT in 5G environment towards smart manufacturing. *J. Indus. Inform. Integr.* 10, 10–19. doi: 10.1016/j.jiii.2018.04.001

Davoudian, A., and Liu, M. (2020). Big data systems: a software engineering perspective. *ACM Comp. Surv.* 53, 1–39. doi: 10.1145/3408314

De Mauro, A., Greco, M., and Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Libr. Rev.* 65, 122–135. doi: 10.1108/LR-06-2015-0061

Deutsches Institut für Normung (2019). *Reference Architecture Model Industrie 4.0 (RAMI 4.0) English Translation of DIN SPEC 91345:2016-04*. Berlin: Deutsches Institut für Normung.

DoDAF (2007). *Department of Defense Architecture Framework, Version 1.5. Tech. Rep*. Washington, DC: Department of Defense.

dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., et al. (2019). Process mining techniques and applications-a systematic mapping study. *Expert Syst. Appl.* 133, 260–295. doi: 10.1016/j.eswa.2019.05.003

Doshi, K. A., Zhong, T., Lu, Z., Tang, X., Lou, T., and Deng, G. (2013). "Blending SQL and NewSQL approaches: reference architectures for enterprise big data challenges," in *2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (Beijing: IEEE), 163–170. doi: 10.1109/CyberC.2013.34

EFNMS (2019). *Body of Knowledge (BoK)*. Available online at: https://www.efnms.eu/about-us/bok/ (December 12, 2020).

EIF (2017). *European Interoperability Framework - Implementation Strategy. Annex II of to the Communication From the Commission to the European Parliament, the Council*. Brussels: European Economic and Social Committee and the Committee of the Regions.

EN/ISO I9439 (2003). *Enterprise Integration—Framework for Enterprise Modelling*. Geneva: EN/ISO.

European Commission (2019). *Ethics Guidelines for Trustworthy AI*. Available online at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed December 03, 2020).

Fraile, F., Sanchis, R., Poler, R., and Ortiz, A. (2019). Reference models for digital manufacturing platforms. *Appl. Sci.* 9:20. doi: 10.3390/app9204433

Galster, M., and Avgeriou, P. (2011). "Empirically-grounded reference architectures: a proposal," in *Proceedings of the joint ACM SIGSOFT conference–QoSA and ACM SIGSOFT symposium–ISARCS on Quality of software architectures–QoSA and architecting critical systems—ISARCS* (Boulder, CO), 153–158. doi: 10.1145/2000259.2000285

Ganz, C. (2018). *Digital Twin – Virtually Identical? ABB Review 02/2018*, 94–95.

German Federal Ministry of Affairs and Energy (2018). *The Standardisation Roadmap of Predictive Maintenance for Sino-German Industrie 4.0/ Intelligent Manufacturing. Sino-German Industrie 4.0/Intelligent Manufacturing Standardisation Sub-Working Group*. Berlin: Federal Ministry of Economics and Technology.

German Federal Ministry of Economic Affairs and Energy (2018). *Details of the Administration Shell: The Exchange of Information between the Partners in the Value Chain of Industrie 4.0 (Version 1.0)*. Berlin: Federal Ministry of Economics and Technology.

Girard, P., and Doumeingts, G. (2004). GRAI-Engineering: a method to model, design and run engineering design departments. *Int. J. Comp. Integr. Manufact.* 17, 716–732. doi: 10.1080/0951192042000237492

Grieves, M., and Vickers, J. (2017). "Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems* (Cham: Springer), 85–113. doi: 10.1007/978-3-319-38756-7_4

Gröger, C. (2018). Building an industry 4.0 analytics platform. *Datenbank-Spektrum* 18, 5–14. doi: 10.1007/s13222-018-0273-1

Guillén, A. J., Crespo, A., Gómez, J. F., and Sanz, M. D. (2016). A framework for effective manage-ment of condition based maintenance programs in the context of industrial development of E-Maintenance strategies. *Comp. Ind.* 82, 170–185. doi: 10.1016/j.compind.2016.07.003

Hand, D. J., and Adams, N. M. (2014). Data mining. *Wiley StatsRef Statist. Reference Online* 2014, 1–7. doi: 10.1002/9781118445112.stat06466

Hankel, M., and Rexroth, B. (2015). *The Reference Architectural Model Industrie 4.0 (rami 4.0). ZVEI*, 2, 4. Frankfurt: ZVEI.

Hribernik, K., von Stietencron, M., Bousdekis, A., Bredehorst, B., Mentzas, G., and Thoben, K. D. (2018). Towards a unified predictive maintenance system- a use case in production logistics in aeronautics. *Proc. Manufact.* 16, 131–138. doi: 10.1016/j.promfg.2018.10.168

IDEAS (2002). *Thematic Network, IDEAS: Interoperability Development for Enterprise Application and Software—Roadmaps, Annex 1—Do,W* (Miami, FL).

IFAC–IFIP Task Force (1999). *GERAM: Generalized Enterprise Reference Architecture and Methodology, Version 1.6.3, IFAC–IFIP Task Force on Architecture for Enterprise Integration* (Berlin).

Industrial Internet Consortium (2017a). *The Industrial Internet of Things Volume G1: Reference Architecture*. Needham, MA: Industrial Internet Consortium.

Industrial Internet Consortium (2017b). *Architecture Alignment and Interoperability: An Industrial Internet Consortium and Platform Industrie 4.0 Joint Whitepaper*. Needham, MA: Industrial Internet Consortium.

International Electrotechnical Commission (2016). *Enterprise-Control System Integration-Part 3. Activity Models of Manufacturing Operations Management*. Geneva: IEC.

International Electrotechnical Commission (2017). *Life-Cycle Management for Systems and Products Used in Industrial-Process Measurement, Control and Automation*. Geneva: IEC.

ISO 15704 (2000). *Industrial Automation Systems—Requirements for Enterprise-reference Architectures and Methodologies* (Geneva).

Leal, G. S., Guédria, W., and Panetto, H. (2020). Enterprise interoperability assessment: a requirements engineering approach. *Int. J. Comp. Integ. Manuf.* 33, 265–286. doi: 10.1080/0951192X.2020.1736636

Lee, J., Bagheri, B., and Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Let.* 3, 18–23. doi: 10.1016/j.mfglet.2014.12.001

Lepenioti, K., Bousdekis, A., Apostolou, D., and Mentzas, G. (2020). Prescriptive analytics: literature review and research challenges. *Int. J. Inf. Manag.* 50, 57–70. doi: 10.1016/j.ijinfomgt.2019.04.003

Li, D., Wang, Y., Wang, J., Wang, C., and Duan, Y. (2020). Recent advances in sensor fault diagnosis: a review. *Sens. Act. Phys.* 2020:111990. doi: 10.1016/j.sna.2020.111990

Maier, M., Serebrenik, A., and Vanderfeesten, I. T. P. (2013). *Towards a Big Data Reference Architecture*. Eindhoven: University of Eindhoven.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., et al. (2017). Big IoT data analytics: architecture, opportunities, and open research challenges. *IEEE Access* 5, 5247–5261. doi: 10.1109/ACCESS.2017.2689040

Ministry of Industry and Information Technology of China and Standardization Administration of China (2015). *National Intelligent Manufacturing Standard System Construction Guidelines*. Beijing: MIIT.

Moghaddam, M., Cadavid, M. N., Kenley, C. R., and Deshmukh, A. V. (2018). Reference architectures for smart manufacturing: a critical review. *J. Manuf. Syst.* 49, 215–225. doi: 10.1016/j.jmsy.2018.10.006

MQTT (2019). *Version 5.0 OASIS Standard Specification*. Woburn, MA: OASIS.

NIST (2019). *Big Data Interoperability Framework*. Available online at: https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-1-definitions (December 12, 2020).

Nunez, D. L., and Borsato, M. (2017). An ontology-based model for prognostics and health management of machines. *J. Ind. Inform. Integr.* 6, 33–46. doi: 10.1016/j.jiii.2017.02.006

Open Group TOGAF (2000). *The Open Group Architecture Framework, Document No. 1910, Version 6* (San Francisco, CA).

Pääkkönen, P., and Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big Data Res.* 2, 166–186. doi: 10.1016/j.bdr.2015.01.001

Panetto, H., Zdravkovic, M., Jardim-Goncalves, R., Romero, D., Cecil, J., and Mezgár, I. (2016). New perspectives for the future interoperable enterprise systems. *Comp. In Ind.* 79, 47–63. doi: 10.1016/j.compind.2015.08.001

Papageorgiou, N., Verginadis, Y., Apostolou, D., and Mentzas, G. (2019). Fog computing context analytics. *IEEE Instr. Meas. Magaz.* 22:, 53–59. doi: 10.1109/MIM.2019.8917904

Pedone, G., and Mezgár, I. (2018). Model similarity evidence and interoperability affinity in cloud-ready Industry 4.0 technologies. *Comp. Ind.* 100, 278–286. doi: 10.1016/j.compind.2018.05.003

Pisching, M. A., Pessoa, M. A., Junqueira, F., dos Santos Filho, D. J., and Miyagi, P. E. (2018). An architecture based on RAMI 4.0 to discover equipment to process operations required by products. *Comp. Ind. Eng.* 125, 574–591. doi: 10.1016/j.cie.2017.12.029

Platform Industrie 4.0. (2019). *Details of the Asset Administration Shell Part 1 - The Exchange of Information Between Partners in the Value Chain of Industrie 4.0 (Version 2.0)*. Available online at: https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/Details-of-the-Asset-Administration-Shell-Part1.html (accessed December 03, 2020).

Rezaei, R., Chiew, T. K.„ Lee, S. P., and Aliee, Z. S. (2014). A semantic interoperability framework for software as a service systems in cloud computing environments. *Exp. Syst. Appl.* 41, 5751–5770. doi: 10.1016/j.eswa.2014.03.020

Romero, D., and Vernadat, F. (2016). Enterprise information systems state of the art: past, present and future trends. *Comp. Ind.* 79, 3–13. doi: 10.1016/j.compind.2016.03.001

Ruschel, E., Santos, E. A. P., and Loures, E. D. F. R. (2017). Industrial maintenance decision-making: a systematic literature review. *J. Manufact. Syst.* 45, 180–194. doi: 10.1016/j.jmsy.2017.09.003

Salma, C. A., Tekinerdogan, B., and Athanasiadis, I. N. (2017). "Domain-driven design of big data systems based on a reference architecture," in *Software Architecture for Big Data and the Cloud*, eds I. Mistrik, N. Ali, R. Bahsoon, M. Heisel, and B. Maxim (Cambridge, MA: Morgan Kaufmann), 49–68. doi: 10.1016/B978-0-12-805467-3.00004-1

Sang, G. M., Xu, L., and De Vrieze, P. (2017). "Simplifying big data analytics systems with a reference architecture," in *Working Conference on Virtual Enterprises* (Cham: Springer), 242–249. doi: 10.1007/978-3-319-65151-4_23

Scheer, A. W. (1994). *Business Process Engineering. Reference Models for Industrial Enterprises, 2nd Edn.* Berlin: Springer-Verlag. doi: 10.1007/978-3-642-79142-0

Schmidt, R., and Möhring, M. (2013). "Strategic alignment of cloud-based architectures for big data," in *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops* (Vancouver, BC: IEEE), 136−143. doi: 10.1109/EDOCW.2013.22

Selway, M., Stumptner, M., Mayer, W., Jordan, A., Grossmann, G., and Schrefl, M. (2017). A conceptual framework for large-scale ecosystem interoperability and industrial product lifecycles. *Data Knowl. Eng.* 109, 85–111. doi: 10.1016/j.datak.2017.03.006

Setiawan, F. A., Budiardjo, E. K., and Wibowo, W. C. (2019). ByNowLife: a novel framework for OWL and bayesian network integration. *Information* 10:3. doi: 10.3390/info10030095

Snydman, S., Sanderson, R., and Cramer, T. (2015). "The international image interoperability framework (IIIF): a community & technology approach for web-based images," in: *IS&T (Digital) Archiving Conference 2015* (Angeles, CA: Society for Imaging Science and Technology), 16–21.

Standardization Council Industrie 4.0 (2020). *DIN and DKE ROADMAP: German Standardization Roadmap Industrie 4.0, Version 4*. Berlin: DIN.

Trinks, S., and Felden, C. (2018). "Edge computing architecture to support real time analytic applications: a state-of-the-art within the application area of Smart Factory and Industry 4.0," in *2018 IEEE International Conference on Big Data (Big Data)* (Seattle, WA: IEEE), 2930–2939. doi: 10.1109/BigData.2018.8622649

Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., et al. (2011). "Process mining manifesto," in *International Conference on Business Process Management* (Berlin, Heidelberg: Springer), 169–194.

Varghese, B., and Buyya, R. (2018). Next generation cloud computing: new trends and research directions. *Fut. Gen. Comp. Syst.* 79, 849–861. doi: 10.1016/j.future.2017.09.020

Vinoski, S. (2006). Advanced message queuing protocol. *IEEE Int. Comp.* 10, 87–89. doi: 10.1109/MIC.2006.116

Wei, S., Hu, J., Cheng, Y., Ma, Y., and Yu, Y. (2017). "The essential elements of intelligent manufacturing system architecture," in *2017 13th IEEE Conference on Automation Science and Engineering (CASE)* (Xi'an: IEEE), 1006−1011. doi: 10.1109/COASE.2017.8256234

Wen, Y., Wu, J., Zhou, Q., and Tseng, T. L. (2018). Multiple-change-point modeling and exact Bayesian inference of degradation signal for prognostic improvement. *IEEE Trans. Automat. Sci. Eng.* 16, 613–628. doi: 10.1109/TASE.2018.2844204

Westerlund, M., Hedlund, U., Pulkkis, G., and Björk, K. M. (2014). "A generalized scalable software architecture for analyzing temporally structured big data in the cloud," in *New Perspectives in Information Systems and Technologies, Volume 1* (Cham: Springer), 559–569. doi: 10.1007/978-3-319-05951-8_53

Williams, T. J. (1994). The Purdue enterprise reference architecture. *Comp. Ind.* 24, 141–158. doi: 10.1016/0166-3615(94)90017-5

Xu, L. D., and Duan, L. (2019). Big data for cyber physical systems in industry 4.0: a survey. *Ent. Inf. Syst.* 13, 148–169. doi: 10.1080/17517575.2018.1442934

Xu, Y., Sun, Y., Wan, J., Liu, X., and Song, Z. (2017). Industrial big data for fault diagnosis: taxonomy, review, and applications. *IEEE Access* 5, 17368–17380. doi: 10.1109/ACCESS.2017.2731945

Ye, X., and Hong, S. H. (2019). Toward industry 4.0 components: insights into and implementation of asset administration shells. *IEEE Ind. Electr. Magaz.* 13, 13–25. doi: 10.1109/MIE.2019.2893397

Yue, M., Hong, T., and Wang, J. (2019). Descriptive analytics-based anomaly detection for cybersecure load forecasting. *IEEE Trans. Sm. Grid* 10, 5964–5974. doi: 10.1109/TSG.2019.2894334

Zachman, J. (1996). *The Framework for Enterprise Architecture: Background, Description and Utility.* Monument, CO: Zachman Institute for Advancement.

Zeid, A., Sundaram, S., Moghaddam, M., Kamarthi, S., and Marion, T. (2019). Interoperability in smart manufacturing: research challenges. *Machines* 7:21. doi: 10.3390/machines7020021

Zezulka, F., Marcon, P., Vesely, I., and Sajdl, O. (2016). Industry 4.0–an introduction in the phenomenon. *IFAC-PapersOnLine* 49, 8–12. doi: 10.1016/j.ifacol.2016.12.002

Zhang, N., Wu, L., Wang, Z., and Guan, Y. (2018). Bearing remaining useful life prediction based on Naive Bayes and Weibull distributions. *Entropy* 20:944. doi: 10.3390/e20120944

Zhong, T., Doshi, K. A., Tang, X., Lou, T., and Lu, Z. (2013). "On Mixing high-speed updates and in-memory queries," in *A Big-Data Architecture for Real-time Analytics, IEEE BPOE Workshop* (Silicon Valley, CA). doi: 10.1109/BigData.2013.6691704

Zhou, A., Yu, D., and Zhang, W. (2015). A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. *Adv. Eng. Inform.* 29, 115–125. doi: 10.1016/j.aei.2014.10.001

Zillner, S., Gomez, J. A., Robles, A. G., Curry, E., Södergård, C., Boujemaa, N., et al. (2018). *Data-Driven Artificial Intelligence for European Economic Competitiveness and Societal Progress: BDVA Position Statement, November 2018.* Brussels: BDVA.

Zimmermann, A., Sandkuhl, K., Pretz, M., Falkenthal, M., Jugel, D., and Wissotzki, M. (2013). "Towards an integrated service-oriented reference enterprise architecture," in *Proceedings of the 2013 International Workshop on Ecosystem Architectures* (New York, NY), 26–30. doi: 10.1145/2501585.2501591

Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., and Li, G. P. (2020). Predictive maintenance in the Industry 4.0: a systematic literature review. *Comp. Ind. Eng.* 2020:106889. doi: 10.1016/j.cie.2020.106889