Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 51 (2020) 1106–1110

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

30th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2021)
15-18 June 2021, Athens, Greece.

Analysis-oriented structure for runtime data in Industry 4.0 asset administration shells

Endre Sølvsberg[a], Christian D. Øien[a], Sebastian Dransfeld[a], Ragnhild J. Eleftheriadis[a]*

Odd Myklebust[a]

[a]*Sintef Manufacturing, Enggata 40, 2830 Raufoss, Norway*

* Corresponding author. Tel.: +47 957 06 599; *E-mail address:* ragnhild.eleftheriadis@sintef.no

Abstract

The need of analysis-oriented structure for runtime data in manufacturing process control units such as PLCs is identified and discussed from an engineering point of view and based on an industrial case. This is evaluated as an important building block for obtaining data from production equipment that can be fully utilized in an Industry 4.0 application. When aligned with an I4.0 asset administration shell, analysis-oriented structuring of runtime data will enable generic, multivariate, inter-asset analysis and allow dealing with unforeseen situations where originally unintended analyses are needed, without an ad hoc effort of data pre-processing. Based on this identified need, we propose the outline of a data structure with focus on column-based formatting, data batches based on process sequence and coordinated sampling rates. Benefits for both data accuracy and data storage architecture are discussed, with the aim of improving the value of analyses of previous events or production periods and increasing the value of I4.0 applications.

Keywords: Industry 4,0; Asset Administration Shell; Multivariate Analysis; Condition Monitoring; Zero Defect Manufacturing; Cyber Physical Production Systems.

1. Introduction

Industrial machines involved in production aiming at Zero Defect Manufacturing are extensively monitored with digital instrumentation and form part of a Cyber-Physical System (CPS) [1]. In an Industry 4.0 (I4.0) perspective, such assets will together with an asset administration shell (AAS), described in the RAMI 4.0 documentation[2], to form an I4.0 component [2-6]. Based on work with condition monitoring in industry cases, we see the importance of purposive structure for runtime data, i.e. process data from sensors, actuators and controllers, that is aligned with the AAS in order to optimize data accuracy and analysis value. We see this as an important building block for an I4.0 transformation to enable new functionality and for optimizing the value and accuracy of the information distributed in the RAMI4.0 framework. Ongoing standardization work related to the AAS is still dealing with a higher level of information, structure and interoperability[3]. For this to bring value concerning operational information and e.g. condition monitoring, we identify the need of structure focused on a lower level from an engineering point of view, as opposed to a more theoretical or laboratory focused perspective. We use an industrial case to exemplify how our proposed analysis-oriented runtime data structure facilitates improved data capture, storage, and a highly detailed data foundation, allowing for more detailed analysis.

This paper is organized as follows: Section 2 summarizes the state-of-the-art context for our viewpoint concerning I4.0

2351-9789 © 2020 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the FAIM 2021.
10.1016/j.promfg.2020.10.155

data structure described in Section 3, together with the industrial case that serves as motivation for this. A discussion is given in section 4 and our conclusions are made in section 5.

2. State of the art

I4.0 is a result of an integration of the Industrial Internet of Things (IIoT) and manufacturing [4, 7]. This integration of IIoT, manufacturing machines and processes requires Cyber-Physical Production Systems (CPPS), where the physical and digital worlds are merged [8]. Data can be considered the heart of IIoT systems. Still, utilization of data is one of the main challenges going towards I4.0, and in a typical factory more than 99% of data collected is discarded [9]. Development towards a Smart Factory requires collecting, analyzing, transporting and storing an increasingly vast amount of data[5]. Until challenges concerning volume, velocity and variety are addressed, companies will typically store data in siloed, unstructured and temporary data lakes. To enable development towards CPPS systems, the data collected must be actionable, and actionability entails consistent formatting, targeting and data structure. Near real-time actionable data can enable advanced analytics, allowing new insights into e.g. anomaly detection, pattern recognition and ultimately system prediction and improved decision-making [10].

The reference architecture model Industrie 4.0 (RAMI 4.0) is meant to describe an asset with enough precision to enable a digital mirror image of the asset[3, 5]. An I4.0 component combines a physical asset with digital information in an administration shell, enabling information flow across the different layers in the hierarchy [11]. A proposed structure for CPS, the 5C architecture, provides a five-step guideline for developing and deploying a CPS in manufacturing. The main functionalities of a CPS can be described as; advanced connectivity ensuring real-time data acquisition from the physical world and information feedback from the digital world and intelligent data management, analytics and computational capabilities that enable construction of the cyber realm [12].

Data acquisition can be described as a somewhat neglected field, compared to plant monitoring and diagnosis, and publications on solutions and strategies for layers above the asset layer are considerably more numerous than publications describing solutions for data acquisition. There is increased understanding concerning the value of data, but there is also a discrepancy between this perspective and the work being done in the dataspace to add value [13]. Developing strategies for monitoring manufacturing systems without considering how to collect and distribute actionable data may be suboptimal [14]. Improving the management of data is crucial when moving towards a Smart Factory, and data must be revalued as a commodity not just as a byproduct [9].

Current automation systems are heterogeneous in that different modules use different control devices, automation network protocols and offer different user information. Runtime data from sensors, actuators and controllers in distributed automation systems are not synchronized, and only selective data is available from specific parts of the system, leading to a lack of ability to identify overall plant status at a specific time and a missing system and synchronized plant perspective [14].

A generic solution to these challenges can be to utilize dataloggers connected to all relevant parts of a real-time network, and listen passively to communication between sensors, actuators and controllers. Precision Time Protocol (PTP) can be applied to synchronize all dataloggers. The OPC Unified Architecture (OPC-UA) can be used to distribute all captured data using a uniform interface [3, 5, 6, 14]. This platform-independent standard (IEC 62541) enables communication between industrial automation systems and systems and can serve as a bridge between off-line engineering and runtime communication of the involved physical and digital resources in the CPPS [3, 15]. An OPC-UA server can integrate data from different dataloggers used in the network, use the OPC-UA information model to enrich the data with semantic information and make the data accessible via standardized protocols such as HTTP or binary protocols for MES and SCADA systems. The semantic information can be sourced from sensors, MES, SCADA, IEC 61131 Engineering tools and using manually added information. This is in line with the perspective presented in the I4.0 Roadmap [11], and allows for the perspectives in DIN 77005-1 and ISO 20140.

Generically, a solution architecture ensures data acquisition from every possible configuration of the plant. Using AutomationML files[6], semantic information can be imported from relevant sources, and proprietary data access can be avoided. The implementation of industrial network protocols like TCP, PROFINET and EtherCAT enables the Network Abstraction Layer to address the heterogeneous and distributed automation systems [14]. Depending on the industrial network protocols being utilized, data traffic can potentially be captured with nanosecond accuracy and synchronization accurate down to 1µs. Using a defined semantic and adding the Unified Modelling Language (UML) notation, data stored can contain information on measured values, real-world interpretation and meaning, synchronized timestamps, data geographic location, type, temperature, active processes, related measurements and so on, both historical and monitored can be displayed and values, identifiers and display names can be reformatted into the environment. A homogeneous information system structure also allows for cross-disciplinary collaboration and value chain information flow [9]. A near real-time information system that maps technical and business processes across disciplines and enables value chain information flow can improve both control and business decision-making [15].

When acquiring process data in distributed systems, there are generally two methods; installing an additional measuring system with sensors for all critical to quality (CTQ) parameters, or instrumentalizing the existing automation system [14]. Today's often heterogeneous automation systems can make the second approach challenging. Capturing Ethernet frames directly, avoids the need for reconfiguring the automation system. Ensuring connectivity and information flow from the ground level through all layers of the system architecture in near real-time can be achieved through three approaches; (I) direct communication, where all sensors and actuators are equipped with active communication modules, enabling IT system connectivity, (II) reprogramming the

PLCs/IPCs to deliver data to the IT system in addition to traditional tasks or (III) by connecting an aggregation point to a gateway inside the control loop [9], negating the need for PLC/IPC reprogramming. Choosing the right approach would depend on a cost/benefit analysis performed at the individual facility. The second alternative is most relevant for greenfield installations, while the third solution is more suited for brownfield application, especially with a considerable amount of PLCs/IPCs to reprogram[6].

3. Proposed structure for runtime data in an industrial case

It is a frequent challenge to analyze data from process control oriented PLCs where data are merely stored for general historical reference, as opposed to a PLC programmed also with focus on data analysis where data are formatted meaningfully and saved at a proper rate [16]. The former typically builds a continuous time series of data leading to large amounts of information that requires separating out interesting segments based on time instead of production process or sequence. Further on, the data in such continuous time series are often compressed by a deadband filter[1]. This reduces accuracy, deteriorates information quality and causes the need of potentially extensive pre-processing of data prior to analysis [10]. Additionally, the data may not be saved with a common timeline, i.e. re-sampling is also necessary prior to multivariate analysis. All of these issues should be avoided, and we see that a prerequisite for I4.0 applications will be to have an analysis focus, programming wise, from the beginning of the process chain in order to define proper data formatting and sampling rates on PLC level. It's also important to have a data analysis focus during designing of ICT solutions for a production chain, so that information loss and unnecessary pre-processing of data is avoided later. Assuring such data structure characteristics seems beneficial in order to optimize accuracy and to enable an AAS to provide useful data for an unforeseen multivariate analysis of data from one or more assets, without the need for an ad hoc work effort with extensive pre-processing.

We propose that a data structure in an AAS would build on the following properties:

- column based data formatting with a common timeline
- small data batches based on process or production sequence as opposed to continuous time series
- a common multiple of sampling rates, so that there exists a lower rate with common time stamps for any two sets of runtime data
- early basic data processing, at asset level, and saving to a compressed format like *parquet*

In practice, this could be facilitated by an IPC connected directly to a PLC acquiring data from one or more assets where the foundation for column-based storage and process-based sequencing is laid by an object-based signal structure defined in the PLC itself. This way, a purposeful format and synchronization is made already at process level. Such data sets are, in turn, more suitable for cross-utilization with other data sources over state-of-the-art network functionality, as it would provide an asset level data structure comparable with the decentralized structure of [14].
In other words, we see in practice that the basis for a functional 5C architecture should be made by ensuring certain technical details at the bottom level [12]. This corresponds with a machine tool centered perspective on data analytics [13].

3.1. Condition monitoring by proxy measurements

Condition monitoring based on secondary or indirect signals is necessary in processes where machine components are unavailable or situated in physical conditions where electrical circuits do not function. Also, in cases where existing instrumentation is being complemented with additional s like vibration sensors placed on casings or otherwise in the surroundings of the production process, indirect process data must be analyzed. In relation to maintenance or condition monitoring, such analyses will by nature be focused on general anomaly detection, since the basis for defining causality is not present[2]. This is also valid for the industrial case described in this paper, where an underlying goal is to find generally transferrable routines for data analysis that are not linked to, or dependent on, explicit understanding of the process at hand. Due to this, usage of AI-methods like Long Short Term Memory are relevant since we do not intend to link the data to a specific process component, characteristic or failure mode [17].

A data structure as described in section 2 becomes valuable in the case of proxy measurements due to the need for multivariate analysis, as one typically needs to monitor more signals, or correlate them to direct measurements, in order to compensate the lost accuracy compared to direct measurements.

3.2. Industry case

In an industrial case, studying degradation of machine components involved in aluminium alloy casting, we work with multivariate analysis of a combination of proxy and direct measurements in order to propose a base for condition monitoring. The investigated components move in liquid aluminium, at temperatures above 600°C, and naturally the possibilities of monitoring are limited. As a supplement to existing sensors, wireless vibration sensors, constituting the proxy measurements in this case, have been installed on interfaces of the machine where temperatures are below 70°C. As in many similar cases of additional instrumentation, the new data have been acquired independently of the original runtime data and data storage system as part of a trial. Therefore, the data have been assembled and re-sampled in a post-processing step in order to have synchronized data points that can be used for multivariate analysis. Due to early deadband compression of the data, the information loss

[1] A logical system that ensures new data samples are only saved in case they represent a significant change since the previous entry.
[2] One will e.g. not try to say exactly which bearing is about to break when monitoring an entire rotor assembly with a vibration sensor, and neither will the data be useful in finding a reason for the wear.

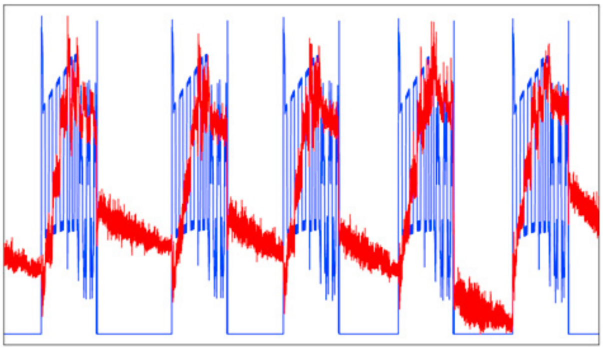

Fig. 1. Time series of signals.

caused by re-sampling and assembling the data is relatively large in our specific analysis. Additionally, the data must be sequenced out from a continuous time series in order to orient them by production sequence such as casting batch in this case, which leads to a specific pre-processing of data that cannot be transferred to other variables or assets in the future.

Figure 1 illustrates the varying signals from a high-volume machine in our industry case. Signals are dependent on the current product and the state of the machine being monitored. For this reason, the patterns are not repetitive. The signals are highly dynamic, so a data rate <<1s is required for optimal signal resolution. The machine being monitored runs 24/7 and produces over 100 million samples per year per variable. 40 variables are considered CTQ. The machine can be stopped and restarted during a process, so a "0" variable value does not necessarily signify a split between products or events.

If we want to look at long time trends in the production, we may want to aggregate the signal or focus on specific variables. Conducting this aggregation before storage runs the risk of incorrect aggregation, and so the experiment must be rerun. With a "Big Data" mindset, the question is not what data we want to save, but rather how much data we are able to save.

In our industrial case, "product" was determined to be a good separator of data. The machine is adjusted for one product at a time. Depending on the product, the process lasts for about 3-4 minutes. One dataset for one product contains 20,000 datapoints, which is considered reasonable. Figure 2 shows a plot of different aggregated values. Having already separated data for each product, the task of creating one data point for each product at a defines point in the process is a simple task. In addition to the variables, each data set contains the state of the process, the id of the product being produced and the process time.

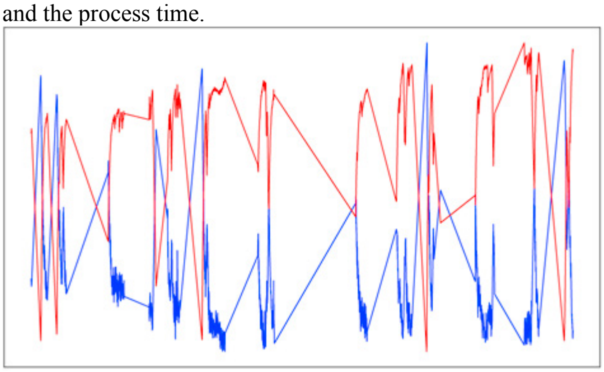

Fig. 2. Aggregated values from time series.

We see that there is a correlation between the signals. In order to be able to analyze this correlation efficiently, all data points must be synchronized, be in pairs. Although it is possible to save storage space by eliminating points with no change, modern columnar data formats like e.g. parquet makes this space-saving negligible. Figure 3 illustrates the correlation between the datapoints shown in Figure 2, with an added color gradient dependent on time. The figure reveals three distinct point clouds; two larger and one smaller in the lower right corner. The smaller cloud describes a serious incident, which affected the stability of the process. Even with a highly detailed data foundation, there are still some outliers. But the data structure makes it simple to locate the data set from where these data comes from, and they can be analyzed separately if required. Based on this incident, it was possible to redo the analysis back in time. Several similar, but much smaller incidents were detected. Without a structured dataset, these incidents would not have been able to be distinguished from the general deviation in the data. In this context, we see the value of a data structure as described in section 2 in the case of implementing I4.0 functionality.

3.3. Benefits

Storage of data for general historical purposes in separate, continuous time series restricts the data's accuracy, analytical value and compatibility with other equipment or systems. In order to be prepared for unforeseen needs of data analysis, implementing a purpose-based formatting and storage of data from PLC level is key in an I4.0 perspective. A data structure based on the fundamental ideas mentioned in section 2 could further on form a basis for a clear architecture, unlike systems with multiple levels of information accuracy, compression and data rates often seen in industry today.

What we see is that a typical process control perspective of PLC programming (Industry 3.0) must be converted into a data analysis perspective in order to create a good base for new solutions and functionality that can be a result of I4.0 transformation.

4. Discussion

Considering a missing analysis-oriented data structure in our industry case, we have proposed what such a structure should contribute to and what prerequisites are needed technically on the asset level for it to be efficient, from an engineering and industry point of view. We see challenges with effectively

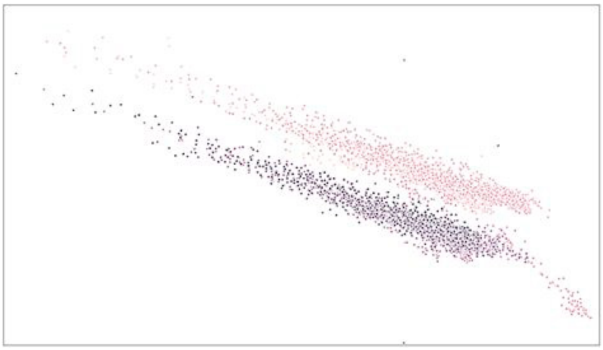

Fig. 3. Data correlation.

defining the data foundation for multivariate anomaly detection without such a structure. Use of proxy measurements especially motivates this.

Another observation in our industry case, and motivation for an analysis-oriented runtime data structure, is that the data used had to be collected from various levels of an enterprise data architecture, where typically each level uses a filter or compression method like sample rate reduction. If common source data were more aligned from the ground up, each compression or filtering step would cause the same effect on multiple signals and this could contribute to reduce loss of data synchronization. Additionally, with a clear focus on data structure when designing such enterprise data architectures, as opposed to general documentation focus making use of historian databases, the complete architecture itself could be simplified. Use of smaller data files based on single process sequences would allow for early compression of data, e.g. by use of the *parquet* file format, which can be done on-the-fly. This would make retrieval of correct data for analysis of previous events easier, and a separate compression step on top level would not be necessary. Compiling a time series out of sequenced files, rather than separating out sequenced data series out of a long time series can save a considerable amount of work,

In terms of standardization, we see the value of dealing with the subject of structure for runtime data as a fundamental building block in I4.0 applications, e.g. for processing of data between assets to lead to improved production performance or new functionality. Standardization on this level may especially be valuable for SME's that could have the automation know-how but lack the analysis competence needed to see the value of a bottom-up definition of the data structure.

Based on the general idea about data structure presented in this article, further work will be made in order to explore ways of implementing such a structure and document its effect in an industry setting.

5. Conclusion

From an engineering point of view, we identify the need for structure for runtime data alignment in manufacturing process control units such as PLC's, in order to obtain data from production equipment that can be fully utilized in an Industry 4.0 application. An analysis-oriented structure communicated through an I4.0 asset administration shell will enable generic multivariate analysis and allow dealing with unforeseen situations where originally unintended analyses are needed, without an ad hoc effort in the data pre-processing. Based on this, we propose the outline of a data structure with focus on column-based formatting, data batches based on process sequence and coordinated sampling rates. We identify benefits of defining such structures already at the PLC level in the case of common source data, improving both data accuracy and data storage architecture, improving the value of analyses of previous events or production periods and increasing the value of I4.0 applications.

Acknowledgements

We wish to thank The Research Council of Norway for their financial contribution to the CPS-Plant project which the R&D work in this article is based upon.

References

[1] R. Baheti and H. J. T. i. o. c. t. Gill, "Cyber-physical systems," vol. 12, no. 1, pp. 161-166, 2011.
[2] *DIN SPEC 91345:2016-04 Reference Architecture Model Industrie 4.0 (RAMI4.0)*, DIN, 2016-04 2016.
[3] H. Bedenbender *et al.*, "Examples of the Asset Administration Shell for Industrie 4.0 Components–Basic Part," *ZVEI White Paper,* 2017.
[4] N. Chilwant and M. S. Kulkarni, "Open asset administration shell for industrial systems," *Manufacturing Letters,* vol. 20, pp. 15-21, 2019.
[5] S. Cavalieri and M. G. Salafia, "Insights into Mapping Solutions Based on OPC UA Information Model Applied to the Industry 4.0 Asset Administration Shell," *Computers,* vol. 9, no. 2, p. 28, 2020.
[6] D. Etz, H. Brantner, and W. Kastner, "Smart Manufacturing Retrofit for Brownfield Systems," *Procedia Manufacturing,* vol. 42, pp. 327-332, 2020.
[7] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," *IEEE,* pp. 3928-3937, 2016.
[8] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & information systems engineering,* vol. 6, no. 4, pp. 239-242, 2014.
[9] IIC, "Smart Factory Applications in Discrete Manufacturing," Industrial Internet Consortium, 2017.
[10] J. Kowal and S. Cramer, "A Practical Way to Get Started in Manufacturing IIoT: Cultivate a "Green Patch" in Your Brownfield," Industrial Internet Consortium, iiconsortium.org, 14.11.2017 2017. [Online]. Available: https://www.iiconsortium.org/pdf/2017-11-14-Cultivate_a_green_patch_in_brownfield_whitepaper.pdf
[11] DIN/DKE, "German Standardisaztion Roadmap Industrie 4.0," DIN e.V., March 2018 2018. [Online]. Available: https://www.din.de/resource/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf
[12] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing letters,* vol. 3, pp. 18-23, 2015.
[13] J. Lenz, T. Wuest, and E. Westkämper, "Holistic approach to machine tool data analytics," *Journal of manufacturing systems,* vol. 48, pp. 180-191, 2018.
[14] F. Pethig, B. Kroll, O. Niggemann, A. Maier, T. Tack, and M. Maag, "A generic synchronized data acquisition solution for distributed automation systems," in *Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, 2012: IEEE, pp. 1-8.
[15] L. Monostori *et al.*, "Cyber-physical systems in manufacturing," *Cirp Annals,* vol. 65, no. 2, pp. 621-641, 2016.
[16] M. Bärring *et al.*, "5G Enabled Manufacturing Evaluation for Data-Driven Decision-Making," *Procedia CIRP,* vol. 72, pp. 266-271, 2018.
[17] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*, 2015, vol. 89: Presses universitaires de Louvain.