An IoT-based Reliable Industrial Data Services for Manufacturing Quality Control

Raul Poler
Research Centre on Production Management and Engineering (CIGIP)
Escuela Politécnica Superior de Alcoy, Universitat Politècnica de València
Alcoy, Spain
rpoler@cigip.upv.es

Anastasios Karakostas
Information Technologies Institute
Centre for Research and Technology Hellas (CERTH-ITI)
Thessaloniki, Greece
akarakos@iti.gr

Stefanos Vrochidis
Information Technologies Institute
Centre for Research and Technology Hellas (CERTH-ITI)
Thessaloniki, Greece
stefanos@iti.gr

Angelo Marguglio
Industry and Security Technologies, Research and Innovation (IS3) Lab
Engineering Ingegneria Informaticaspa
Palermo, Italy
angelo.marguglio@eng.it

Santiago Gálvez-Settier
Research and Development Services
Instituto Tecnológico de Informática (ITI)
Valencia, Spain
sgalvez@iti.es

Paulo Figueiras
CTS,
UNINOVA
Caparica, Portugal
paf@uninova.pt

Ana Gómez-González
Control and Monitoring Area
Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA)
Arrasate, Spain
ana.gomez@ikerlan.es

Benny Mandler
IBM Haifa Research Laboratory
Haifa, Israel
mandler@il.ibm.com

Stefan Wellsandt
ICT applications for production
BIBA – Bremer Institut für Produktion und Logistik GmbH
Bremen, Germany
wel@biba.uni-bremen.de

Robert Trevino
Chair of Quality Science
Technische Universität Berlin
Berlin, Germany
robert.trevino@tu-berlin.de

Shukri Bassoumi
EXOS Solutions S.L.
Valencia, Spain
shukri.bassoumi@exos-solutions.com

Carlos Agostinho
Knowledgebiz
Caparica, Portugal
https://orcid.org/0000-0002-2884-776X

2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) | 978-1-6654-4963-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICE/ITMC52061.2021.9570203

***Abstract*— This paper presents a complete solution consisting of sustainable IoT-based Reliable Industrial Data Services (RIDS) able to manage the huge amount of industrial data coming from cost-effective, smart, and small size interconnected factory devices for supporting manufacturing online monitoring and control. The i4Q Framework guarantees data reliability with functions grouped into five basic capabilities around the data cycle: sensing, communication, computing infrastructure, storage, and analysis and optimisation. With the i4Q RIDS, factories will be able to handle large amounts of data, achieving adequate levels of data accuracy, precision and traceability, using it for analysis and prediction as well as to optimise the process quality and product quality in manufacturing, leading to an integrated approach to zero-defect manufacturing. The i4Q Solutions efficiently collect the raw industrial data using cost-effective instruments and state-of-the-art communication protocols, guaranteeing data accuracy and precision, reliable traceability and time stamped data integrity through distributed ledger technology and provide simulation and optimisation tools for manufacturing line continuous process qualification, quality diagnosis, reconfiguration and certification for ensuring high manufacturing efficiency and optimal manufacturing quality.**

***Keywords*— Data Quality, Data Reliability, Blockchain, Product Quality, Process Quality, Digital Twin, Process Simulation, Process Optimization, Zero-defect Manufacturing**

I. Introduction

Manufacturing companies are continuously facing the challenge of redesigning and adjusting their manufacturing systems to adapt their process to produce goods adapted to specific requirements and produced under the minimum required production rate, guaranteeing high quality and limiting the use of resources in order to reduce production costs. Therefore, reducing waste, scraps and defects, as well as production costs and lead times is crucial to increase productivity and hence, to pursuit manufacturing excellence.

In this context, the implementation of zero-defect strategies plays a decisive role. During the last decade, several R&D efforts have targeted on zero defect approaches with the purpose of developing solutions to improve performance of process control by incorporating enhanced quality control solutions. Nevertheless, current solutions need further developments on:

- Data management: Thanks to the increase in the use of sensors, actuators and instruments, manufacturing lines collect a huge amount of data during the manufacturing process, which is very valuable for the improvement of quality in manufacturing but, for most of the European factories, it is not possible to analyse the data generated in the process on a daily basis.

- Complexity of current solutions: Requiring heavy statistical and technology training and support, making them not accessible for SMEs. Now, users are demanding access to insights from advanced analytics, without requiring them to have IT or data science advanced skills. Most of the current solutions lack of easy-to-use advanced data preparation, production reporting and advanced analytics and prediction.
- Dynamic behaviour of the manufacturing factories: Complex systems of diverse, connected, interdependent entities which need suitable modelling and simulation approaches and data fusion techniques to interpret the collected data.

A successful smart factory should be able to manage data-related processes along the entire data life cycle, including data collection, storage, distribution, analysis, use, and deletion, to ensure high data quality at all times. This includes processes related to: (i) the design, deployment, and use of hardware and software; (ii) the planning, implementation, and monitoring of intra-organisational procedures; (iii) and the inter-organisational practices in the value chain. The comprehensive quality control of all important factors is an effective measure against unfit, erroneous, unintelligible, or otherwise unreliable data.

This paper presents the i4Q Project, whose approach targets a complete solution consisting of sustainable IoT-based Reliable Industrial Data Services (RIDS) able to manage the huge amount of industrial data coming from cost-effective, smart, and small size interconnected factory devices for supporting manufacturing online monitoring and control. In section 2, the authors present a literature review of the most relevant concepts and research in the area. Section 3 describes the i4Q Project proposition. Section 4 discusses the advances provided by i4Q beyond the start-of-the-art technologies. Section 5 presents the conclusions and future work.

II. LITERATURE REVIEW

A. *Manufacturing Data Quality*

Data Quality is a term coined in the 1990s. It adopted the ideas of product quality management and the basic notion for quality as: “degree to which a set of inherent characteristics of an object fulfils requirements” [1]. The reference object is data, though oftentimes, it is more useful in practice to widen the scope to information, i.e. data with semantics (meaning). The main difference between data quality and information quality is that the former focuses on the technical means to collect and store data (e.g. databases), while the latter focuses on the application of meaningful data and the fit-for-purpose aspect, i.e. how well information meets information needs. The operationalisation of data quality requires the specification of relevant inherent characteristics of the data. While the exact characteristics typically depend on the data user and the application purpose of the data, several researchers and practitioners developed quality models that describe common data quality characteristics. Acknowledged models are the ones described in [2], [3], and the ISO 25012 standard [4]. The latter proposes 15 characteristics including, for instance, accessibility, accuracy, confidentiality, completeness, compliance, credibility, recoverability, and understandability. Besides these quality models, the literature also identified data quality management frameworks that suggest a management scope and procedures aligned to basic management processes, such as Plan-Do-Study-Act. These frameworks include early approaches, such as Total Data Quality Management [5], and later ones that focus, e.g., on Big Data [6]. The data quality management typically needs to specify data quality characteristics as measurable indicators and based on the data application context. In manufacturing, the quality of data is important in most decisions taken by humans (e.g., planning, and operations) or machines (e.g., artificial intelligence). Successful data analytics depend on data with sufficient quality for the individual analytics task [7], [8]. This typically concerns characteristics, such as accuracy, precision, completeness, and timeliness, because they influence algorithms results significantly. Besides the selection of relevant quality characteristics, also the factors that influence these characteristics over the entire data life cycle require attention. [9] suggested to use cause-effect diagrams to identify and analyse data quality factors related to the collection, organisation, presentation, and application of data. Understanding and controlling these factors can be a key to manage data quality comprehensively.

B. *Manufacturing Data Collection*

Modern manufacturing produces high volumes of data [10] up to the point that the concept of “Smart Manufacturing” is itself tightly intertwined to that of data-driven manufacturing [11], allowing companies, for instance, to visualise, analyse and react to both collected and real-time (or near real-time) information, relevant to many areas of manufacturing, ranging from production to maintenance, order management and supply chain. Additionally, data can be used in periodic analysis and strategic/business planning.

As highlighted in various research, e.g., [12], [13], [14], the quality of data plays a critical role in business applications under various aspects including performance, decision-making (management) and cooperation. As already highlighted by [2] it is, of course, important to define what ‘data quality’ actually means. In fact, there are several dimensions (and measures) related to the concept. In their book on Data Quality [12] and then in [15] examine in great detail the dimensions of data quality, highlighting how literature does not always agree on the definitions of such dimensions and measurements. Furthermore, [16] provide a detailed overview of data quality assessment methodologies and frameworks. Nonetheless, common attributes/measures defining a basic set of data quality attributes can be identified in the dimensions of accuracy, completeness, consistency, timeliness. Regarding accuracy, two types can be defined: syntactic and semantic. The former essentially assesses how close a data value is to a set of values defined in a domain considered syntactically correct. The latter assesses closeness from a semantic point of view. Completeness assesses the degree to which a given data collection includes data describing the corresponding set of real-world objects. In certain domains (especially databases), completeness has to do with the presence (and meaning) of null values, therefore some authors suggest that during quality assessment a Boolean value should be associated with a field. Consistency assesses the adherence to semantic rules defined over a set of data i.e., answering the question: “are data consistent across the data sets?” and “are the data representing conflicting information?”. Currency is a time dimension which relates to how often the data is updated. Related time dimensions are volatility which represents how frequently data changes in time (e.g., a birth date has volatility equal to zero), and timeliness which specifies the currency of data with respect to a given task/usage e.g., data could be current, but still late for a certain usage. Accessibility relates to the capability of users to access data within their own context, including physical status/functions, technologies and culture. [17] focus on data (quality) dimensions in Digital Manufacturing and consider information as manufacturing 'product' itself. [18] highlight how in real world industrial application it is not uncommon to face situation of low quality data including non-accuracy, non-completeness due to data loss, non-consistency due to different vendors, software versions, etc. To this end a set of data cleaning strategies are proposed [19] such as duplicate identification and elimination, data transformations, schema matching and data mining approaches. Additionally, domain ontologies are suggested as a tool to improve data quality management tasks [20]. [21] report several efforts to propose ontologies, semantics, and semantic web technologies within manufacturing and [22] underline how semantically rich descriptions can provide benefits in Industry 4.0 scenarios. Indeed, several ontologies for manufacturing exist such as PSL (Process Specification Language), MASON (Manufacturing's Semantics Ontology), SIMPM (Semantically Integrated Manufacturing Planning Model), among others. In dealing with data quality, especially in the context of industrial applications and manufacturing it is important to mention the ISO 8000 [23], the international standard for data quality. ISO 8000 considers data as a representation of information and quality data as instrumental to capture, store and share information. Data portability is one of the core characteristics of ISO 8000. On the one hand this facilitates preservation and exchange, but (within IT systems) also allows to separate data from the software being used to manage it. Another fundamental principle of ISO 8000 is the use of dictionaries for both consistency and portability.

C. *Manufacturing Data Analytics*

Since the third Industrial Revolution, which was characterised by the emergence of the digital information age, that manufacturers all over the world are embracing the notion of convergence of the digital and physical worlds [11]. Mainly due to this convergence and to technological advances achieved throughout the last two decades, manufacturing-related data is being generated at exponentially growing rates [24]. Still, there are few manufacturing sectors that truly capitalise on such amount of collected data, by extracting meaningful insights for supporting improvements on their businesses, processes and products [25]. Recently, the application of Data Analytics to manufacturing data has been presented as a solution for the issue of capitalising on ever-growing manufacturing data [26]. Manufacturing Data Analytics can be defined as the process of finding useful information from analysing manufacturing-generated raw data, whether for decision-making support or for optimisation of business and production processes, among other objectives [15]. [27] present the main objectives for applying Big Data Analytics (BDA) in smart manufacturing. It is envisioned that BDA applications will be able to assist enterprise managers to learn everything about what they did today and to predict what they will do tomorrow. This future vision is based on a taxonomy of data analytics approaches for manufacturing, which entails four types of analytics processes: descriptive, diagnostic, predictive and prescriptive analytics [28], [24]. Both descriptive and diagnostic analytics methods are reactive while predictive and prescriptive analytics approaches are proactive. Descriptive analytics is an exploratory analysis of historical data to tell what happened. During this stage, most of data mining and statistical methods can be used to reveal the data characteristics, recognise patterns and identify relationships of data objects. Diagnostic analytics is a deeper look at data to attempt to understand the causes of events and behaviours. The diagnostic analysis of machines and other equipment can help to identify the possible faults and predict the failures to reduce the machine down-times. Predictive analytics mainly utilises historical data to anticipate the trends of data (i.e., what will occur in the future). Finally, prescriptive analytics extends the results of descriptive, diagnostic and predictive analytics to make the right decisions in order to achieve predicted outcomes. The prescriptive methods typically include simulation, decision-making, optimisation and reinforcement learning algorithms. Although the three first types of data analytics are not new research trends, the fourth, prescriptive analytics, is seen as a future challenge in Manufacturing Data Analytics [28], [24], and is closely linked to simulation (Digital Twins) and optimisation.

D. *Manufacturing Data Trustworthiness*

Trust in data may rely on a blockchain-based platform. Blockchain, at heart, is a distributed ledger which maintains ordered records of all transactions that occur over a network. Participants in a blockchain network maintain a replicated shared state through consensus algorithms [29]. A transaction recorded in a blockchain signifies information pertaining to an exchange of an asset, data or monetary value, and each transaction is governed by a smart contract. Transaction information is grouped together in blocks which get appended to the blockchain, with transaction validation and immutability guaranteed, through algorithmic consensus among participants in a blockchain network. Since each block is linked to its predecessor block, altering any contents of a single block incurs alterations in all subsequent blocks, and algorithmic consensus ensures any corrupted copies of the ledger are detected and corrected. Blockchain is a fully decentralised technology that does not rely on single trusted authorities for record-keeping between transacting parties, thus blockchains create a trusted environment for conducting transactions. The blockchain technology is currently being applied in numerous application domains, including financial services, and supply-chain management [30]. For traceability blockchain can be used to provide an audit trail for assets exchange, use and associated data [31]. Transactions across multiple parties, protected by a security and privacy layer, are immutable offering transparency and trust in data and transactions. Blockchain provides for immutability, finality, consensus, and provenance of its state and all the transactions therein. The primary use of blockchains is the transfer of assets (tokens) among different parties. The same infrastructure can be used for conserving data reliability. New trends in Blockchain technology lead to an integration with the Internet of Things domain, providing a powerful combination that can cause significant transformations across several industries, paving the way for new distributed applications. Being able to provide guaranties of the authenticity of data is important to enhance trust in incoming data and act accordingly. Inability to provide data reliability guarantees might endanger the credibility of the decision maker.

E. *Manufacturing Process Qualification and Reconfiguration*

Qualification involves evaluating a complete process, consisting of many individually certified activities, to determine whether the process can perform at the appropriate level when the activities are linked together [32]. Process qualification is a component of the process validation, which analyses the data gathered throughout the design and manufacturing of a product in order to confirm that the process can reliably output products of a determined standard. Process validation has been widely applied and documented in the pharmaceutical, medicine, food and drugs research areas [33] [34]. Process validation can be broken down into 3 steps: process design, process qualification, and continued process verification. In this regard, manufacturing process qualification enables to assess if the process is able to meet determined manufacturing targets. In this stage all production processes and manufacturing equipment are proofed to confirm quality and output capabilities. Critical quality attributes are evaluated and critical process parameters taken into account to confirm product quality [33]. The process performance qualification (PPQ) protocol is a key element of process qualification, that allows recording and having available for review essential conditions, controls, testing, and expected manufacturing outcome of a production process. The following criteria must be considered in the PPQ protocol: (i) manufacturing conditions, including operating parameters, equipment limits, and component inputs; (ii) the data that should be recorded and analysed; (iii) the tests that should be performed to ensure quality at each production step; (iv) a sampling plan to outline sampling methods both during and between production batches; (v) an analysis methodology that allows for data scientific and risk oriented decision making based on statistical data; and (vi) variability limits should be defined and contingencies in the event of non-conforming data established.

Reconfiguration aims at modifying the manufacturing process to rapidly and cost efficiently (i) adapt the production to market changes, (ii) increase production flexibility for mass customisation, or (iii) react upon unpredicted events such as machine faults or quality degradation [35]. Process reconfiguration relies on reconfigurable machines and controllers, and methodologies for the systematic design of new process configurations [36]. Such methodologies must account for the current customer specifications, current product quality, and available resources. Therefore, they feed on many heterogeneous data sources, such as sensors, customer demands or diagnostic analytics. In this context, the Digital Twin (DT) concept is an attractive solution to supply the required data in a systematic way [37]. According to [38], the "DT is an integrated multi-physics simulation of a system, that uses the available physics-based models, sensors and fleet history to replicate the behaviour of its real counterpart". Through these models, the DT behaves as a one-to-one replica of a system, running parallel to it and with the same known operational conditions. Such a replica, may thus give insight into the system's behaviour, for instance, by (i) providing virtual sensors to extend available data [39], (ii) generating synthetic data to infer the system's behaviour under unknown conditions [40], (iii) providing descriptive features [41], or simulating operational decisions to aid in the decision making.

III. i4Q Conceptual Approach

The design of the i4Q IoT-based Reliable Industrial Data Services (RIDS) is based on the i4Q Reference Architecture (RA). The conceptualisation of the i4Q RA (Fig. 1) follows the ISO/IEC/IEEE 42010 "Systems and software engineering – Architecture" and it is aligned with the most common reference architectures in the manufacturing domain: IIRA, RAMI4.0, IDSA, and IMSA. i4Q RA is structured around five basic layers: physical, network, middleware, database, and application.

The i4Q digital representation of data produced during manufacturing and supply chain processes is described considering well-known standards and ontologies adopted in the manufacturing domain: B2MML, AutomationML, CAEX, PLCOpen, COLLADA, MTConnect, MIMOSA, PSL, MASON, SIMPM, etc., in order to ease data interoperability, exchange and processing. The exchange of digital data across i4Q Solutions is supported by a range of data models and ontologies providing the interoperability specification used in the different viewpoints of the i4Q RA.

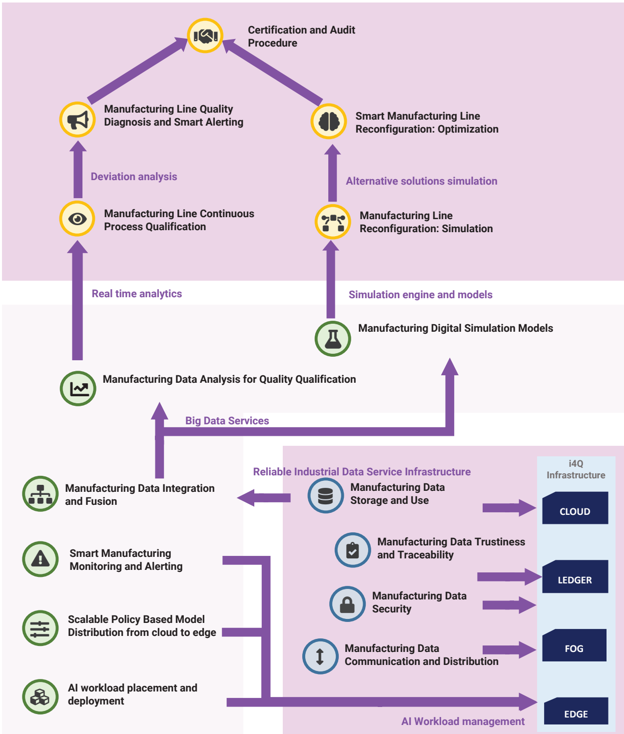

Fig. 1. i4Q Conceptual Architecture

The design of the i4Q RA is realised taking into account the analysis across four key viewpoints: business, usage, functional and implementation.

- The *Business viewpoint* has the objective of avoiding the 'technology-centric' view in the design phase. This view allows to incorporate in the early design the stakeholders' requirements and needs which are closer to real world.
- The *Usage viewpoint* includes the tasks, roles, activities and parties, taking into account both human beings and software systems. One level deeper defines the functional map, the implementation maps, the role responsible for the execution of the tasks, the triggers that start the activity, workflows that define the organisation of the tasks within the activity and the

978-1-6654-4963-2/21/$31.00@2021 IEEE

effects that will produce the execution of the activity on the system and the constraints of its execution.

- The *Functional viewpoint* decomposes the i4Q RA into its Control, Operations, Information, Application and Business domains and identify the data, decision and command/request flows circulating among them, and the controls, coordination and orchestration exercised from each of these domains as well as on the different typical operations from these domains.
- The *Implementation viewpoint* technically describes the different components of the i4Q RA, how they are interconnected and the technologies required for its proper implementation. On the other hand, it provides a detailed architecture based on cloud computing patterns, considering edge computing, microservice applications, and Function as a Service (FaaS) platforms.

A. Reliable Industrial Data Services Infrastructure

The i4Q RIDS infrastructure provides the necessary strategies, methods, and key technologies to ensure data quality, which is influenced by many factors including human error, communication issues or inaccuracies. Monitoring systems are expected to be reliable enough for decision-making, but sensors are susceptible to provide unreliable information. Trusted networks allow ensuring the reliability, integrity and privacy of the data exchanged. The heterogeneity in technologies leads to a complex ecosystem of data models that is difficult to contextualise, leading to misinterpretations and incomplete data analysis. Setting the guidelines for information models, needed metadata for traceability and interoperability is crucial for ensuring data quality. The increasing amount of collected data requires flexible data storage and data analysis infrastructure able to process information without affecting data quality.

The i4Q RIDS provides an easier, trustable and traceable access to data coming from many different sources by employing a blockchain based data service. This enhances trust and acceptability by providing security and trust in the data that flows directly to the blockchain, serving as a single point of truth, preserving provenance and supporting non-repudiation. Information stored on the blockchain cannot be changed or erased and can be proved to be authentic. Blockchain based services ensure that the information is not tampered with. Moreover, differential visibility scopes of information are supported for cases in which a subset of the participating entities needs to share some information. The blockchain capabilities are exposed to other platform components and microservice applications via REST interfaces. Smart contracts are provided to govern the actions that take place upon the arrival of new data, and to ensure that the required participants in the network approve incoming transactions. The REST interfaces provided enable the invocation of transactions to add data to the blockchain and a query mechanism to retrieve data previously stored in the blockchain.

The i4Q RIDS implements reliable data collection, providing connectivity to industrial data sources through Trusted Networks able to assess and ensure precision, accuracy, and reliability. Technologies such as TSN (Time Sensitive Networks) for wired communications, and wireless access networks (e.g. Industrial WSN, LPWAN, ad-hoc connections...), are integrated with other solutions such as Software Defined Network, Network Function Virtualisation, or Network Slicing in order to improve reliability of the communication infrastructure and therefore the integrity and reliability of data collected. This includes software-defined wireless industrial interfaces for data communication, paying special attention to requirements such as predictability and determinism, high reliability and trustability and low consumption, while reducing the installation cost of new-wired infrastructure. Time-sensitive transmission of data over deterministic Ethernet networks is also required for applications that require very low transmission latency and high availability and can use the floor plant wired network infrastructure.

Collected data can be delivered by diverse types of devices and be transmitted through and processed by a significant number of layers and technologies. This translates into the necessity of recommendations and guidelines to enable multilayer cyber security features in Industrial Internet of Things (IIoT), as well as the tools to implement these recommendations, enabling IIoT devices to interact with the platform securely in all stages of a manufacturing scenario. The i4Q RIDS provisions signed certificates with Hardware Security Module (HSM) and trusted material to devices' Trusted Platform Module (TPM) using asymmetric encryption architecture. Furthermore, it includes the use of software-Trusted Execution Environment (TEE) in order to set the boundaries between the security and non-security processes running in the devices. Security mechanisms may differ between communication solutions or Distributed Ledger Technology (DLT) tools, which means that it is needed to apply security by design during development, adjusting security, and safety policies at different levels to ensure the trustability and privacy of data.

The i4Q RIDS implements a distributed storage system taking into account those aspects unique to the Industry 4.0 paradigm. One of the main aspects to consider is the high degree of digitisation expected in companies, resulting in most manufacturing devices acting as sensors or actuators and generating vast amounts of data. The infrastructure has to be, then, able to absorb large volumes of data coming into the system at high speeds. Similarly, it has to be as elastic as possible, to adapt the computing resources it requires to the existing demand and be ready to use additional resources, either local to the factory or from remote systems like public or private clouds if needed, although bearing in mind possible data privacy restrictions. In addition to the storage capabilities, the i4Q storage system provides easy ways to access this data so other platform components and microservice applications can easily consume and use it to improve the efficiency of the system.

B. Big Data Services

The i4Q RIDS data services provide the mechanisms for the analysis of manufacturing data by combining simulation and real data, while employing data fusion techniques. Microservice applications are able to use manufacturing data processing services, data streams and artificial intelligence (AI) models, scaling up resources in an efficient and transparent way.

The i4Q Data Integration and Transformation Services prepare manufacturing data so that microservice applications can process it. This includes all the elements required for manufacturing data stream management: reading, cleaning, storing, indexing, enriching, searching & retrieving, maintaining, and correspondence of open APIs. The key characteristics of the data to be managed are: (i) variety – supporting different data types from different sensors; (ii) velocity – most of the data includes more or less intensive data streams to be processed in real-time; (iii) volume – data size to be processed is typically large (GBs) to very large (TBs). Big Data analytic databases are used for supporting the required specifications and being adapted/parameterised to support the required complexity of higher-level services.

The i4Q Manufacturing Data Analysis for Quality Qualification is a set of specialised analytic functions on top of the data infrastructure, implementing several incremental algorithms (i.e., operating on data streams with fast incremental updates) suitable for analytic processing of high-speed data streams. The core functions are related to clustering, regression, classification, anomaly detection and temporal correlation. The key properties of these implementations are speed and ability to support intensive data streams.

The i4Q RIDS includes a multi-tier infrastructure to address the management of AI-based workloads in a hybrid cloud edge manufacturing environment. Scalability is a major concern for smart manufacturing environments that involve multiple sites. On the one hand, a discovery component keeps track of all components of the distributed system and, on the other hand, a policy-based mechanism eases the task of the controller by enabling the specification of rules for eligible targets in a simplified manner (such as the model type, model version, geographic area, or the existence of specific resources). The AI model distribution is coordinated with the workload distribution mechanism to ensure that the right set of AI models is made available for the workload that uses them.

Analysis capabilities utilise the potential to deploy and run AI workloads on the edge computing environments prevalent in manufacturing facilities. The i4Q RIDS enables workloads to execute efficiently on the edge, including placement and deployment services. Target deployment environments may be very heterogeneous and dynamic; thus, deployment needs to take a variety of criteria into consideration. The environment is dynamic thus re-deployment of the entire workload or the adaptation of the underlying model may be required while the workload is running. A Cloud/edge architecture provides efficient and flexible management of edge workloads with deployment on an orchestrator.

The i4Q RIDS provides scalable monitoring tools designed for the manufacturing edge environment. Monitoring is performed on applications running on the edge and the resources they consume, and specific data of the AI models. The information that is collected can be used in smart workload assignment to edge nodes and AI models evaluation. Monitoring is performed at various levels, such as the infrastructure, available resources, and the active workloads. The collected information provides an updated view of resources consumption and availability, in turn enabling better workload distribution and deployment. Collected data can be used for retraining of the model, which later shall be used for updating the running workloads. This task shall develop monitoring tools designed for smart manufacturing workload orchestration and predictive failure alerting, including monitoring the health of workloads and productively alerting and taking corrective actions when a predicted problem is detected.

The i4Q Digital Twin implements digital twins representing the complete production line to control the full traceability and to streamline and digitalize the entire production process. Supervised classification techniques (artificial neural networks, support vector machines, boosting, pattern recognition, etc.) and time series analysis (estimate uncertain events of the past, predict future events) are used to virtually replicate various stages of the manufacturing process (machine, tools and process parameters and product characteristics) to improve further process control and to avoid bottlenecks. In addition, model-based strategies are exploited for the estimation of virtual sensors and quality-sensitive features. To this end, physics-based models of the plant and model updating strategies are included to achieve a connected 3D production simulation, with a digital twin for manufacturing enabling virtual validation/visualisation and productivity optimisation using pre-existing and data from different factory levels (small cell to entire factory).

C. *Manufacturing Line Qualification and Reconfiguration*

Process qualification in flow process manufacturing is an essential step during ramp-up and after reconfiguration of production processes. The i4Q Data-Driven Continuous Process Qualification overcomes the classical approaches and methods for Continuous Process Qualification (CPQ) for ensuring best practice for manufacturing high quality goods and assure final product quality through adequate control over processes, collection of multivariate data and statistical procedures for evaluation of process stability and process performance. It implements an automated CPQ system based on real-time data provided AI algorithms combined with improved smart data analytics and algorithms results in faster process approval and in-line continuous process validation after process reconfiguration.

The i4Q Manufacturing Line Quality Diagnosis and Smart Alerting tools provide the diagnosis of the condition of the manufacturing line by evaluating data fidelity, product-quality, and process condition. The smart alerting system includes auto alerting systems and data visualisation tools to achieve zero defect manufacturing. Based on the diagnosis, action recommendations are provided, such as sensor/data processing recalibrations, process line/machine reconfiguration, or maintenance actions.

The i4Q Prescriptive Analysis Tools use simulation strategies to investigate whether small changes in the control can reduce or even eliminate the defects on the production, taking as input the identified process condition, available resources and current production planning, and proposing process configuration parameters. This would open the way to an opportunistic process handling strategy. The Digital Twin's models used in this strategy should be updated as often as possible to ensure that they properly reflect the current process condition. Another key aspect is to consider intelligibility of the proposed upgrades, ensuring that non-simulation experts may also exploit the prescriptive analyses. Thus, this task will result in a micro-service consisting of simulation models as a service.

The i4Q Manufacturing Line Reconfiguration tools are optimisation microservices that use simulation to evaluate different possible scenarios and propose changes in the configuration of the manufacturing line to achieve improved quality targets. After the proposed configuration parameters are confirmed, these optimisation microservices evaluate the new process output characteristics to validate the success of the optimisation and/or adapt its reasoning rules according to the achieved results. Based on this learning approach, the tools develop strategies for machine parameter calibration, line setup and line reconfiguration in order to increase productivity and reduce the efforts for line reconfiguration through AI, considering both automated approaches and collaboration with humans.

The i4Q Manufacturing Line Data Certification Procedure provides a certification and audit procedure to be applied to the manufacturing resources (machine, cell or manufacturing line) to ensure that the data resulting from the manufacturing processes are accurate and reliable. In addition, it provides recommendations for process reconfiguration, audit strategies, certificates and regulations. The procedure describes the logical sequence of the activities to be performed, elements of the manufacturing resources to be audited (sensors, controls, software, etc.), calibration devices to be used and tests to be performed as well as the frequency with which the procedure is to be performed. This procedure also serves as a basis to complement existing quality certifications (i.e., ISO 9000) introducing as a new factor to consider: the quality of the data generated during manufacturing processes. The procedure addresses: definition and vocabulary, frame and application areas, prerequisites, planning, implementation, controlling, improvement and documentation of data driven qualification, reconfiguration, and quality control.

IV. Discussion and Potential Impact

The i4Q IoT-based Reliable Industrial Data Services (RIDS) focuses on the prescriptive analytics challenge, which entails several smaller challenges, such as close-loop integration between data analytics and simulation processes (in order to bring simulation and digital twin models closest to reality and to capitalise on the insights gathered from such models) and by leveraging data analytics workloads between edge and cloud computing (so as to implement a hybrid cloud/edge computing scheme, to not only exploit the strength of cloud computing to process the complicated tasks but also harness the benefit of edge computing in short latency, consequently obtaining the better performance).

Data quality is also more pervasive thanks to the i4Q Rapid Manufacturing Line Qualification and Reconfiguration tools, in particular in the context of Manufacturing Line Quality Diagnosis and Certification and Auditing. i4Q RIDS covers the wider, strategic and the more narrow, operational aspects of manufacturing data quality with the systematic identification of the various factors that influence data quality in manufacturing. This ranges from the hardware-related aspects, such as measurement system characteristics, to the analytics processes including AI methods, and organisational procedures that drive manual data collection. The i4Q Ontology makes the identified knowledge accessible to organisations and serves as a basis for a public knowledge base about the factors that influence data quality in manufacturing.

Furthermore, the i4Q RIDS enhances blockchain technologies, including placing the necessary smart contracts in place, to enable full control over data. Access to specific data is governed by the blockchain itself, using appropriate smart contracts, by which evidence of the data to enable traceability, provenance and verification is kept in the blockchain itself, providing full control to the data owner. It is supported a direct interaction from the manufacturing devices to the blockchain application to maintain a high level of trust in the data. The blockchain is used as a source of trust and integrity between multiple parties. In addition, there are aspects of privacy and confidentiality, whereas there might be data which only a defined set of participants could be eligible to access, therefore this might require exploration of encryption techniques to be able to grant access based on a given access policies list.

One of the major challenges in Manufacturing Process Qualification is the reconsideration and adaptation of the classical approaches and methods for process validation to the new technologies. This step change would allow manufacturers to automatically collect real time data and use advanced statistical procedures for evaluation of process stability and process performance. To this end i4Q RIDS provides an automated manufacturing process qualification method enable to manage real-time data provided by architectures, algorithms and smart data analytics to faster the continuous process validation after process reconfiguration, monitoring the stability, capability and performance of manufacturing processes.

Although the idea of the Digital Twin (DT) has received considerable attention in the last years, few real cases in manufacturing environment have successfully implemented this concept. The i4Q RIDS integrates the DT concept into existing production lines and legacy components, using it as a driver for line reconfiguration. Among others, the DT is used to obtain virtual sensors that increase available data, to supply Key Performance Indicators of the product quality, and to simulate potential line reconfigurations aimed at correcting deviations on the process. Thus, it will aid in moving towards intelligent automatic production line reconfiguration.

V. Concluding Remarks and Future Work

This paper presents the i4Q RIDS (Reliable Industrial Data Services), integrating a set of 22 i4Q Solutions, targeting the manufacturing sector and aimed at improving the digital manufacturing through more reliable and effective data. It is founded on a unified yet modular Framework, rooted in a consistent Reference Architecture which encompasses the following core layers: physical, network, middleware, database, and application. The i4Q Reference Architecture is based on current standards in manufacturing (e.g., IIRA, RAMI4.0, IDSA, and IMSA) and incorporates all fundamental viewpoints involved in the process: business, usage, functional and implementation. The i4Q RIDS, therefore, aims to support the complete flow of industrial data, starting from data collection to data analysis, simulation and prediction. It provides solutions to ensure data quality, security and trustworthiness, especially tailored for manufacturing, such as blockchain-based data services and distributed storage.

The i4Q RIDS also includes a set of services for data integration and fusion, data analytics and data distribution. Execution of AI workloads (including at the edge) is enabled and effectively managed through dedicated services which enable the dynamic deployment scenarios based on a cloud/edge architecture. Monitoring at various levels is provided in the i4Q RIDS through scalable monitoring tools and the collected monitoring data are used for a variety of activities including resource monitoring and management, workload assignment, smart alerting, predictive failure and model (re)training.

Digital Twins are extensively used, enabling full digitisation of the manufacturing process and providing simulation and modelling capabilities. Digital twins are used for process qualification, in particular, to analyse how process parameters affect final product quality and obtain virtual sensors, as well as to explore potential upgrade actions and extend existing process data. Additionally, digital twins support quality diagnosis of the manufacturing line. Typical process qualification methods are improved in the i4Q RIDS thanks to automated continuous process qualification and the use of real-time data.

In order to facilitate wide and agile deployment, the i4Q RIDS adopts a modular, microservices-based approach, allowing the framework (and individual components) to be adapted and integrated in different manufacturing scenarios, for diverse companies and at varying maturity levels.

Future work will be dedicated to the implementation and validation of the i4Q RIDS into real use cases of different industrial sectors (white goods, wood equipment, metal machining, ceramics pressing, plastic injection and metal equipment), in order to demonstrate the applicability and the impact of the i4Q Solutions and its results in the market environment under real-world conditions, and the creation of a start-up for the commercialisation of the i4Q RIDS.

Acknowledgment

The research leading to this work has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No: 958205.

References

[1] ISO 9000:2015: Quality management systems — Fundamentals and vocabulary.

[2] R. Y. Wang, and D. M. Strong. "Beyond accuracy: What data quality means to data consumers." Journal of management information systems 12.4 (1996): 5-33.

[3] M.J. Eppler. Managing information quality: Increasing the value of information in knowledge-intensive products and processes. Springer Science & Business Media, 2006.

[4] ISO/IEC 25012:2008: Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model

[5] R.Y. Wang "A product perspective on total data quality management." Communications of the ACM 41.2 (1998): 58-65.

[6] I. Taleb, M.A. Serhani, and R. Dssouli. "Big data quality: A survey." 2018 IEEE International Congress on Big Data (BigData Congress). IEEE, 2018.

[7] J. Moyne, and J. Iskandar. "Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing." Processes 5.3 (2017): 39.

[8] J. Lee, H. Davari, J. Singh, and V. Pandhare "Industrial Artificial Intelligence for industry 4.0-based manufacturing systems." Manufacturing letters 18 (2018): 20-23.

[9] L. Liu, and L. Chi. "Evolutional Data Quality: A Theory-Specific View." ICIQ. 2002.

[10] N.H. Robertson, and T. Perera. "Feasibility for automatic data collection." Proceeding of the 2001 Winter Simulation Conference (Cat. No. 01CH37304). Vol. 2. IEEE, 2001.

[11] F. Tao, Q. Qi, A. Liu, and A. Kusiak "Data-driven smart manufacturing." Journal of Manufacturing Systems 48 (2018): 157-169.

[12] C. Batini, and M. Scannapieco. Data Quality: Concepts, Methodologies and Techniques. Data-Centric Systems and Applications. Springer, 2006.

[13] N. Laranjeiro, S.N. Soydemir, and J. Bernardino. "A survey on data quality: classifying poor data." 2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC). IEEE, 2015.

[14] Data Warehousing Institute. "Data quality and the bottom line: Achieving business success through a commitment to high quality data." Retrieved from http://www.dw-institute.com/ Accessed on January 2020

[15] S. H.Bang, R. Ak, A. Narayanan, Y.T. Lee, and H. Cho. "A survey on knowledge transfer for manufacturing data analytics." Computers in Industry 104 (2019): 116-130.

[16] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino. "Methodologies for data quality assessment and improvement." ACM computing surveys (CSUR) 41.3 (2009): 1-52.

[17] K.Q. Wang, S.R. Tong, L. Roucoules, and B. Eynard. "Analysis of data quality and information quality problems in digital manufacturing." 2008 4th IEEE International Conference on Management of Innovation and Technology. IEEE, 2008.

[18] T. Hubauer, S. Lamparter, M. Roshchin, N. Solomakhina, and S. Watson. "Analysis of data quality issues in real-world industrial data." Annual Conference of the Prognostics and Health Management Society. 2013.

[19] E. Rahm, and H.H. Do. "Data cleaning: Problems and current approaches." IEEE Data Eng. Bull. 23.4 (2000): 3-13.

[20] S. Brüggemann, and F. Grüning. "Using domain knowledge provided by ontologies for improving data quality management." Proceedings of I-Know. 2008.

[21] E. Järvenpää, N. Siltala, O. Hylli, and M. Lanz. "The development of an ontology for describing the capabilities of manufacturing resources." Journal of Intelligent Manufacturing 30.2 (2019): 959-978.

[22] V.J. Ramírez-Durán, I. Berges, and A. Illarramendi. "ExtruOnt: An ontology for describing a type of manufacturing machine for Industry 4.0 systems." Semantic Web Preprint: 1-23.

[23] ISO 8000. Data Quality

[24] G. Shao, S.J. Shin, and S. Jain. "Data analytics using simulation for smart manufacturing." Proceedings of the Winter Simulation Conference 2014. IEEE, 2014.

[25] S. Jain, G. Shao, and S.J. Shin. (2017). "Manufacturing data analytics using a virtual factory representation." International Journal of Production Research 55.18 (2017): 5450-5464.

[26] P. Lade, R. Ghosh, and S. Srinivasan. (2017). Manufacturing analytics and industrial internet of things." IEEE Intelligent Systems 32.3 (2017): 74-79.

[27] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C.M. Almeida. (2019). "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions." Journal of cleaner production 210 (2019): 1343-1365.

[28] H.N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran. (2019). "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies." Enterprise Information Systems 14.9-10 (2020): 1279-1303.

[29] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, et al. "Hyperledger fabric: a distributed operating system for permissioned blockchains." Proceedings of the thirteenth EuroSys conference. 2018.

[30] S.A. Abeyratne, and R.P. Monfared. "Blockchain ready manufacturing supply chain using distributed ledger." International Journal of Research in Engineering and Technology 5.9 (2016): 1-10.

[31] A. Ramachandran, and D. Kantarcioglu. "Using blockchain and smart contracts for secure data provenance management." arXiv preprint arXiv:1709.10000 (2017).

[32] H.J. Harrington. "P. American Society for Quality Control Quality, Business process improvement: the breakthrough strategy for total quality, productivity, and competitiveness." (1991).

[33] FDA, Food and Drug Administration. Retrieved from https://www.fda.gov/ Accessed on January 2020

[34] FDA, U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry Process Validation: General Principles and Practices. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/process-validation-general-principles-and-practices Accessed on January 2020

[35] W. Lepuschitz, A. Zoitl, M. Vallée, and M. Merdan. "Toward self-reconfiguration of manufacturing systems using automation agents." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41.1 (2010): 52-69.

[36] Y. Koren, U. Heisel, F. Jovane, T. Moriwaki, G. Pritschow, G. Ulsoy, and H. Van Brussel. "Reconfigurable manufacturing systems." CIRP annals 48.2 (1999): 527-540.

[37] J. Leng, Q. Liu, S. Ye, J. Jing, Y. Wang, C. Zhang, et al. "Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model." Robotics and Computer-Integrated Manufacturing 63 (2020): 101895.

[38] E. Glaessgen, and D. Stargel. "The digital twin paradigm for future NASA and US Air Force vehicles." 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA. 2012.

[39] H. Van Der Auweraer, T. Tamarozzi, E. Risaliti, M. Sarrazin, J. Croes, B. Forrier, B, et al. "Virtual Sensing Based on Design Engineering Simulation Models." ICEDyn2017, Date: 2017/07/03-2017/07/05, Location: Ericeira, Portugal. 2017.

[40] U. Leturiondo, O. Salgado, L. Ciani, D. Galar, and M. Catelani. "Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data." Measurement 108 (2017): 152-162.

[41] M. González, O. Salgado, X. Hernandez, J. Croes, B. Pluymers, W. Desmet. "Model-based condition monitoring of guiding rails in electro-mechanical systems." Mechanical Systems and Signal Processing 120 (2019): 630-641.