Measurement: Sensors 18 (2021) 100206

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

FAIR sensor services - Towards sustainable sensor data management

ARTICLE INFO

Keywords
Internet of Production
FAIR data
Sensor data
Model-based software engineering

ABSTRACT

Sustainable collection, processing and storage of sensor data in industrial context are essential requirements for gaining long-term knowledge on product and process quality. Today, collected data is often stored in arbitrary data formats without appropriate metadata describing the data content and is therefore lost for future reuse because crucial information on how to find, access and interoperate with the data is missing. Moreover, insufficiently described or missing data can lead to wrong decisions. In a scientific context, the FAIR data principles are an emerging design guideline to provide all collected data with rich metadata that allow for finding, accessing, interoperating with and reusing the data effectively later. This publication proposes an approach to implement the FAIR principles for industrial sensor data. It analyses major deficits and challenges of making industrial sensor data FAIR and outlines a system architecture for acquisition of FAIR industrial sensor data, called *FAIR sensor services*.

1. Introduction

Data is the key driver of modern production processes and Industry 4.0 [1]. Powered by overarching infrastructures, such as the Internet of Production (IoP) [2], connecting all available sensors and machines of the production process, decisions with subject to product and process quality are based on substantially large amounts of collected data. To fully realize the potential and benefits of the IoP, sharing data and knowledge is an essential requirement [3]. Nowadays, the use of acquired sensor and process data is shifting from retrospective use for process and part qualification to prescriptive approaches. By utilizing modern tools and algorithms, such as machine learning, reusing sensor data exposes further potential for value creation. A prominent example is monitoring the condition of machining tools to reduce maintenance cost and down time or analysing and improving the overall process quality and efficiency [2]. Following this development, the applications' context of data collection shifts from short-term, local usage to a long-term, global usage increasing the set of involved stakeholders and imposing higher requirements on persistency and interpretability of the collected data.

However, currently collected data is usually stored in arbitrary, often proprietary formats and on local storages. Thus, the data can neither be accessed nor processed by machines to gain knowledge from that data and support decisions [4]. Usually, industrial sensor data lacks consistent interoperable metadata describing the data content and meaning, so that the data often cannot be interpreted and (re-) used, even within big companies. Thus, such poorly documented data can be considered lost [5].

To foster sustainable data aggregation and data management by extensive documentation with appropriate metadata the implementation of the *FAIR Guiding Principles* [6] is required. Documenting and managing data conform to the FAIR principles results in unambiguously interpretable data, increases trust in the data and support all kinds of data-driven applications [7,8]. While the implementation is broadly discussed for research data the applicability and significance in industrial scenarios is largely unexplored. Therefore, this paper gives one of the first analysis of the challenges of implementing the FAIR guiding principles for industrial data. Furthermore, it introduces *FAIR sensor services* as a novel architecture draft to enable the implementation of the FAIR guiding principles and introduces an approach for bringing the concept into practice by model-based software engineering.

Based on scientific literature and practical experience from industry, Section 2 analyses the current situation in industrial data management and derives major challenges for making industrial sensor data FAIR. Section 3 summarizes related work regarding *FAIRification* [9], i.e. processes of making data FAIR, of industrial sensor data. In Section 4 a draft for a system architecture enabling the collection of FAIR sensor data is proposed and approaches for implementing the architecture are introduced. The paper closes with a summary and outlook in Section 5.

2. Deficits and challenges

Major deficits and challenges in the FAIRification of industrial sensor data arise due to the heterogeneity of the data itself and also of the involved stakeholders. This section analyses the key issues related to the technical realisation of a FAIRification process for industrial sensor data.

2.1. Deficits of industrial (meta-) data management

The deficits of industrial (meta-) data management cover all aspects of the data lifecycle, as well as model-related, technical and legal concerns. In the following, three deficits are presented, which turned out being most significant during the authors studies on practical implementation of FAIRification processes in industrial production context.

https://doi.org/10.1016/j.measen.2021.100206

Available online 4 October 2021

D1 *Missing interoperability:* Industrial (sensor) data lacks consistent, agreed terms, formats and protocols. The techniques for data description, storage and transmission depends on the manufacturer of the sensors and the company using it. Thus, data can often not be shared with and understood by a third party [10]. Whereas standardizing formats and protocols can be considered as *technical interoperability*, using controlled terms and vocabulary can be considered as *semantical interoperability* [11].

D2 *Poor contextual embedding:* Collected data usually misses a detailed description of the context within which the data has been created, i.e. the documentation of related entities, such as the process the sensor is integrated in or involved tools and stakeholders [8]. Thus, crucial information regarding the reusability of the data is missing.

D3 *Absence of FAIRness criteria:* The FAIR principles alone are only abstract design guidelines. To implement the FAIR principles for industrial sensor data, quantifiable metrics have to be defined to measure the *FAIRness*. Associated with that, it has to be clarified, which exact information is required for making data findable, interoperable and reusable [12].

2.2. Challenges of the FAIRification process

The challenges of enabling the acquisition of FAIR industrial sensor data are illustrated in a three-step approach: First, the requirements and criteria for FAIR sensor data must be established. Based on that, proper models must be defined to record metadata and achieve a rich documentation of the data. Finally, the models and tools developed must be implemented appropriately in order to be applicable in practice. Whereas the first and second are broadly discussed in literature for scientific data, investigation and conclusions regarding industrial data is only sparsely available. Moreover, bringing the knowledge and methods into practice has not been addressed properly so far [12]. Issues arising related to step three are directly hindering application of developed methods. Hence, developing applicable solutions for bringing FAIRification methods to practice is crucial. For that, the following key challenges must addressed:

C1 *Data heterogeneity:* Industrial data is characterised by high quantity and high heterogeneity. Thus, the metadata associated with the data is also of high heterogeneity. This makes it practically impossible to define a standardised model covering all aspects of the data, which are required to achieve a rich description as demanded by the FAIR principles [10].

C2 *Stakeholder heterogeneity:* The number of potential applications and use cases for industrial sensor data is growing. Accordingly, the group of stakeholders is growing in size and diversity [6]. To maximize findability the metadata must contain appropriate information in an indexable format which stakeholders or algorithms will most probably use when searching for the data. Moreover, the requirements for reusability differ between use-cases and therefore between different stakeholders.

C3 *Unknown (heterogeneous) applications:* At time of data creation, i.e. when the measurement is taken, the group of applications potentially reusing the data is unknown. Generally, one has applications in mind to use data for, but one cannot imagine every possible application the measured data might be used for in future. Hence, the metadata and the format of these required for such applications is also unknown.

C4 *System heterogeneity:* The system architecture and network setup differs from company to company and often even from site to site [10]. Hence, the models and tools developed for a sensor will require manual adaptation at each deployment to be technically and semantically interoperable with the existing architectures. This implies costly, time-consuming and error-prone adaptations.

C5 *Interdisciplinary deployment:* Deployment requires different types of expertise and knowledge from at least two domains [9]: One have to be familiar with the sensor and the process, which should be FAIRified, and with the requirements and process of FAIRification itself. Because the expertise in FAIRification is usually not core business of production companies, external experts have to be consulted [13]. This becomes even more complex, when the deployment covers more issues than FAIRification only.

C6 *Distributed data:* In Industry 4.0 scenarios, a freshly acquired measurement might be untraceably distributed within an interconnected network and to innumerable consumers or storages [14]. Furthermore, data consumption might take place immediately, e.g. within control loops, or at some point later (up month or years). Thus, the FAIRification of the measurement must be happen at edge level before any transmission into a network takes place. Consequently, manual curation and FAIRification of data is not applicable due to the high effort which is associated with the manual process [15]. Consequently, the FAIRification of sensor data must be automatable.

3. Related work

Approaches and literature on FAIRification of industrial (sensor) data are sparse and a comprehensive solution was not found [8]. A generic FAIRification workflow has been proposed by Jacobsen et al. [9]. It suggests a general procedure for making existing data FAIR, by applying tools for assessing the FAIRness of existing data, defining metadata models, collecting metadata and linking those with the data itself. Although this procedure is generally applicable for all kinds of data, there are no implemented tools for the most cumbersome steps. Moreover, the applicability for *FAIRification-by-design*, i.e. the process of making data FAIR at the point of creation, is yet unclear. Prominent tools for FAIRification, like *FAIRifier* [12], are only applicable for the FAIRification of previously collected data, which is not applicable for the case considered here, due to **C6**.

Approaches like *FactDAG* [10] aim to improve interoperability of industrial data by keeping track of data provenance with persistent identifiers and directed-acyclic graphs. With that, information can be semantically linked, however it is not covered what information has to be linked to the measurement data to document the provenance appropriately, leaving **C2** unaddressed.

Technical interoperability of sensor data can be addressed by using standardized data formats, such as JSON or XML, as well as data transmission and communication protocols for Industry 4.0, like MQTT, OPC UA or similar. Nevertheless, the implementation of the protocols and formats has to be done manually for each sensor and each different system architecture, imposing high effort with respect to **C4**. Semantical interoperability can be achieved by using ontologies, offering defined terminology and relations between defined concepts. Ontologies like *Semantic Sensor Network* (SSN) [16] or *IoT-Light* [17] define basic reusable concepts of the domain, but missing some domain-specific terms. Furthermore, defining specific models for a certain sensor comes with high effort and requires deep knowledge of ontology-based modelling often hindered by **C5**.

4. FAIR sensor services

To address the issue of collecting FAIR industrial measurement data, *FAIR Sensor Services* are proposed as essential cornerstone of *Distributed Sensor Services* [14]: Each sensor is considered as service being encapsulated and interchangeable by definition of a standardized and interoperable interface, comparable to the microservice approach from software development. Given the standardized interface, applications consuming the data provided be the sensor service can be easily implemented against the sensor interface, such that those application must not be known in advance, which directly contributes to a solution for **C3** *unknown applications.* Following the service concept, the sensor is also responsible for advertising the offered service: yielding well documented measurement data. Because documentation is significantly provided through FAIRification, the sensor service has to take care of the FAIRification of its measured data, which directly tackles **C6** *distributed* *data*. Furthermore, the service approach comes with a shift in responsibilities. Today the purchase of a sensor usually covers the hardware and a useable low-level programming interface (API), which is then used by experts to manually integrate the sensor into the buyer's system architecture. The attraction and applicability of sensor is usually determined by its technical performance. While today the factory and data owner is responsible for curating the data to be useable for its intended purpose, in the future the sensor's manufacturer or distributor might be responsible for providing a standardized, easy-to-use sensor, which produces well documented data and can be easily integrated into the buyer's systems. With this theoretical concept the challenges **C1** *data heterogeneity* and **C5** *interdisciplinary deployment* are addressed, as the heterogeneity of data is reduced by standardization and the issue of implementing the sensor interface is decoupled from the future application.

However, the most significant contribution of *FAIR sensor services* is the acquisition of interoperable measurement data. For achieving full interoperability both issues, technical and semantical interoperability, must be achieved. To decouple the solution for both and achieve maximum modularity, the software controlling the sensor and providing the communication interface is sliced into a three-layer architecture, as depicted in Fig. 1. Thus, separation-of-concerns can be achieved by dedicating each layer to a specific concern requiring knowledge regarding one domain only. The lowest one, the *sensor layer*, encapsulates the implementation of the sensor itself, i.e. the access to device API provided by the manufacturer and the driver for the physical functioning of the sensor. The second layer, called *FAIRification layer*, augments the data with a rich description of the measurement context and detailed metadata in general to optimize the measurement for findability, semantical interoperability and reusability according to the FAIR principles. The third and outer *network layer*, implements the communication interface with standardized and widely used communication protocols and data formats, realising accessibility and technical interoperability. In the following the modelling and implementation approach for the service architecture are explained in detail.

4.1. Provenance rich modelling

Concerning **C2** *stakeholder heterogeneity* and **C3** *unknown applications* it can be assumed that all relevant applications and stakeholders are concerned with an entity or a case that is directly related to context of the measurement. Thus, the provenance of the data and the context which the data is created in must be documented in detail. An extensive analysis of the context is still to be done, but in essence the major entities of the context are (i) the sensor or measuring system itself, (ii) the product the measured quantity is associated with, (iii) any machine which interacts with the product directly before or at the time of the measurement, (iv) the overall process the sensor, product and machine are integrated in, and (v) if the process is not fully automated, workers involved in the production or measurement process. In a first step, the data can be linked to those entities by using persistent, unique identifiers. To describe the provenance of the sensor data semantically formal ontology capturing the measurement context can be used. Such an ontology can be based on established ones, like the PROV ontology [18], which provides high-level concepts for describing the provenance of entities. The temporal context can be captured by making extensive use of timestamps.

4.2. Implementation approach

To address **C5** *interdisciplinary deployment* when realising the network and FAIRification layer, model-driven software engineering (MDE) is employed. In MDE one uses predefined domain-specific programming or modelling languages (DSLs) with reduced complexity and high specificity. From models or programs written in such a DSL one usually generates an implementation in a general-purpose language

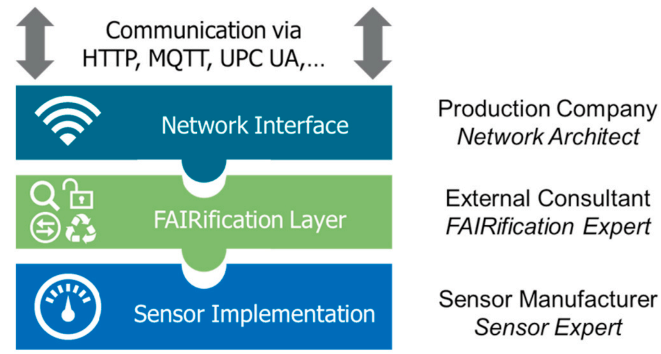

Fig. 1. Three-layer architecture of a FAIR sensor service and the stakeholders responsible for each layer.

(GPL), such as C++, Java or Python. Specific, non-generalizable aspects such as the sensor implementation and hardware access, which differ from device to device, has to be implemented directly in the GPL after generation. In contrast, generalizable parts, like the network and FAIRification layer are included into the DSL and the generation step. Thus, the DSL is developed once by the FAIRification and network expert and then used by the sensor expert to implement FAIR sensor service for all sensors. By including several different communication protocols into the network layer **C4** *system heterogeneity* can be solved effectively. This principle has been proven successfully by implementation of a prototypical modelling language. The *SensOr Interfacing Language* (*SOIL*) [19] decouples the network and sensor layer. Based on unified meta-models the sensor expert defines functional models for a specific sensor from which a basic implementation of the sensor and network interfaces for HTTP, MQTT or OPC UA are generated. By applying the *generation gap pattern* [20], the access to the sensor hardware is then implemented by hand.

5. Conclusion

Sustainable and unambiguously interpretable data is a prerequisite for Industry 4.0. Therefore, this publication proposes the implementation of the FAIR guiding principles for industrial sensor data as enabling key approach. Three primary deficits and six major challenges that arise when dealing with the implementation of a FAIRification process of industrial sensor data are discussed. To address the challenges, the concept of *FAIR sensor services* is proposed. Considered as a counterpart to the service concept of software development, the architectural draft of the sensor service separates the different aspects of the sensor software - hardware access, FAIRification of the data, and communication - into distinct layers to achieve maximum modularity. As enabling technology model-driven software engineering is used. By that complex, generalizable parts like data FAIRficiation are concealed in the generation step of the service implementation process. A preliminary study has shown the applicability of the envisaged implementation approach for the networking layer.

In the future, a detailed analysis on the requirements for FAIR data in an industrial context is planned and the derivation of corresponding domain-specific FAIRness criteria and metrics is envisaged. Furthermore, it is intended to deeply analyse the provenance context of measurement data and develop an appropriate semantic description of the context. Last, the development and implementation of a DSLs using the semantic description model is planned to provide a tool for FAIRification of industrial sensor data in practice.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023

Internet of Production – 390621612 and Project-ID 432233186 – AIMS.

References

[1] M. Reis, G. Gins, Industrial process monitoring in the big data/industry 4.0 era: from detection, to diagnosis, to prognosis, Processes 5 (4) (2017) 35.
[2] J. Pennekamp, et al., Towards an infrastructure enabling the Internet of production, in: 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), IEEE, 2019, pp. 31–37.
[3] S. Jeschke, C. Brecher, H. Song, D.B. Rawat (Eds.), Industrial Internet of Things. Cybermanufacturing Systems. Springer Series in Wireless Technology, Springer International Publishing, Cham, 2018.
[4] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, Industrial Internet of things: challenges, opportunities, and directions, IEEE Trans. Ind. Inf. 14 (11) (2018) 4724–4734.
[5] T. Tanhua, et al., Ocean FAIR data services, Front. Mar. Sci. 6 (2019).
[6] M.D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific data 3 (2016).
[7] M. Rychlik, et al., Ensuring food integrity by metrology and FAIR data principles, Front. Chem. 6 (2018) 49.
[8] M. van Reisen, M. Stokmans, M. Basajja, A.O. Ong'ayo, C. Kirkpatrick, B. Mons, Towards the tipping point for FAIR implementation, Data Intelligence 2 (1–2) (2020) 264–275. Data Intellegence 2, 1-2, 264–275.
[9] A. Jacobsen, et al., A generic workflow for the data FAIRification process, Data Intelligence 2 (1–2) (2020) 56–65. Data Intellegence 2, 1-2, 56–65.
[10] L. Gleim, et al., FactDAG: formalizing data interoperability in an Internet of production, IEEE Internet of Things Journal 7 (4) (2020) 3243–3253.
[11] G. Guizzardi, Ontology, ontologies and the "I" of FAIR, Data Intellegence 2 (1–2) (2020) 181–191.
[12] M. Thompson, K. Burger, R. Kaliyaperumal, M. Roos, L.O.B. da Silva Santos, Making FAIR easy with FAIR tools: from creolization to convergence, Data Intelligence 2 (1–2) (2020) 87–95. Data Intellegence 2, 1-2, 87–95.
[13] F.W. Baumann, U. Breitenbücher, M. Falkenthal, G. Grünert, S. Hudert, Industrial data sharing with data access policy, in: Cooperative Design, Visualization, and Engineering. 14th International Conference, CDVE 2017, Mallorca, Spain, September 17-20, 2017, Proceedings. Lecture Notes in Computer Science 10451, Springer International Publishing, Cham, 2017, pp. 215–219.
[14] M. Peterek, B. Montavon, Prototype for dual digital traceability of metrology data using X.509 and IOTA, CIRP Annals 69 (1) (2020) 449–452.
[15] H. van Vlijmen, et al., The need of industry to go FAIR, Data Intellegence 2 (1–2) (2020) 276–284.
[16] M. Compton, et al., The SSN ontology of the W3C semantic sensor network incubator group, Journal of Web Semantics 17 (2012) 25–32.
[17] M. Bermudez-Edo, T. Elsaleh, P. Barnaghi, K. Taylor, IoT-lite: a lightweight semantic model for the Internet of things, in: UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld 2016, IEEE, Piscataway, NJ, 2016, pp. 90–97.
[18] Timothy Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, Jun Zhao, PROV-O: the PROV Ontology. W3C Recommendation, World Wide Web Consortium, United States, 2013.
[19] M. Bodenbenner, M.P. Sanders, B. Montavon, R.H. Schmitt, Domain-specific language for sensors in the Internet of production, in: Production at the Leading Edge of Technology. Proceedings of the 10th Congress of the German Academic Association for Production Technology (WGP), Dresden, 23-24 September 2020. Lecture Notes in Production Engineering, Springer, Berlin, Heidelberg, 2021, pp. 448–456.
[20] M. Fowler, R. Parsons, Domain-specific Languages. A Martin Fowler Signature Book, Addison-Wesley, Upper Saddle River, NJ, Boston, Indianapolis, San Francisco, New York, Toronto, Montreal, London, Munich, Paris, Madrid, Sydney, Tokyo, Singapore, Mexico City, 2011.

Matthias Bodenbenner*
WZL | RWTH Aachen University, Aachen, Germany

Benjamin Montavon
WZL | RWTH Aachen University, Aachen, Germany

Robert H. Schmitt
WZL | RWTH Aachen University, Aachen, Germany
Fraunhofer Institute for Production Technology IPT, Aachen, Germany

* Corresponding author.
E-mail address: m.bodenbenner@wzl.rwth-aachen.de (M. Bodenbenner).