

Review

Data Science on Industrial Data—Today's Challenges in Brown Field Applications

Tilman Klaeger * , Sebastian Gottschall and Lukas Oehm 

Fraunhofer Institute for Process Engineering and Packaging (IVV), Division Machinery and Processes, Heidelberger Str. 20, 01189 Dresden, Germany; sebastian.gottschall@ivv-dd.fraunhofer.de (S.G.); lukas.oehm@ivv-dd.fraunhofer.de (L.O.)

* Correspondence: tilman.klaeger@ivv-dd.fraunhofer.de

Abstract: Much research is done on data analytics and machine learning for data coming from industrial processes. In practical approaches, one finds many pitfalls restraining the application of these modern technologies especially in brownfield applications. With this paper, we want to show state of the art and what to expect when working with stock machines in the field. The paper is a review of literature found to cover challenges for cyber-physical production systems (CPPS) in brownfield applications. This review is combined with our own personal experience and findings gained while setting up such systems in processing and packaging machines as well as in other areas. A major focus in this paper is on data collection, which tends to be more cumbersome than most people might expect. In addition, data quality for machine learning applications is a challenge once leaving the laboratory and its academic data sets. Topics here include missing ground truth or the lack of semantic description of the data. A last challenge covered is IT security and passing data through firewalls to allow for the cyber part in CPPS. However, all of these findings show that potentials of data driven production systems are strongly depending on data collection to build proclaimed new automation systems with more flexibility, improved human–machine interaction and better process-stability and thus less waste during manufacturing.



Citation: Klaeger, T.; Gottschall, S.; Oehm, L. Data Science on Industrial Data—Today's Challenges in Brown Field Applications. *Challenges* **2021**, *12*, 2. <https://doi.org/10.3390/challe12010002>

Received: 24 September 2020

Accepted: 20 January 2021

Published: 25 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: industrial communication; industrial informatics; cyber-physical production system; machine to machine communication; OPC UA

1. Introduction

Many studies show the possibilities for cyber-physical production systems (CPPS) in contexts of Industry 4.0 [1]. Major benefits expected include the potential to be self-adaptive to changing environments and product properties and thus should overall produce less waste, but also an improved resilience is expected from these systems [2]. To do research in this area on current production machines in the brownfield and to adapt older systems with these new technologies, interfaces for machines and data transfer need to be created. A lot of projects presented are using the latest machines the most recent technologies or circumventing interaction that is too direct with existing infrastructure by applying Industrial Internet of Things (IIoT) architectures.

Very little is written about working with older machines in brownfield applications and doing data science on data to be collected from those machines in real live production settings and not in the laboratory and learning factories. This paper represents a review of challenges and pitfalls documented in scientific papers as well as our personal perspective and experience with data science in industrial settings and CPPS mainly in packaging and processing machinery, though other industries are very comparable. Some of these findings generalize very well and are also well known when talking to other people doing research on CPPS, Industrie 4.0 and IIoT but only a few are actually written down. Finding appropriate work to cite is not always easy as these contents usually hide in sections like “lessons learned”. Other papers keep silent on technical details for implementations and

often the efforts put into the technical solution are not scientifically published. According to some researchers, these problems may be seen as a technical problem only, but it impedes doing further research and thus is, in our opinion, relevant for people in the field to know and to reference. The aim of this paper is therefore to help people new to the subject to understand the current situation and the challenges to face when interacting with these older machines and to share our experience with the community.

Looking at typical machine life-cycles of ten to sometimes more than twenty years, the situation described in the following parts will change but slowly. Looking at the limitations in current, state-of-the-art automation platforms, many machines developed today will not contain features proclaimed in the scientific community today. Thus, a major focus is on data collection and interfacing with existing machinery including challenges local information technology (IT) infrastructures may provide when transmitting the data. Another topic covered is the expectable data quality and content. The basic steps needed and where to expect challenges are also shown in Figure 1.

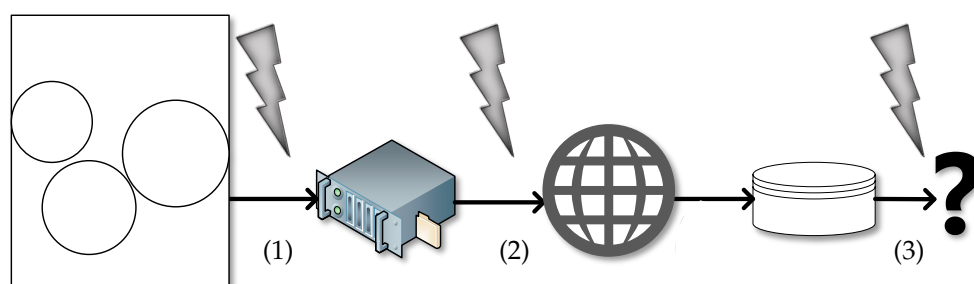


Figure 1. Process steps when doing industrial data science with the machine on the left and the typical way data may flow from some kind of edge device over the internet or cloud to a database. In addition, potential positions for pitfalls are shown: (1): Data acquisition from the machine, (2): Transferring data to the internet/ research institute, (3) understanding and labeling the data.

There are major differences between industrial and academic data science [3]. Some topics noted are currently under intensive research and development, whereas other topics seem to be unresolved at the moment. Various research opportunities are well known [4] and many challenges are still open for brownfield as well as for greenfield applications [5]. In this paper, we will mostly focus on brownfield applications and summarize existing literature as well as personal experience gained in the last years.

2. Raw Data Collection on the Field

A data science project in an industrial context usually starts at the field level to collect data from existing machines and sensors adding some kind of edge or IIoT device. Generally speaking, there are some different approaches from OPC UA, field bus integration to debugging interfaces, and external IoT-Devices to read the data. A combination of those techniques is usually possible but causing more efforts in data acquisition and merging.

2.1. Open Platform Communication Unified Architecture—OPC UA

Many machine controllers (programmable logic controller, PLC) set up in the last years support standard “OPC UA” created by the “OPC Foundation” as machine-to-machine interface. It not only provides a standardized protocol but also adds semantic data models to the data in order to have a description of the data available. Depending on the configuration of the OPC UA server in the controller, selected reading and writing or all values of the PLC is possible. For reading values, not only polling—periodic requesting of values—is possible but also a subscription mode. In this mode, new data will be pushed to the subscribing data acquisition client at the moment the values change. Connecting those to an edge device or even to a complete machine network should therefore be easy. With OPC UA TSN data, transfer over Ethernet will even be possible in hard real time and

thus provide a good solution for time critical applications [6]. The key is the application of “Time Sensitive Networks” (TSN) to warrant configurable time slots just for the transfer of real-time data like OPC UA data.

One major drawback for current applications is the time needed to read the data which has been shown by [7] and correlates well with our own experience. Reading a variable from a Soft-PLC in polling-mode can take up to 20 ms. Sometimes, over 1000 parameters need to be read from the machine. This will severely slow down the process of data acquisition. If this turns out to be too slow, it is possible to build a cache in the controller and read the data as an array. However, this will result in the loss of one of OPC UA’s main features: Semantic description of the data. With good implementations, more performance is possible [8], but you cannot always expect to have those in the machine chosen for the project.

Another major pitfall in OPC UA is the incomplete implementation of the standard in many PLCs [9]. At CERN, this caused engineers to implement automatic testing of OPC UA implementations as there is no guarantee for a complete and bug free implementation of the standard [10]. Polling data seem to be available always, but you cannot expect subscription mode to work. If this is available, one may experience sudden outages with no more data arriving at the data acquisition tool.

This may slowly be changing with newer PLCs being available to machine manufacturers and, after thorough testing, in the field replacing older machines. However, even for the latest machines, in the machine manufacturer’s laboratory, special solutions are built to access the data as current standards are not suitable for all needs [11].

2.2. Integration in Field Bus

If data in higher frequency is needed, then OPC UA can provide it, or if OPC UA is not available, one may try to read the field bus directly. Here, one will face different protocols by different vendors [12,13]. When looking at a manufacturing line, the machines may even use different protocols making the process of data acquisition even more difficult. The proposed solution in many edge devices providing field bus support is to add those devices as an additional slave in the system and have the PLC write the wanted data to this new slave. It therefore needs to provide the same protocol as the existing field bus. For widely used protocols, there are solutions available, but older systems like the fiber-optic based Sercos 2 or 1970s Arcnet solutions are hard to find if available at all.

Depending on the complexity and age of a machine, there is often no possibility to add a new edge device as a slave in the automation bus. Automation environments for programming the PLC changed over the years and the old version may not be compatible with current operating systems. In addition, reprogramming with the control software’s source-code available at the machine manufacturer may overwrite minor but crucial improvements or fixes done in the field. This is completely independent of adding any additional, untested logic. It seems not to be uncommon for a service technician to alter the machine code slightly to make up for some local problem and not always return the changes to the company’s server [14]. Another source of diverging code bases is working with an internal code platform and adapting the resulting code to each customer. Adding, even minor, features may make changes to old machines costly when an update to a new code base is needed [15]. Source code versioning systems common in the software development are only slowly evolving in current PLC programming environments. Especially for graphic programming from the IEC 61131 PLC languages version control with concurrent versions and merge strategies is a complex task [14]. Thus, the usage of those techniques seems to be even more limited in a practical approach. Thus, accessing the data, one often has to cope with “No changes to a running system” issued by the PLC programmer for a fair reason.

However, one option to access the field bus is to build “sniffers” that capture traffic in a passive mode on the bus and decode the signals in an edge device. Solutions like this can be purchased for some systems (to our knowledge for ProfiNet and EtherCAT),

others worked on Profibus [16]. To capture Modbus RTU, a 1970s protocol still widely used in the field, we developed a sniffer ourselves. Off-the-shelf solutions may be difficult to build as there are always surprising solutions in the practice like usage of a bidirectional point-to-point RS422 connection instead of the more common asynchronous RS485 for Modbus data transfer. Building a suitable gateway, it is possible to create a Modbus to OPC UA mappings [17] to address some brownfield issues especially in the area of semantics.

Nevertheless, the accessible data are implementation dependent. Notifications about machine faults e.g., may be sent over the field bus from PLC to the panel showing the human-machine interface (HMI). In some integrated systems, however, this may not be the case if the panel and controller are one device. In other settings, the HMI may be connected to the PLC with some other interface than the field bus. Before implementation, a thorough analysis of the available automation solution and its interfaces is needed.

2.3. Using the Debugging Interface of the PLC

Most PLCs have a so-called “Online Mode” during programming. What is available for development may also be used to read out data in production if the protocol is known.

With Siemens PLCs, this is sometimes referred to as S7 protocol with a popular open source implementation called “Snap7”. Reading out the memory (process image) of a PLC is not difficult using this method [18]. However, accessing the memory more or less raw requires to have a mapping available. This in turn requires cooperation with the programmer of the machine controller. Even Siemens’ own “MindConnect Nano” designed for brownfield data acquisition needs to know the addresses of the variables [19]. Depending on the project, this may lead to conflicts of interest as some manufacturers are not able or not willing to share any details about their control program with their customers. No regularities to whom the data belong that is created in the machine are available at the moment. For old machines, the controller code or tools to view the code may not be available and thus getting access to the required mapping is impossible even if the machine manufacturer is willing to support the project.

If the addresses of the variables are not specifically set during programming, these may easily change on software updates or largely differ between various machine revisions. Thus, this will not result in a plug and play solution for effortless adoption to different machines.

Similar to Siemens S7 protocol is Beckhoff PLC’s support for “ADS”. Reading out the names of variables is possible, provided you can set up an “ADS Route” which requires administrator permissions in the PLC. Gaining this permission again depends on a trustful collaboration with the manufacturer or programmer of the machine. The reading speed is not fast enough for high frequency data as each reading takes 1 ms to 4 ms response time and additional, unpredictable time for the Ethernet data transfer [20].

Smaller PLC vendors may use different protocols. Those are rarely documented, so available features are unknown and predicting the transfer speed is impossible. Hence, reading the debugging interfaces on those PLCs is mostly not an option in one-time projects as it requires reverse engineering the protocol. Implementation seems only worth it when adaption of the system is planned for many machines.

2.4. Third Party Device as Sensor Interface

Another option to read data in the field is to set up a dedicated IIoT structure with devices connected directly to existing sensors or adding new ones [21,22]. This provides major benefits as one has complete control over data formats and protocols being used to transfer the data. However, on the other hand, much data can not be acquired at all: There is no possibility to read internal states of the PLC or errors detected by the PLC. In addition, reading out error codes, motor currents, or lag errors are difficult if not impossible.

Some sensors like PT100 temperature sensors connected in a three- or four-wire circuit or sensors like resistance strain gauge connected with a bridge circuit can not be read out by

two analog-digital converters at the same time without major efforts with custom designed circuits.

2.5. Data Collection from Business Intelligence

Another source one might want to access is the area of business intelligence with data from manufacturing execution systems (MES) or enterprise resource management tools (ERP) like SAP. A lot of these tools may be called open, but what is deployed in the field is highly heterogeneous following different standards [5]. Common standards are REST-APIs providing the data in Java Script Object Notation (JSON) over simple HTTP-Protocol. This is easy to read out, but of course requires adjusting the own data acquisition tools to work with the API. In a similar manner, data can be transferred over TCP/IP servers in XML-format.

Thus, implementing tools to read out and analyze this data is technically not difficult but requires lots of manual work. Sometimes, additional licenses also have to be bought in order to open specific external interfaces.

2.6. Summarizing Data Collection

Each of these solutions has its benefits and drawbacks need to be carefully considered before setting up a system for data acquisition in brownfield applications. In most cases, some data will not be readable with reasonable effort and costs. Efforts are put towards standardization, but, for current projects, one has to deal with tedious work not bringing scientific progress in order to acquire data for research.

3. Data Quality

3.1. Lack of Ground Truth to Train Models

Producing lots of (raw) data is easy in industrial processes [23]. What is much more difficult is obtaining a reasonable ground truth to build machine learning models, which is addressed less frequently [24]. In particular, classification data for fast running processes are difficult to produce [25]. Different methods used are very suitable for research, but, in our opinion, are not fully adoptable to live industrial processes.

One approach is to provoke longer lasting faults and classifying the resulting data [26–28]. Not looking for classification but for anomalies, many pieces of data are needed, which can still be produced in a local experiment, as shown by [11]. Building models to detect longer lasting anomalies like wear or malfunctions in bearings is much different. Looking at steadier processes than discrete manufacturing, some problems are easier to solve, even though more machines need to be available for a proper modeling [29,30]. Detection anomalies in discontinuous processes as against steady ones is not much easier than classification as ground truth often needs to be determined for single products to validate the generated models.

When working with continuous or batch processes like food or chemical production, one may be able to work on a lot of historical data. Data-Logging in this area is much more common already since data need to be available in a central control room [31,32]. Data in such processes come at a slower rate than those originating in discrete processes. If data are used to detect rare and hard to find events, the ground truth may also not be available. One such event is the so-called fouling which will produce stains especially in heat exchangers of industrial processes. If this occurs, operators notice it has happened with some time delay not knowing when it actually got severe. Detecting such an event is hard and thus application of new methods like machine learning seems suitable. Training those models is hard as no secured ground truth is available to accompany the historic data.

One way to build a ground truth may be to manually watch the process. This has some severe drawbacks: First, it is very labor intensive to monitor the process, especially if events do not occur very frequently. On other processes, the processing speed is simply too high for a human to monitor manually [25].

Building models to predict events in the (near) future has one severe benefit as opposed to classification or anomaly detection in historic data: Annotation of training data will

happen automatically based on events happening, at least as long as the events can be detected and recorded. This technique can of course be used for weather forecasts but has also been proven to work in industrial settings like production of plastic films [33].

When setting up projects, the focus at an early state should therefore look towards the ground truth data acquisition.

3.2. Obtainable Training Data Quality

A major concern for all models and especially machine learning models is generalization: “Will the model work on new data as well as on collected training data?” One key issue are spurious relationships leading to over-fitted models [34]. During training, those models “memorize” all correlations independent of causality and thus perform well on the provided data but not future data. Statistically, it can easily be proved that storks deliver babies [35]. There is no causality behind this finding, but the statistical model does show this clearly. For offline-learning, applying a good design of experiments is possible. This allows for creating statistical independent data for training of machine learning models. For collecting live industrial data, controlling the process is not always possible and thus the knowledge over generated data is smaller and may contain various spurious correlations that are hard to find and eliminate.

Closely related to this issue are concept drifts. In this case, the surrounding effects, not being part of the model, change over time. Causes for this may be manual adjustments to the machine or wear in measuring equipment or electrical drives. In the data collected for [25], we could notice such behavior: The model working well on data for one month did not generalize for data of another month. Using different features and adjusting the model provided a better model. However, it still can not be guaranteed to work in the future.

Determining the difference between a model drift or over-fitting on spurious correlations is one of the most difficult tasks in data science. This turns out to be even harder if one is not in complete control over the process being monitored and modeled, whereas there are many ideas on handling concept drift [36,37], finding spurious correlations that can only be done with process and data understanding.

Other challenges are not only common in industrial processes but are for all measured data: Noise in the data and measurement errors like sensors with incorrect or no reading at all. Furthermore, one often has to cope with imbalanced data looking at many well produced parts and having only very few with defects.

4. Semantic Description of Available Data

Much research effort is put into semantic description of data to make it machine-readable and especially -understandable. Efforts are going in different directions using linked data approaches [38,39], description of devices like the Industry 4.0 reference architecture (RAMI 4.0) and similar device description approaches [40,41], and various other ways to describe sensors and connected machines [42–44]. The most promising seems to use semantics already designed in OPC UA. It offers a variety of possibilities to annotate the data for special purposes like in application-to-application interfaces [45]. Even more valuable for future applications are so-called “Companion Standards”. Their aim is to provide a standardized data model for many applications. Models like this are not new, like the “Weihenstephaner Standards” to provide for production data acquisition (PDA) in food processing plants, and are implemented in many machines [46]. For many applications, companion standards are works in progress [45]. Many other data formats available including Modbus are transferred to OPC UA data models [17,47]. All of these works can help to reduce the tedious manual work needed e.g., for feature selection in machine learning applications [48,49].

Going one step further, semantic description and process modeling could be used for an easier if not automated feature engineering. Some process modeling techniques are even sophisticated to generate PLC code automatically [50–52]. All of these ideas are not

new, but, in particular, discrete manufacturing is very different in terms of process models opposed to continuous processes like chemical production [31].

In practical applications, these ideas are not in use yet. PLC programming is mostly manual work and the PLC programmer is supplied with various information sources, mainly informal ones [53,54]. Process models are mostly not used when designing control programs. The topic of incomplete OPC UA implementations arises again when looking to add semantic description to the data model in the automation framework. Tools provided today only have limited possibilities to create OPC UA data models, and more sophisticated models are mostly not available. Consequently, one will face a number of more or less systematically named variables that need to be mapped to the targeted solution.

5. IT Infrastructure

Heavily depending on IT infrastructures, IT security has gained an increasing amount of attention even at the field level. A common technology is network segmentation using virtual local area networks (VLAN). With this, it is possible to assign each network jack or device a separate network independent of the actual network switch to which it is connected. These separate networks can be inter-connected by simple routers or more sophisticated firewalls depending on the specific need. Some companies even go one step further having completely split networks for administrative tasks, machines, and guests at their facility with separate internet connections for each network.

IT security at an infrastructure level is definitely needed as PLCs can have security flaws [55,56]. Software updates may mitigate the risk of security incidents but increase the risk of control software malfunctions. In particular, safety relevant areas proven to work or even certified solutions are not easy to update. When using SoftPLCs, as many vendors offer, the PLC is running on a host operating system in a special real-time setup. Having the control system based on widely available operating systems brings features like easy integration into corporate networks and possibilities to install custom software for instance to log process data. Due to the update risks mentioned earlier, the risk of security incidents is high. In addition, more integrated control hardware is often closely enmeshed with embedded operating systems like Microsoft Windows CE [57]. IT security has to be in focus, but, nevertheless, this can severely slow down data science projects when deploying local edge devices or IIoT hardware and connecting it to the internet. For our own research projects attached to existing infrastructure, the security of own devices has to be kept in mind.

Some research is done in the area of IT security with the aim to provide guidelines for infrastructures suitable for modern cyber-physical systems [58,59]. In practice, one will find different setups at every location to be coped with.

One major and lasting trend in IT is outsourcing. Increasing complexity of IT systems and moving resources to the cloud are causing many companies to have centralized or outsourced IT departments. This may conflict with the need for flexible solutions on location. Data acquisition is often needed in different areas: field level for sensors data up to the ERP to collect data about the currently manufactured product [60]. Sending data to external cloud storage may be even more cumbersome and, based on our experience, may produce serious slow downs in the project.

In the end, this may result in special solutions, and we experienced this ourselves more than once: Consumer-style networks using consumer-style routers for internet access. To have some freedom to operate, these are sometimes even unknown to the central IT department.

6. Conclusions

Doing data science on data collected in industrial processes requires a lot of tedious manual work at the moment. It is not only the often proclaimed “80% of the time needed for data preparation” but it also easily extends by the time needed to just collect the data to prepare. There are exciting new technologies that can be integrated into future

control software solutions but are not available at the moment. Looking at typical lifetimes of machines, these technologies will only emerge slowly. To obtain the benefits like a transition to more flexible production, adaptive and resilient systems, proclaimed in the fourth industrial revolution and its CPPS, research can not only be limited to learning factories but also live production with all the hassles shown in this paper.

The difference in life cycles for software products and machines cast in iron and steel will be a major challenge even in the future. So-called retrofits may be a solution and are provided by machine producers already to support their service business. Looking at scientific projects that want to push one step further while waiting for the next retrofit is not always possible. Not wanting to wait to apply machine learning and other data scientific approaches to discrete manufacturing processes, one has to cope with the challenges. Thus, whenever setting up projects aiming at real data, we recommend planning enough time for the simple but much time-consuming task of data acquisition.

From a software and automation perspective, modular micro service architectures with the possibility to add new, special data acquisition modules for every different machine may be the way to go at the moment. With a higher penetration of sufficient OPC UA servers, this task will become easier. Until then, one has to expect pitfalls during implementation. Thinking in terms of Open Data, Open Science, and Open Source Software, a possible way to go could be a common framework for machine interfacing and data science. We developed such a framework, which has yet to prove applicable to more projects.

From an IT security perspective, companies need to find a secure but flexible way to integrate new cyber-physical systems more easily. Special machine networks with limited but quick access to the internet and outbound-only traffic for safety might be a solution to circumvent current hassles.

Funding: This particular paper received no external funding but is based on various research-projects funded by the German Federal Ministry of Education and Research, the German Federal Ministry for Economic Affairs and Energy, Development Bank of Saxony, the European Regional Development Fund and various other smaller fundings.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, X.; Goepp, V.; Siadat, A. Concept and Engineering Development of Cyber Physical Production Systems: A Systematic Literature Review. *Int. J. Adv. Manuf. Technol.* **2020**, *111*, 243–261. [\[CrossRef\]](#)
2. Monostori, L.; Kádár, B.; Bauernhansl, T.; Kondoh, S.; Kumara, S.; Reinhart, G.; Sauer, O.; Schuh, G.; Sihn, W.; Ueda, K. Cyber-physical systems in manufacturing. *CIRP Ann.* **2016**, *65*, 621–641. [\[CrossRef\]](#)
3. Menzies, T.; Bird, C.; Zimmermann, T.; Schulte, W.; Kocaganeli, E. The Inductive Software Engineering Manifesto: Principles for Industrial Data Mining. In Proceedings of the International Workshop on Machine Learning Technologies in Software Engineering, Lawrence, KS, USA, 12 November 2011; pp. 19–26. [\[CrossRef\]](#)
4. Bordeleau, F.É.; Mosconi, E.; Santa-Eulalia, L.A. Business Intelligence in Industry 4.0: State of the Art and Research Opportunities. In Proceedings of the Hawaii International Conference on System Sciences, Hilton Waikoloa Village, HI, USA, 3–6 January 2018. [\[CrossRef\]](#)
5. Wang, J.; Zhang, W.; Shi, Y.; Duan, S.; Liu, J. Industrial Big Data Analytics: Challenges, Methodologies, and Applications. *arXiv* **2018**, arXiv:1807.01016.
6. Gogolev, A.; Braun, R.; Bauer, P. TSN Traffic Shaping for OPC UA Field Devices. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki-Espoo, Finland, 23–25 July 2019; Volume 1, pp. 951–956. [\[CrossRef\]](#)
7. Wazny, T. *Configuration and Performance Test of TwinCAT 3 OPC-UA Server on Beckhoff IPC*; Technical Report; CERN: Geneva, Switzerland, 2015.
8. Profanter, S.; Tekat, A.; Dorofeev, K.; Rickert, M.; Knoll, A. OPC UA versus ROS, DDS, and MQTT: Performance Evaluation of Industry 4.0 Protocols. In Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia, 13–15 February 2019; pp. 955–962. [\[CrossRef\]](#)

9. Weyer, S.; Schmitt, M.; Ohmer, M.; Gorecky, D. Towards Industry 4.0—Standardization as the Crucial Challenge for Highly Modular, Multi-Vendor Production Systems. *IFAC Pap. Online* **2015**, *48*, 579–584. [[CrossRef](#)]
10. Farnham, B. Automated testing of opc servers. In Proceedings of the 13th International Conference on Accelerator and Large Experimental Physics Control Systems, Grenoble, France, 10–14 November 2011; pp. 985–988.
11. Kammerer, K.; Hoppenstedt, B.; Pryss, R.; Stöckler, S.; Allgaier, J.; Reichert, M. Anomaly Detections for Manufacturing Systems Based on Sensor Data—Insights into Two Challenging Real-World Production Settings. *Sensors* **2019**, *19*, 5370. [[CrossRef](#)] [[PubMed](#)]
12. Mahalik, N.P.; Yen, M. Extending Fieldbus Standards to Food Processing and Packaging Industry: A Review. *Comput. Stand. Interfaces* **2009**, *31*, 586–598. [[CrossRef](#)]
13. Bader, S.R.; Wolff, C.; Vössing, M.; Schmidt, J.P. Towards Enabling Cyber-Physical Systems in Brownfield Environments: Leveraging Environmental Information to Derive Virtual Representations of Unconnected Assets. In *Exploring Service Science*; Satzger, G., Patrício, L., Zaki, M., Kühl, N., Hottum, P., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 331, pp. 165–176. [[CrossRef](#)]
14. Khudyakov, P.Y.; Kisel'nikov, A.Y.; Startcev, I.M.; Kovalev, A.A. Version Control System of CAD Documents and PLC Projects. *J. Phys. Conf. Ser.* **2018**, *1015*, 042020. [[CrossRef](#)]
15. Lettner, D.; Angerer, F.; Prähofer, H.; Grünbacher, P. A Case Study on Software Ecosystem Characteristics in Industrial Automation Software. In Proceedings of the 2014 International Conference on Software and System Process, ICSSP, Nanjing, China, 26–28 May 2014; ACM: New York, NY, USA, 2014; pp. 40–49. [[CrossRef](#)]
16. Mamo, F.T.; Sikora, A.; Rathfelder, C. Legacy to Industry 4.0: A Profibus Sniffer. *J. Phys. Conf. Ser.* **2017**, *870*, 012002. [[CrossRef](#)]
17. Tunkkari, J. Mapping Modbus to OPC Unified Architecture. Master's Thesis, Aalto University, Espoo, Finland, 2018.
18. Zheng, B.; Xu, J.; Li, H.; Xing, J.; Zhao, H.; Liu, G. Development of Remotely Monitoring and Control System for Siemens 840D SI NC Machine Tool Using Snap 7 Codes. In Proceedings of the 2017 2nd International Conference on Electrical, Automation and Mechanical Engineering, Shanghai, China, 23–23 April 2017; Atlantis Press: Paris, France, 2017. [[CrossRef](#)]
19. Siemens, A.G.; Division Digital Factory. *MindSphere with MindConnect Nano and MindConnect IoT2040*; Getting Started: Nürnberg, Germany, 2017.
20. Beckhoff GmbH & Co KG. *Fieldbus Networks Workshop—Performance Comparison*; Workshop Presentation; Beckhoff GmbH & Co KG: Verl, Germany, 2003.
21. Strauß, P.; Schmitz, M.; Wöstmann, R.; Deuse, J. Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1474–1483. [[CrossRef](#)]
22. Burggräf, P.; Wagner, J.; Koke, B.; Manoharan, K. Sensor Retrofit for a Coffee Machine as Condition Monitoring and Predictive Maintenance Use Case. In Proceedings of the Wirtschaftsinformatik, Siegen, Germany, 23–27 February 2019; pp. 62–66.
23. Oliveira, M.; Afonso, D. Industry Focused in Data Collection: How Industry 4.0 Is Handled by Big Data. In Proceedings of the 2019 2nd International Conference on Data Science and Information Technology, DSIT 2019, Seoul, Korea, 19–21 July 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 12–18. [[CrossRef](#)]
24. Obdenbusch, M. Referenzarchitektur für Cloudbasiertes Condition Monitoring am Beispiel von Verpackungsmaschinen. Ph.D.-Thesis, RWTH Aachen University, Aachen, Germany, 2018, ISBN 978-3-86359-573-9.
25. Klaeger, T.; Schult, A.; Oehm, L. Using Anomaly Detection to Support Classification of Fast Running Packaging Processes. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN'19), Helsinki-Espoo, Finland, 23–25 July 2019; Volume 1, pp. 343–348. [[CrossRef](#)]
26. Zurita, D.; Carino, J.A.; Picot, A.; Delgado, M.; Ortega, J.A. Diagnosis Method Based on Topology Codification and Neural Network Applied to an Industrial Camshaft. In Proceedings of the 2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Guarda, Portugal, 1–4 September 2015; pp. 124–130. [[CrossRef](#)]
27. Klaeger, T.; Schult, A.; Majschak, J.P. Lernfähige Bedienerassistenz für Verarbeitungsmaschinen. *Industrie 4.0 Management*; GITO mbH Verlag für Industrielle Informationstechnik und Organisation: Berlin, Germany, 2017; Volume 33, pp. 25–28.
28. Brecher, C.; Obdenbusch, M.; Buchsbaum, M. Optimized State Estimation by Application of Machine Learning. *Prod. Eng.* **2017**, *11*, 133–143. [[CrossRef](#)]
29. Smart, E.; Brown, D.; Axel-Berg, L. Comparing One and Two Class Classification Methods for Multiple Fault Detection on an Induction Motor. In Proceedings of the 2013 IEEE Symposium on Industrial Electronics Applications, Taipei, Taiwan, 28–31 May 2013; pp. 132–137. [[CrossRef](#)]
30. Rapur, J.S.; Tiwari, R. Experimental Time-Domain Vibration-Based Fault Diagnosis of Centrifugal Pumps Using Support Vector Machine. *ASCE ASME J. Risk Uncertain. Eng. Syst. Part A Mech. Eng.* **2017**, *3*, 044501. [[CrossRef](#)]
31. Mersch, H.; Behnen, D.; Schmitz, D.; Epple, U.; Brecher, C.; Jarke, M. Gemeinsamkeiten Und Unterschiede Der Prozess- Und Fertigungstechnik. In *AT Automatisierungstechnik*; De Gruyter: Berlin, Germany, 2011; Volume 59. [[CrossRef](#)]
32. Müller, R.; Oehm, L. Process Industries versus Discrete Processing: How System Characteristics Affect Operator Tasks. *Cogn. Technol. Work.* **2019**, *21*, 337–356. [[CrossRef](#)]
33. Kohlert, M. Multi-Sensory Data Analysis and On-Line Evaluation for Advanced Process Control and Yield Optimization in Polymer Film Industry. Ph.D. Thesis, Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2015.

34. L'Heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A.M. Machine Learning with Big Data: Challenges and Approaches. *IEEE Access* **2017**, *5*, 7776–7797. [\[CrossRef\]](#)
35. Matthews, R. Storks Deliver Babies ($P = 0.008$). *Teach. Stat.* **2000**, *22*, 36–38. [\[CrossRef\]](#)
36. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under Concept Drift: A Review. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 2346–2363. [\[CrossRef\]](#)
37. Webb, G.I.; Lee, L.K.; Petitjean, F.; Goethals, B. Understanding Concept Drift. *arXiv* **2017**, arXiv:1704.00362.
38. Graube, M.; Pfeffer, J.; Ziegler, J.; Urbas, L. Linked Data as Integrating Technology for Industrial Data. In Proceedings of the 2011 14th International Conference on Network-Based Information Systems, Tirana, Albania, 7–9 September 2011; pp. 162–167. [\[CrossRef\]](#)
39. Folmer, J.; Kirchen, I.; Trunzer, E.; Vogel-Heuser, B.; Pötter, T.; Graube, M.; Heinze, S.; Urbas, L.; Atzmüller, M.; Arnu, D. Big Und Smart Data. *ATP Ed.* **2017**, *59*, 58–69. [\[CrossRef\]](#)
40. Smart Manufacturing—Reference Architecture Model Industry 4.0 (RAMI4.0). Available online: <https://webstore.iec.ch/publication/30082> (accessed on 25 January 2021).
41. Gössling, A. Device Information Modeling in Automation—A Computer-Scientific Approach. Ph.D.Thesis, TU Dresden, Dresden, Germany, 2014.
42. Bunte, A.; Diedrich, A.; Niggemann, O. Integrating Semantics for Diagnosis of Manufacturing Systems. In Proceedings of the 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6–9 September 2016; pp. 1–8. [\[CrossRef\]](#)
43. Nilsson, J.; Sandin, F. Semantic Interoperability in Industry 4.0: Survey of Recent Developments and Outlook. In Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), Porto, Portugal, 18–20 July 2018; pp. 127–132. [\[CrossRef\]](#)
44. Dibowski, H.; Ploennigs, J.; Wollschlaeger, M. Semantic Device and System Modeling for Automation Systems and Sensor Networks. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1298–1311. [\[CrossRef\]](#)
45. Graube, M.; Hensel, S.; Iatrou, C.; Urbas, L. Information Models in OPC UA and Their Advantages and Disadvantages. In Proceedings of the 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 12–15 September 2017; pp. 1–8. [\[CrossRef\]](#)
46. Flad, S.; Weißenberger, B.; Chen, X.; Rösch, S.; Voigt, T. Automatische Generierung von Fertigungs-Managementsystemen. In *Handbuch Industrie 4.0 Bd.2: Automatisierung*; Vogel-Heuser, B., Bauernhansl, T., ten Hompel, M., Eds.; Springer Reference Technik; Springer: Berlin/Heidelberg, Germany, 2017; pp. 349–368. [\[CrossRef\]](#)
47. Seilonen, I.; Vyatkin, V.; Atmojo, U.D. OPC UA Information Model and a Wrapper for IEC 61499 Runtimes. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN'19), Helsinki-Espoo, Finland, 23–25 July 2019; Volume 1, pp. 1008–1013. [\[CrossRef\]](#)
48. Ringsquandl, M.; Lamparter, S.; Brandt, S.; Hubauer, T.; Lepratti, R. Semantic-Guided Feature Selection for Industrial Automation Systems. In *The Semantic Web—ISWC 2015*; Arenas, M., Corcho, O., Simperl, E., Strohmaier, M., d'Aquin, M., Srinivas, K., Groth, P., Dumontier, M., Heflin, J., Thirunarayan, K., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2015; Volume 9367, pp. 225–240. [\[CrossRef\]](#)
49. Diedrich, C.; Hadlich, T.; Thron, M. Semantik durch Merkmale für Industrie 4.0. In *Handbuch Industrie 4.0 Bd.2*; Vogel-Heuser, B., Bauernhansl, T., ten Hompel, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 417–432. [\[CrossRef\]](#)
50. *Formalisierte Prozessbeschreibung—Informationsmodell*; VDI/VDE-Richtlinie VDI/VDE 3682; Verein Deutscher Ingenieure: Düsseldorf, Germany, 2015.
51. Fay, A.; Scholz, A.; Hildebrandt, C.; Schröder, T.; Diedrich, C.; Dubovy, M.; Wiegand, R.; Eck, C.; Heidel, R. Semantische Inhalte Für Industrie 4.0. *ATP Ed.* **2017**, *59*, 34. [\[CrossRef\]](#)
52. Arroyo, E.; Schulze, D.; Christiansen, L.; Fay, A.; Thornhill, N.F. Derivation of Diagnostic Models Based on Formalized Process Knowledge. *IFAC Proc. Vol.* **2014**, *47*, 3456–3464. [\[CrossRef\]](#)
53. Colla, M.; Leidi, T.; Semo, M. Design and Implementation of Industrial Automation Control Systems: A Survey. In Proceedings of the 2009 7th IEEE International Conference on Industrial Informatics, Cardiff, UK, 23–26 June 2009; pp. 570–575. [\[CrossRef\]](#)
54. Holowenko, O. *Assistenz der Steuerungsentwicklung Produktionstechnischer Anlagen*; InnoTeam-Zeitung; TU Dresden: Dresden, Germany, 2017.
55. Karnouskos, S. Stuxnet Worm Impact on Industrial Cyber-Physical System Security. In Proceedings of the IECON 2011—37th Annual Conference of IEEE Industrial Electronics, Melbourne, Australia, 7–10 November 2011; pp. 4490–4494. [\[CrossRef\]](#)
56. Klick, J.; Lau, S.; Marzin, D.; Malchow, J.O.; Roth, V. Internet-Facing PLCs as a Network Backdoor. In Proceedings of the 2015 IEEE Conference on Communications and Network Security (CNS), Florence, Italy, 28–30 September 2015; pp. 524–532. [\[CrossRef\]](#)
57. Colla, M.; Leidi, T.; Semo, M.; Strasser, T. A Survey of Methods and Technologies for Developing Industrial Control Applications. *Msaum J. Rev. Surv.* **2009**, *1*, 259–271.
58. Diemer, J. Sichere Industrie-4.0-Plattformen auf Basis von Community-Clouds. In *Handbuch Industrie 4.0 Bd.1*; Vogel-Heuser, B., Bauernhansl, T., ten Hompel, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 177–204. [\[CrossRef\]](#)

-
59. Fallenbeck, N.; Eckert, C. IT-Sicherheit und Cloud Computing. In *Handbuch Industrie 4.0 Bd.4*; Vogel-Heuser, B., Bauernhansl, T., ten Hompel, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 137–171. [[CrossRef](#)]
 60. Pérez, F.; Irisarri, E.; Orive, D.; Marcos, M.; Estevez, E. A CPPS Architecture Approach for Industry 4.0. In Proceedings of the 2015 IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA), Luxembourg, 8–11 September 2015; pp. 1–4. [[CrossRef](#)]