Challenges Building a Data Value Chain to Enable Data-Driven Decisions: A Predictive Maintenance Case in 5G-Enabled Manufacturing

(article starts on next page)

28th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018), June 11-14, 2018, Columbus, OH, USA

Challenges Building a Data Value Chain to Enable Data-Driven Decisions: A Predictive Maintenance Case in 5G-Enabled Manufacturing

Magnus Åkerman[1*], Camilla Lundgren[1], Maja Bärring[1], Mats Folkesson[2], Viktor Berggren[2], Johan Stahre[1], Ulrika Engström[2], Martin Friis[3]

1: Industrial and Materials Science, Chalmers University of Technology, Gothenburg, Sweden
2: Ericsson Research, Ericsson, Gothenburg, Sweden
3: Manufacturing Development Centre, SKF, Gothenburg, Sweden

Abstract

Improvements in data storage and processing technologies have led many managers to change how they make decisions, relying less on intuition and more on data. This trend is especially notable for the manufacturing industry where Big Data applications, i.e. data analytics, are mentioned as an important enabler of value creation with the event of the fourth industrial revolution. Designing and building the entire data value chain that enables Big Data applications in manufacturing requires new knowledge about digital technologies combined with already established knowledge about the specific manufacturing processes. This paper focuses on the convergence of these different knowledge spaces applied to a specific case of implementing a Big Data application for predictive maintenance. Every step of building the data value chain from data acquisition to system feedback is presented and discussed in terms of the major challenges that were observed during the project. Results show that, just as the literature suggests, the knowledge gaps between different domains is a key component to manage for succeeding when building Big Data applications in the context of future manufacturing and maintenance.

1. Introduction

Recent improvements in data storage and processing technologies have led many managers to change how they make decisions, relying less on intuition and more on data [1]. This trend is especially notable for the manufacturing industry where Big Data applications, i.e. data analytics, are mentioned as an important enabler of value creation with the event of the fourth industrial revolution (Industry 4.0) [2]. The goal is to build Cyber-Physical Production Systems (CPPS) that can sense and adapt to its environment, which requires sensors and actuators to work together in complex emerging systems [3]. With CPPS it is possible to achieve self-adaptation and more dynamic automation. An important step towards these adaptable systems is to achieve predictive capabilities by improved decision making from data analytics [3, 4]. Manufacturing companies collect production data today but few utilize Big Data applications [5].

Maintenance is one area that is currently getting attention for data analytics applications. In digitalized manufacturing, maintenance must take a key role to reduce the risk and minimize the consequences of unplanned stops and disruptions [6, 7]. Data analytics can help to predict deviations in data, to trigger preventive actions to avoid failures or to reduce their consequences [8].

Even though technology provides the opportunities and enables better and more data to the decision-maker, developer and business leaders tend to focus on the potential of the technology that can collect and analyze big volumes of data rather than on the outcome of applying the technology [9]. To identify the right problems to address in an organization, the domain expert stays vital. The persons within an organization that have a deep expertise within an area are the ones that can identify where the biggest opportunities and challenges are [10].

Designing and building the entire data value chain that enables Big Data applications in manufacturing requires new knowledge about digital technologies combined with already established knowledge about the specific manufacturing processes. This study focuses on the convergence of these different knowledge spaces applied to a specific case of implementing a Big Data application for predictive maintenance. The data value chain is defined and every step is discussed regarding the diversity of different technologies and the specific domain knowledge needed.

1.1. 5G-Enabled Manufacturing

In the Swedish research project, 5G-Enabled Manufacturing, researchers and engineers work together to answer the question "What if we had unlimited free connectivity on the shop floor?". To answer this question, the shop floor at one factory in Gothenburg was supplied with an LTE network with 5G technologies. The idea is that a 5G network, which is a future cellular technology targeted for 2020, can be used to achieve low-cost mobile and stationary connectivity that the future factory needs.

Within the project, three specific demonstrators were created to show the capabilities the connectivity can enable. The first demonstrator is to have a cloud infrastructure and data analytics capabilities, meaning a data-center for data storage and data distribution as well as software with applicable libraries to achieve useful analytic results. The second demonstrator was a specific object to connect and get real-time data from, which is a grinding machine. Ball screw, slide, and motors are critical components in a grinding machine. These components were in focus during the project regarding predicting deviations. The third demonstrator is a mobile decision support system that the technicians, managers, and operators are using to some extent in the factory. The system is today connected using a Wi-Fi network and it is to some extent integrated with some parts of the factory to supply operators with mobile production data. This system is also used as a digital and mobile way to share information such as disturbances and instructions. Together the three demonstrators form the specific case of predictive maintenance on a grinding machine using LTE network with 5G technologies on the shop floor, which is further described in chapter 3.

2. Big Data in Manufacturing

This chapter aims to explain the connections between different enabling technologies and put Big Data in perspective to Industry 4.0.

2.1. Industry 4.0 and CPPS

By examining the literature about "how to do Industry 4.0", Hermann, et al. [11] identified four main design principles: interconnection, information transparency, decentralized decisions, and technical assistance. Interconnection, enabled by standards and modularization, increase collaboration between humans and machines. This enables new data and information and here is where information transparency becomes important. All the data needs to be refined to add value, enabled by Big Data, cloud computing, and smart devices [12]. Decentralized decision-making is a cornerstone for CPPS, where every entity acts as autonomous as possible. This includes the human part of the system that can be supported by technical assistance to allow for correct and autonomous decisions.

The Industry 4.0 maturity index [4] models a system in six separate steps that are needed to reach the capabilities of CPPS. The first two steps are computerization and connectivity, which requires digitized and interoperable systems. These steps allow data to be collected from sensors, machines, or other systems and sent to where it is needed. The third step, visibility, refers to the collection of all the raw production data. In the fourth step, transparency, the data is aggregated, correlated, and analyzed to understand why something has happened. The transparent information can be used to achieve predictive capacity, which is the fifth step. The idea is to project the collected data into the future and predict different scenarios. In the last step, adaptability, the system constantly adapts to the predictions.

A third model for CPPS is the 5C architecture model [13]. In this model, a CPPS consists of five levels: Smart Connection, Data-to-Information Conversion, Cyber, Cognition, and Configuration. The smart connection level means connectivity of the physical data sources. Data-to-Information conversion refers to the data processing or data analytics stage. The cyber level is the sum of all the information, but instead of focusing on all the raw, this level concerns with groups of analyzed data e.g. machine fleets or data over time. At the cognition level, the information needs to be presented in the correct way to experts so that it helps them in their decision-making. The last level, configuration, is the feedback from the virtual world back to the physical world.

These three different models of Industry 4.0 and CPPS are very similar (see Figure 1) even though they emphasize different things. They all put Big Data and analytics as central parts of the system but at the same time, it requires much more than just applying algorithms to raw data.
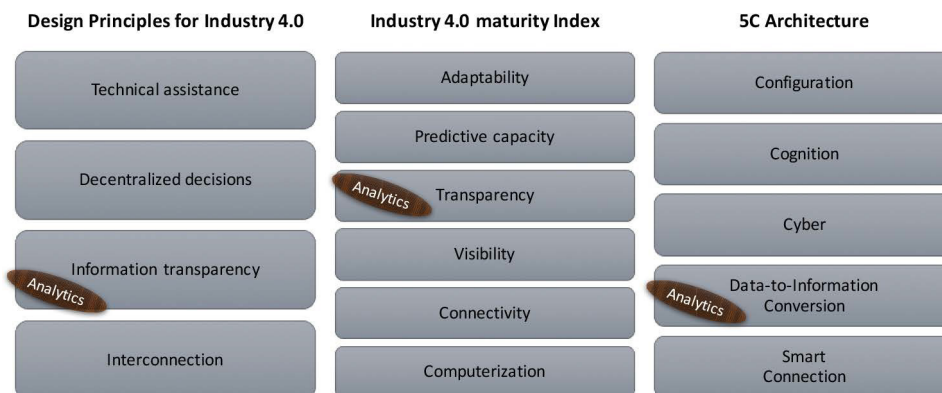
Figure 1: Three models of Industry 4.0 and CPPS and where they place data analytics [4, 11, 13].

2.2. Big Data applications

Big Data applications need to manage the four important V's of data: Volume, Variety, Velocity, and Veracity [14], which has been enabled by new technologies in distributed computing and data storage. According to Chen, et al. [15], any Big Data system requires the following six subsystems: data generation, data acquisition, data transportation, data pre-processing, data storage, and data analytics. When looking at how to specifically implement Big Data in the context of manufacturing and maintenance, Li, et al. [16] demonstrates an architecture that combines the CPPS view and the Big Data subsystems with required technologies (see Figure 2). It is a modular model that

includes feedback to the system in a CPS module, which is the physical part i.e. the manufacturing system. Data is acquired in the IoT module and pre-processed and analyzed in a Data Mining (DM) module. Like in the 5C architecture model, the results are distributed as services on the Internet of Services (IoS) module.

Figure 2: Simplified view of an architecture for predictive maintenance, adapted from [16].

Data analysis in manufacturing and maintenance has mainly been used to explain "what happened", but also to some extent for prediction. Vibration monitoring and analysis is common practice for condition-based maintenance (CBM). Predictive maintenance by vibration analysis is not commonly used but it was demonstrated already in 1985 [17]. Prognostic and Health Management (PHM) is a term used for strategies and techniques u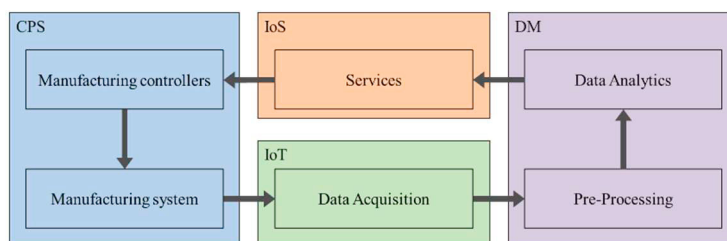sing big data sets, focusing on failure detection, current health assessment and prediction of remaining useful life [18, 19], root cause analysis of failures, and support maintenance planning [20]. In previous research, data sets from controlled environments have been used to validate PHM methods and algorithms [21]. Data from real factories is less frequently used to develop and validate PHM methods and algorithms due to complications with uncontrollable external factors, which makes the validation difficult [20]. In addition, Pellegrino, et al. [22] states the following research challenges: (1) system-level analytics focusing on the entire production system, (2) interoperability between systems and equipment, (3) lack of awareness, experience, and training to apply PHM principles and tools.

Big data and advanced algorithms are getting more explored in manufacturing companies to be used for fact-based decisions [5], as the algorithms can help to find more complex correlations in the data. The use of big data in maintenance has been shown by Li, et al. [23] for data-driven bottleneck detection. Algorithms have been used to identify bottlenecks from real-time production data to (among other) prioritize maintenance activities [24, 25].

3. Case Description

This chapter describes the case of building the data value chain for the predictive maintenance case, monitoring a grinding machine and utilizing results using a decision support system. The data value chain is defined in Figure 3 and every step is described with a focus on the challenges experienced and/or observed.
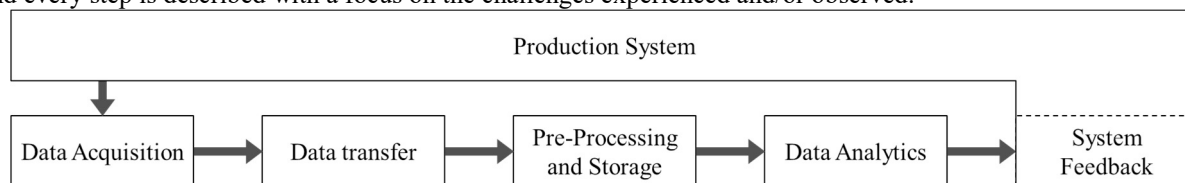
Figure 3: The data value chain.

3.1. Data Acquisition

There were two parallel approaches acquire data. One approach was to identify some key data sources that "should" be relevant when doing root cause analysis. The other approach was to add as much data as possible to test connectivity and perhaps get synergies during data processing. A small team of experts decided on a list of specific parts of the process that they wanted to be monitored. It was also decided that a couple of extra sensors should be mounted on the grinding machine. This resulted in four different ways to collect data divided into two main groups: internal machine

parameters and externally mounted sensors. Internal machine parameters are sensor signals and program system parameters that accessed from the internal machine computer. There are three externally mounted sensors or sensor systems. A vibration measuring system called IMX, two externally mounted sensors measuring cooling fluid and temperature connected with the communication protocol IO-Link, and a temperature sensor embedded on a Raspberry Pi.

A generic challenge regarding data acquisition is with what frequency the data should be collected. In this case, most data is collected every 100 milliseconds.

3.1.1. Grinding machine onboard computer

The industrial grinding machine has an onboard dedicated computer that controls the grinding process, user interfaces, and external connections. 39 parameters were decided on in an early stage, but the work to get all of them accessible was not an easy one. Mainly three types of challenges were encountered at this stage. First was the limitation of recourses. There was only one person available that could reprogram the machine to allow access to the wanted parameters. The second was the limitation of the machine control system. The system supported OPC UA that was the chosen way of communication but it turned out that the implementation did not allow access to all wanted parameters through OPC UA. The third challenge comes from the nature of working with a manufacturing system, and that is the limited time that it is possible to implement changes in a producing machine on the shop floor. The managers running the manufacturing process will not allow changes to a working machine without good reasons to do so. In the end, 43 different tags are acquired from the machine computer including values for the position of axis, torque, drive load, time of processes etc.

3.1.2. Externally mounted sensors

The advantage to collect data from externally mounted sensor systems is that it avoids any issues regarding the internal control system, such as limitations or access to limited recourses. A disadvantage is of course that there are more systems to connect and keep track of. Choosing between the two options is sometimes only a matter of preference, like with the IO-Link sensors or embedded temperature sensor. These could have been connected to the machine internal computer and accessed that way. However, for the IMX system that option would not achieve the same results since this system does some processing or aggregation of the vibration data before sending it further. Just as with the internal computer, mounting external sensor systems is also limited by the manufacturing process that always gets priority.

3.2. Data Transfer

Each digital system supports different, and sometimes several, types of communication protocols. The grinding machine computer and the IMX system both support the standard OPC UA, and this was chosen as a preferred communication because of its being future proof with a Service Oriented Architecture (SOA) [26]. The IO-Link sensors send its data through the IO-Link master as a TCP data stream according to each sensors specification, published in an IO-Link database [12]. The temperature sensor can be read locally in the Raspberry Pi computer and then sent by any communication system with support for the Linux platform. On top of the described systems, the IoT platform Calvin [27] was used to send the data to the data-center. This solution makes it possible to unify the data close to the network edge and transfer the data in the same way over the Calvin platform. The challenges at this stage were related to the commonly known problem of the difference between industrial networks and IP networks. Sending data to the data-center is part of the software engineering domain while industrial networks are part of the manufacturing domain. Therefore, the local experts with knowledge of the process and industrial systems have little knowledge about available communication options. While the experts of the communication systems don't understand the specifics about the industrial related data such as frequency, size, data types etc. Another challenge is to choose the different communication protocols, which often is limited to what is available depending on the system. Furthermore, using an IoT platform is naturally optional, it this case it was useful, but that is also an architectural decision to make.

3.3. Pre-Processing and Storage

At the data-center, the data is stored in a document database [28] which is a NOSQL [17] database that can handle different types of data. This is important for Big Data applications, variety being one of the four V's of Big Data. To handle two other V's, volume and velocity, the Apache Hadoop [29] and Kafka [30] platforms are used for distributing the data. This cloud infrastructure is common and what is needed for Big Data applications today. The data is stored as documents in the database as JavaScript Object Notation (JSON). To be able to use the data together some parts of these documents should be similar, even if the datatypes are different and the data comes from different sources and means of transportation. The challenge here is to understand what specific part of the data that is important for future processing. Also, to identify important metadata that is needed to correctly connect the data together.

3.4. Data Analytics

The major challenge for the data analytics step was very much aligned with the literature in that it is difficult for the data analysis experts to choose correct methods and data sources without having the deep knowledge of the domain experts [1]. In this case, since the approach was to collect a mix of known relevant data and some extra data with potential synergies, the first issue was to choose what scenario to start with. The choice was to focus on anomaly detection in the vibration sensor data. A specific challenge about anomaly detection based on vibration data is that it lacks a "correct" value. Vibration data is easy to understand as a concept but can be affected by many external events that are not monitored, any anomaly needs to be manually correlated with the knowledge of experts on-site.

There are several open source libraries available with machine learning algorithms. In this case, the anomaly detector from Twitter's Luminol library, suited for continuous and seasonal data, was used [31]. Anomalies were identified for the eight different vibration sensors, and during a working day, the number of anomalies for the vibration sensors varied between ca 5–50 (Figure 4).

The following challenge was to find the reason for the anomalies. Root Cause Analysis (RCA) was done based on other data sources/parameters besides the vibrations, for example, torque and temperature. The data is very frequently collected (100 ms), and to get a generic visual understanding of the nature of the data, the data was first aggregated in five-minute periods. The min-, mean-, max-, sum- and standard deviation of the periodical data were visualized in heat maps. It is a tool that can graphically aid the process of working with the data.

A Decision Tree algorithm (Figure 4) is used for the RCA to find the combinations of values from other parameters. It also provides the possibility to graphically display the value-combinations as a 'tree', which is important to increase the understanding of the results of the machine learning algorithms and how data relate to each other.

Figure 4: Left: Detected anomalies marked in red for one vibration sensor, Kanal1ENVTotal (data blue, smoothed data green). Right: Decision Tree for a sensor with five detected anomalies (deviations). Blue boxes show on value combination for deviations.

A major problem with applying algorithms for continuous data in a discrete manufacturing process is that the process is not continuous. Unfortunately, the parameter that state the current specific process was one of the data sources that were difficult to access early. Eventually, this was accessed and a new approach was staged. The different working phases, e.g. rough grinding, fine grinding, waiting, idle, etc., was now separated from each other. The phase lengths vary between about one second to around ten seconds and the data used in the heatmaps and Decision Tree

was re-calculated into the discrete phase-time periods that are continuous in time but of different length. The phase was now included in the heat maps as well as the Decision Trees.

3.5. System Feedback

As mentioned above, the system feedback was planned to be through the mobile decision support system already used to some extent on the shop floor. Technically, with the current setup and infrastructure, it was very easy to send data to this system. It is a modern system that supports the publish-subscribe protocol MQTT [32], which has become popular for many IoT applications.

Since the actual data analytic stage is an ongoing process it was difficult to visualize the results in any meaningful way. It is too difficult at this stage to understand what the anomalies represent and how they should be interpreted. However, some real-time data can be visualized more directly and still achieve some added value for operators and technicians. E.g. by visualizing the thickness of the grinding wheel it is possible to manually predict when it should be changed (allowing more proactive work instead of being only reactive). It was also decided that the vibration data should be presented in real-time so that a machine operator could see if something unexpected happens. This does not provide any predictive capability but it can still be useful information. The issue with this vibration data is to know how to visualize it for humans. The sensor system measures vibrations on several axes and then creates a spectrum analysis before sending it further, this data is not obviously interpreted.

4. **Discussion and Conclusions**

This paper describes the challenges encountered when building the data value chain for predictive maintenance of a grinding machine in 5G-Enabled Manufacturing. The data value chain has been implemented in an industrial context and development of algorithms for predictive maintenance has been started. As stated by Jin, et al. [20], uncontrollable external factors are making the validation difficult, and lack of experience to apply and use PHM [22]. In this case, we have started to develop the experience by combining the domains of manufacturing, information technology, and data analytics.

Just as emphasized in the literature, the knowledge gap between different domain experts cannot be overestimated [10]. These domain differences are not limited to the analysis phase but do to some extent exist along the entire data chain. With the described challenges in mind, the authors suggest the following changes for future Big Data implementations:

- Agile work cycle. Meaning that there should be very short iterations between new data acquisitions, analysis, and utilization. This includes following the entire value chain for every new data source. Adding everything at once can create too many questions at every step of the chain, halting any progress.
- Know what parts of the process that are self-comparable, and what parts are not relevant. In a discrete manufacturing flow, there are lots of different phases and some are just idle and simply cannot influence the process.
- Connect the data to relevant metadata depending on products, components, machines, batches, or product families etc.
- Experiment with the data but let the manufacturing process experts guide these experiments, do not suffice with letting them comment on results.

References

[1] E. Brynjolfsson and K. McElheran, "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review,* vol. 106, pp. 133-139, 2016.

[2] H. Kagermann, W. Wahlster, and H. Johannes, "Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0," acatech2013.

[3] L. Monostori, B. Kádár, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reinhart, *et al.*, "Cyber-physical systems in manufacturing," *CIRP Annals - Manufacturing Technology,* vol. 65, pp. 621-641, 2016.

[4] G. Schuh, R. Anderl, J. Gausemeier, M. t. Hompe, and W. Wahlster, "Industrie 4.0 Maturity Index: Managing the Digital Transformation of Companies," acatech STUDY, Munich2017.

[5] P. O'Donovan, K. Leahy, K. Bruton, and D. T. J. O. Sullivan, "Big data in manufacturing : a systematic mapping study," *Journal of Big Data,* 2015.

[6] J. Bokrantz, A. Skoogh, C. Berlin, and J. Stahre, "Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030," *International Journal of Production Economics,* vol. 191, pp. 154-169, Sep 2017.

[7] J. Bokrantz, "On the Transformation of Maintenance Organisations in Digitalised Manufacturing," Licentiate thesis, Chalmers University of Technology, Gothenburg, 2017.

[8] C. Lundgren, A. Skoogh, B. Johansson, J. Stahre, and M. Friis, "The Value of 5G Connectivity for Maintenance in Manufacturing Industry," *Proceedings of the 2017 Winter Simulation Conference,* 2017.

[9] M. Patel, J. Shangkuan, and C. Thomas. (2017). *What's new with the Internet of Things?* Available: https://www.mckinsey.com/industries/semiconductors/our-insights/whats-new-with-the-internet-of-things

[10] A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review,* 2012.

[11] M. Hermann, T. Pentek, and B. Otto, "Design Principles for Industrie 4.0 Scenarios," in *49th Hawaii International Conference on System Sciences (HICSS)*, Hawaii, 2016.

[12] H. Kagermann, "Change Through Digitization—Value Creation in the Age of Industry 4.0," in *Management of Permanent Change*, H. Albach, H. Meffert, A. Pinkwart, and R. Reichwald, Eds., ed Wiesbaden: Springer Gabler, 2015.

[13] 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.

[14] J. Gantz and D. Reinsel, "Extracting Value from Chaos," 2011.

[15] M. Chen, S. Mao, and Y. Liu, "Big Data: A survey," *Mobile Netw Appl,* vol. 19, pp. 171-209, 2014.

[16] Z. Li, Y. Wang, and K.-S. Wang, "Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario," *Advances in Manufacturing,* vol. 5, pp. 377-387, 2017.

[17] J. T. Renwick and P. E. Babson, "Vibration Analysis-A Proven Technique as a Predictive Maintenance Tool," *IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS,* vol. IA-21, 1985.

[18] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry,* vol. 57, pp. 476-489, 2006.

[19] J. Lee, M. Ghaffari, and S. Elmeligy, "Self-maintenance and engineering immune systems: Towards smarter machines and manufacturing systems," *Annual Reviews in Control,* vol. 35, pp. 111-122, 2011.

[20] X. Jin, D. Siegel, B. A. Weiss, E. Gamel, W. Wang, J. Lee, *et al.*, "The present status and future growth of maintenance in US manufacturing: results from a pilot survey," *Manufacturing Review,* vol. 3, p. 10, 2016.

[21] G. W. Vogl, B. A. Weiss, and M. Helu, "A review of diagnostic and prognostic capabilities and best practices for manufacturing," *Journal of Intelligent Manufacturing,* pp. 1-17, 2016.

[22] J. Pellegrino, M. Justiniano, A. Raghunathan, and B. A. Weiss, "Measurement Science Roadmap for Prognostics and Health Management for Smart Manufacturing Systems," National Institute of Standards and Technology2016.

[23] L. Li, Q. Chang, and J. Ni, "Data driven bottleneck detection of manufacturing systems," *International Journal of Production Research,* vol. 47, pp. 5019-5036, 2009.

[24] M. Subramaniyan, A. Skoogh, M. Gopalakrishnan, and A. Hanna, "Real-time data-driven average active period method for bottleneck detection," *International Journal of Design & Nature and Ecodynamics,* vol. 11, pp. 428-437, 2016.

[25] M. Subramaniyan, A. Skoogh, M. Gopalakrishnan, H. Salomonsson, A. Hanna, D. Lämkull, *et al.*, "An algorithm for data-driven shifting bottleneck detection," *Cogent Engineering,* vol. 3, 2016.

[26] (2017-11-21). *Unified Architecture*. Available: https://opcfoundation.org/about/opc-technologies/opc-ua/

[27] (2015, 2017-12-15). *Calvin - Lets Things Talk to Things*. Available: https://www.github.com/EricssonResearch/calvin-base

[28] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-Physical Systems: The Next Computing Revolution," *Proceedings of the 47th Design Automation Conference,* pp. 731 - 736, 2010.

[29] (2017-11-28). *Open CV (Open Source Computer Vision Library)*. Available: https://opencv.org/

[30] B. Esmaeilian, S. Behdad, and B. Wang, "The evolution and future of manufacturing: A review," *Journal of Manufacturing Systems,* vol. 39, pp. 79-100, 2016.

[31] (2018-02-11). *Luminol*. Available: https://github.com/linkedin/luminol

[32] G. Weichhart and D. Wachholder, "On the interoperability contributions of S-BPM," in *6th International Conference on Subject-Oriented Business Process Management*, 2014, pp. 3-19.