Data analytics framework for Industry 4.0: enabling collaboration for added benefits

Sanja Lazarova-Molnar[1] ✉, Nader Mohamed[2], Jameela Al-Jaroodi[3]

[1]*Faculty of Engineering, University of Southern Denmark, Odense 5230, Denmark*
[2]*Department of Mathematics, Computer Science and Information Systems, California University of Pennsylvania, California, PA, USA*
[3]*Department of Engineering, Robert Morris University, Moon, PA, USA*
✉ *E-mail: slmo@mmmi.sdu.dk*

Abstract: Industry 4.0 is a promising vision for advancing the manufacturing sector through the recent innovations in information and Communication Technologies that enable collecting, storing, and processing detailed and accurate data about industry processes. This data enables manufacturers for data-driven decision making to significantly improve their operations and profitability. Most of the large manufacturing enterprises can benefit from this as they can collect more data that can be utilised to enhance their decision-making processes. Small and medium enterprises (SMEs) have limited data and resources, thus reducing the possible gains. However, if SMEs and small manufacturing facilities collaborate and share data, which is then jointly analysed, feasibility and quality of their data analytics and decision-making processes could be significantly enhanced. This study discusses collaborative data analytics (CDAs) in Industry 4.0, summarising findings into a novel CDA framework that can be used by manufacturing enterprises of any size and scale to enable and enhance the mutual benefits of CDAs and decision-making processes. The CDA framework can enhance the key factors and performance metrics of manufacturing facilities such as reliability, availability, and efficiency. The study also provides a preliminary benefit analysis of utilising the proposed CDA framework for manufacturing SMEs.

1 Introduction

The fourth Industrial Revolution or Industry 4.0 is the widely adopted term to refer to the phenomenon of instilling the latest information and communication technologies (ICTs) and corresponding infrastructures into automation and manufacturing processes [1]. The results are more data-driven and more intelligent manufacturing and manufacturing-related processes. Through the emergence of Industry 4.0, future manufacturing systems are expected to become highly flexible, modular and much more efficient, producing intelligent products that control their own production processes [2]. This anticipation will be facilitated by the availability of data and advancements in data analytics processes, which are considered to be vital to the development of the Industry 4.0.

Small and medium enterprises (SMEs) are anticipated to face serious challenges in adopting this fourth Industrial Revolution [3–7]. Frequently, they have been termed as victims in this industrial revolution. Data analytics and simulations, central to the Industry 4.0, are often luxuries, almost exclusively utilised by large corporations that can afford for them [8]. On top of this, high-quality data analytics relies heavily on the quantity and quality of available data, neither of these being easily available to SMEs. There is also the increased pressure on adopting *predictive manufacturing*, which is termed as the next transformation [9], adding to the list of challenges. Predictive manufacturing implies that manufacturing systems possess technologies and intelligence to proactively implement mitigating solutions to prevent efficiency loss in manufacturing operations. This includes predicting equipment performance, as well as faults and anomalies, and inferring future fault events and even diagnosing potential root causes of problems. For this reason, SMEs are in dire need to establish feasible and cost-effective approaches to data analytics to keep themselves in good shape to complete on the market [10].

Collaborative data analytics (CDAs) can facilitate some of the data analytics processes that are by far unfeasible to SMEs due to them being costly to be configured and run. CDA works by manufacturers jointly processing and analysing data to the advantage of all participants. In some of our previous works, we studied the potential of CDA in the context of building management systems (BMSs) [11, 12], where we identified very significant benefits, in terms of obtaining better quality decision support. We believe that our findings can be extended to manufacturing systems, at least in some of the data analytics problem areas, as we identify further on in this paper. The advantages of using CDA can be applied beyond SMEs, i.e. for large corporations as well.

Future manufacturing machines are expected to be very flexible and configurable, implying that the same manufacturing modules might be in use by different manufacturers. This fact creates possibilities for a vast potential in joining the data to develop accurate models of the manufacturing machines within very short times. In that sense, this data could be utilised for improved machines' configurations toward the better performance of the manufacturing systems. Provided the typical rivalry among manufacturers, CDA will anticipate challenges regarding privacy and data protection, which will need to be taken into consideration.

As a result of our analysis and study of the CDA in the context of Industry 4.0, we have developed the CDA framework that we present in this paper. The preliminary work that forms the basis for this CDA framework is presented in [13].

The paper is structured as follows. We provide a summary of the current developments in Industry 4.0 in Section 2, with a special focus of collaborative approaches. In Section 3, we present the significance of CDA in the context of Industry 4.0, focusing on potential data analytics tasks. In Section 4, we introduce the framework for enabling CDA, along with a case study for one of the data analytics aspects and discussion of its challenges and benefits. In Section 5, we analyse the implications of the use of collaborative processes, which is then followed by a general discussion in Section 6. Finally, we conclude the paper in Section 7.
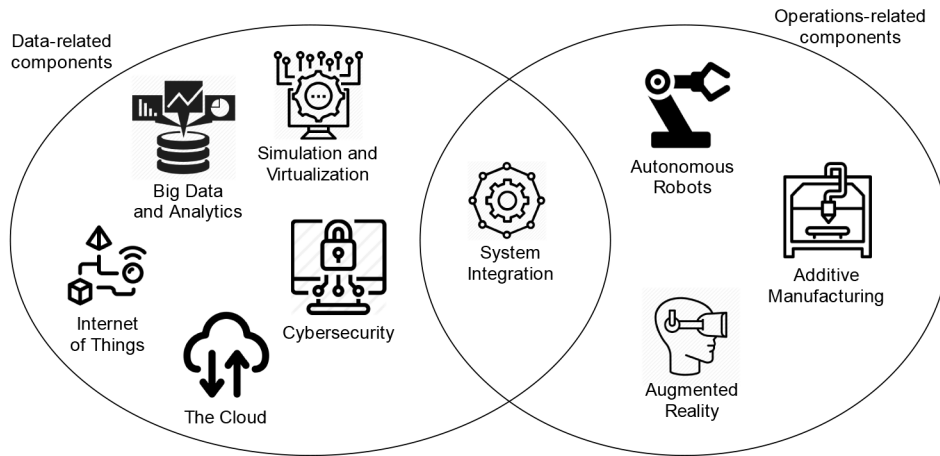
Fig. 1 *Enabling components of industry 4.0*

2 Background

Even though Industry 4.0 is currently at the level of a concept and it is still very vaguely defined, it is very much happening and changing the world. Nine components have been identified as enablers for the Industry 4.0 [14], such as Big Data and Analytics, system integration, simulation and virtualisation, The Cloud, internet of things, cybersecurity, augmented reality, autonomous robots and additive manufacturing. We have analysed all of them and divided into data- and operations-related components, as illustrated in Fig. 1, to resort at the fact that most of these components (six out of nine) are centred around the data and data analytics to enable automation and optimisation of manufacturing processes. The main function of the aforementioned six components is to enable data collection, storage and processing.

The transformations that Industry 4.0 imposes on manufacturing facilities carry a number of substantial challenges, of both technical and societal characters [15]. Some of the identified challenges are listed as follows:

- The necessity to improve intelligence level of manufacturing equipment, to cope with the expectations of flexible manufacturing.
- Needs for higher standards of data acquisition, to support fine-grained and efficient heterogeneous data acquisition.
- Need for deep integration networks, to facilitate the high transmission rate of data, low duty cycle and IP network availability.
- Need for platforms that relate data to decision-making processes, to automate and enable processes like sales forecasting and quality analysis.
- Need for re-adjustment of education programmes and intakes, as there will be a change in the needed competencies [16].

The hardest hit in this latest revolution will be SMEs, who are not prepared for the Industry 4.0 as many SMEs are still not utilising latest ICTs due to it being too costly for them [8]. To cope with the new circumstances and challenges, SMEs will need to collaborate in the data analytics processes.

In the following, we identify the different roles that data analytics for Industry 4.0 can play, as well as the different challenges that it can address. We, furthermore, provide definition and scope of the CDA in the context of cyber-physical systems.

2.1 Data analytics – The core of Industry 4.0

The process of making use of data to obtain knowledge by applying mathematical, statistical and computational techniques is known as data analytics. Insights are the typical representation form of the extracted knowledge, which are then utilised to facilitate different decision-making processes. Data analytics plays a central role in facilitating Industry 4.0 [17, 18]. Based on findings throughout the literature, we identify four main data analytics

needs of manufacturing facilities that target manufacturers' performance [13, 19–21], listed as follows:

(i) Customer analytics,
(ii) manufacturing and operations planning (MOP) analytics,
(iii) prognostics and health management (PHM) analytics, and
(iv) logistics optimisation (LO) analytics.

In the following, we elaborate on each of the four data analytics aspects. The aspect of the data analytics needs for Industry 4.0 that focuses on customers' behaviour and relationships, we term as *Customer Analytics*. This aspect focuses on data analytics focused on enhancing customers' experiences and their relationships with manufacturers. Customer analytics processes enable manufacturing companies to design their products and supply chains to better match customers' expectations and needs. Furthermore, customer analytics can also support processes for mass customisation of products [22], which aims to limit production products that customers want and are most probably going to purchase, thus generating less waste and supporting sustainable and environmentally friendly production.

Improving the planning of manufacturing and operations processes forms the second aspect of data analytics in Industry 4.0, which we term as *MOP analytics*. The objective of this aspect is to utilise collected data to configure manufacturing operations in order to facilitate that each individual machine/module performs as efficiently as possible and participates in a layout that both increases throughput and decreases bottlenecks in manufacturing.

Obtainability of data has uncovered the usually undistinguishable inefficiencies and degradations of manufacturing systems and their components. Data on these degradations and inefficiencies have further enabled preventive management procedures aimed at optimising maintenance processes. This forms the third aspect of data analytics, focused on the health monitoring of manufacturing machines with an aim to enhance their and their systems' reliability and availability. This aspect is referred to as PHM analytics [23].

LO analytics is the fourth aspect of data analytics processes for Industry 4.0. Logistics and its optimisation forms the focus of LO analytics [24]. LO data analytics methods focus on improving supply chain management to better coordinate supply and demand, and thus, reduce costs for sourcing, shipping, storing, stockout, and disposal. Typically, LO analytics has applications in strategic sourcing (supply chain network design, and product design and development) and demand planning (procurement, production, inventory and logistics). LO analytics practices are inherently collaborative as supply chains are systems of large numbers of entities that collaborate to enable moving products or services from suppliers to customers.

Table 1 Benefits and challenges associated with the four aspects of CDAs for industry 4.0

	Benefits	Challenges
customer analytics	- more accurate customer profiles	- customers' consent needed
	- more customer-tailored products	- privacy protection mechanisms needed
	- reduced waste	
MOP analytics	- faster obtaining of optimal layouts	- needs ways to protect competing manufacturers
	- more effective and timelier avoidance of bottlenecks	
PHM analytics	- more accurate fault models of modules (beneficial to all)	- machine/robot manufacturers will be more challenged to provide higher quality machines and more realistic pricing plans
	- higher availability and reliability of machines	
	- higher quality machines	
LO analytics	- improved inventory management	- suppliers and transportation organisers might be challenged to deliver better services
	- timelier elimination of bottlenecks in logistics processes	
	- better delivery times	

2.2 Collaborative data analytics for systems

CDA, as usually referred to throughout literature, implies the processing of data through collaboration of humans. One popular and successful example of this is Wikipedia [25], where the content is developed through a large-scale contribution of people. Thus, CDA usually has a 'human' dimension. We extended this definition to include other contributors. We describe CDA as joint analytics of data from multiple and varied sources, by multiple and varied processors (human or hardware), for accomplishing synergistic outcome, i.e. an outcome that is greater than the sum of the individual outcomes of the contributors' [12].

We already investigated the potential of CDA for cyber-physical systems, for the case of BMSs, an enabling constituent of smart buildings [11, 12]. The studied benefits of CDA in smart buildings motivated us to look into applying the concept to the context of Industry 4.0 [13]. The high similarity between BMS and manufacturing systems, in particular, as both are cyber-physical systems that generate a lot of data with predefined performance metrics and with high level of human interaction, encouraged us to do so. Additionally, BMS and manufacturing systems share some of the performance metrics and goals, such as energy efficiency. This led us to believe that the benefits of performing CDA can also be comparable and motivated us to study and analyse the potential of CDA for manufacturing systems, as well as identify the ways in which it could be made feasible.

With respect to the collaborative data processing and analytics, not many substantial efforts for manufacturing systems have been documented. Schuh *et al.* studied the potential of collaboration with respect to productivity and competitiveness, where they also see Industry 4.0 as enabler for collaboration [26]. They point four global enablers for collaboration productivity in the context of Industry 4.0: IT-globalisation, single source of truth, automation and cooperation. The authors of this paper clearly identify the advantages of collaboration in terms of a reference system that consists of two indicators: return on engineering and return on production. Regarding the practical implementation issues, Yue *et al.* provide an insight into the role of the cloud for the cyber physical production systems [27] and Morgan and O'Donell present an implementation of cloud manufacturing monitoring systems [28]. In both contributions, sharing among factories is noted as one of the goals.

In this paper, we provide an exhaustive study of the different data analytics aspects that can be targeted by CDA, and we identify the ways in which these data analytics aspects can be targeted along with the potential challenges and benefits.

3 Collaborative data analytics for Industry 4.0

CDA at the level of systems and machines, if properly facilitated, can offer a number of benefits for both manufacturers and customers. In the following sections we identify the benefits and challenges of CDA for each of the four aspects presented in Section 2.1. We, furthermore, describe the models in which CDA can be performed.

3.1 Collaborative data analytics: benefits and challenges

Names and affiliations should immediately follow the title. To avoid confusion, the family name must be written as the last part of each author name and the first name should be spelt out rather than abbreviated (e.g. John A.K. Smith). Author details must not show any professional title (e.g. Managing Director), any academic title (e.g. Dr.) or any membership of any professional organisation

We summarised the benefits and challenges associated with each of the four data analytics aspects in Table 1. We explain each of them as follows:

Customer Analytics is the first aspect that we look at. The relationship customer-factory is typically many-to-many, meaning that, the same customers are often shared by many production facilities. Thus, collaboration manufacturers could build more comprehensive and accurate models of customers, which would support their relationships to customers and yield more personalised products and services. This collaboration among manufacturers is also very challenging, considering existing and forthcoming legislatures for data protection. These challenges imply a need for CDA to be rationalised to and accepted by customers, and additionally encompass data anonymisiation processes, safeguarding identities of customers.

The advantages of PHM analytics stem from the large amounts of data on faults of manufacturing machines. Both faults and failures are usually rare events. Therefore, data on faults is always very sparse. On top of this, proper event logging of faults is a typically intrusive process, implying that information about faults needs to be manually entered in most of the cases. For this reason, combining data from different manufacturing facilities would make a lot of sense for fault-related data. As a result, highly accurate fault and degradation models for manufacturing machines can be built within short times. This, in turn will lead to more accurate and timelier fault detection and diagnostics, as well as more effective preventive maintenance schedules, resulting into higher reliability and availability of manufacturing machines [29, 30]. The privacy in this context is not of great concern, as the data is not very sensitive. The only stakeholders who might not benefit from these CDA processes are machine manufacturers. Machine manufacturers will be actually pressed to produce better quality products and apply more realistic pricing.

The benefits of collaborative MOP analytics have already been recognised [31]. Some of the benefits that have been identified are shorter production lead time due to effective collaboration process planner and numerical control programmer, quick time-to-volume with effective collaborative manufacturing capability, and lower product cost with efficient collaborative process planning and numerical control programming. Further to this, sharing of data could also imply timelier fine-tuning of manufacturing operations and layouts, as well as removal of bottlenecks.

LO analytics is concerned with the use of data analytics for supply chain performance [32–34]. It is apparent that the supply chain typically links different manufacturers among each other, as well as with other businesses. In this task, collaboration is inevitable. However, by providing automated models that facilitate data sharing and CDA while respecting legislatures, collaboration processes could be greatly enhanced. Even existing manufacturers could share data and findings with new manufacturers and possibly even developing a business model that would attach a price to such services. Furthermore, by CDA, independent manufacturers can

Table 2 CDA models for Industry 4.0

Model	Options	Benefits	Challenges/ limitations
data collection model	1. static data collection	• allows for more controls by the manufacturer on their data • secure • stable	• slow observations • difficult to change/upgrade applied data collection processes
	2. dynamic data collection	• situation-based data collection • continuous monitoring • flexible adjustments • faster observations	• involves extra technical solutions • involves more security requirements
offered model	1. private	• provides more customisable data analytics • provides better security and privacy measures	• high cost of ownership • limited experience and capabilities
	2. community	• low to medium cost of ownership • more data is collected, thus faster and more accurate observations can be found • more specialised data analytics services can be offered to the community	• data sharing limitations • medium need for flexibility and scalability requirements • strong need for security and privacy requirements
	3. public	• low cost of ownership • more data is collected and faster new knowledge observations • better access to high-performance resources	• data sharing limitations • high need for flexibility and scalability requirements • high need for security and privacy requirements
processing model	1. in-house solution	• provides better security and privacy • higher levels of control	• requires in house development, configuration and support • high cost of ownership
	2. cloud-based solution	• limited technical support needed • more cost-effective • new analytics functions can be easily added • more data can be collected, stored and analysed for better findings	• high security and reliability requirements • big data analysis requirements

also better predict demand and even better understand the prediction. The challenge might be, again, that companies are competitors, but if properly handled and possibly priced, such services could actually benefit all manufacturers.

Finally, to protect data and support privacy of competitors, there could be another model to enable CDA, i.e. through a third party. That means that there would be companies that enable data protection and privacy and facilitate data analytics, thereby ensuring that each participant obtains the outcomes of the CDA processes.

In the following sections, we present the models in which CDA can be performed for manufacturing facilities.

3.2 Collaborative data analytics models for Industry 4.0

There are many models that can be used for CDAs for Industry 4.0 applications. These models are based on different aspects including how data collection for CDA is achieved, how CDA is offered, and how CDA is processed. There are two options for data collection for CDA: static data collection and dynamic data collection. With static data collection, each manufacturer will agree on the types of gathered data and the rate at which this data will be gathered. On the other hand, with dynamic data collection, it is possible to change the collected data from any participating manufacturer and/or change the collection rates at any time. This allows for managing the data collection process to achieve better results in discovering new knowledge. In addition, CDA for Industry 4.0 can be offered for private, community, or public use. The private use can be for multiple manufacturing units owned by a single company. The community use can be for several manufacturing companies or industries that have similar production business and share closely related processes. The public use can be for all manufacturers in a region, country or worldwide while they have different types of production businesses. Moreover, there are two options for how to conduct or process CDA: in-house or using cloud-based services provided by a third party. Table 2 provides summaries of CDA models, their options, and their benefits and challenges/limitations.

4 Framework for collaborative data analytics for Industry 4.0

Based on our findings related to the four Industry 4.0 data analytics areas where CDA can be performed, as well as the different types and models for their implementations, presented in the previous sections, we developed a CDA framework. In the following sections we describe our framework, and further illustrate it for one of the four aspects.

4.1 Framework description

To enable the CDAs processes, considering our findings described in the previous sections, we have developed a framework that incorporates all processes, from data collection, to decision support, into a series of concrete steps. The framework that describes the manufacturers' CDA process workflow is illustrated in Fig. 2. In the following we describe all framework steps in more detail in the same sequence in which they are illustrated in the figure:

- *Goals*: The first step in the Industry 4.0 CDA framework is the determination of data analytics goals and security expectations. The data analytics goals determine the data analytics problem domain that is targeted (one of the four aspects shown in Fig. 2 and described in Section 2), resulting in a final outcome in form of a concrete scheme for participation in the CDA. Security expectations define which parts of data are sensitive to sharing and therefore, need special protection mechanisms. These CDA goals and security expectations, once formulated lead to the next process, i.e. requirements.
- *Requirements*: Based on the defined CDA goals, the type of decision support is defined (e.g. if the goal is to increase the number of customers, the decision support would be in terms of the number of employees, the mode of reaching out to customers, opening hours, etc.). Furthermore, when CDA goals are combined with the security expectations, this yields more precisely the type and model of participation (as described in

Section 3, e.g. if it has high-security expectations, an in-house private solution might be preferred). Finally, CDA goals also lead to defining the required data streams. For instance, if a manufacturer's goal is to enhance reliability of a manufacturing facility, examples of corresponding data streams that will be needed are event logs on faults from manufacturing machines, repair times, qualifications of repairmen, etc. The three activities presented in this *Requirements* step will need many back-and-forth alignments as they all depend on each other. Based on the manufacturer's security specifications and other relevant circumstances, it will be defined whether the manufacturer's participation will be in-house or cloud-based, and whether it will be as private, community or public. These decisions will form both assumptions and limitations to the decision support that the manufacturer can expect and benefit from. Explicitly, if the manufacturing facility does not avail some of these data streams, it might not be able to obtain certain kind of decision support, or it might need to accept a lower level of accuracy and quality. Furthermore, some types of decision support will assume availability of certain data streams, and would not be offered, unless the required data streams are in place. This explains the next process, explained as follows.

- *Refinement*: Due to the discovery of limitations in the *Requirements* step, the goals will need to undergo refinement based on these new limitations. The output of this step is concrete agreements between the collaborators with respect to the decision support that can be expected and provided.
- *Security*: Once the goals are clarified and redefined to reflect the real situation at hand, the next step is to define the security requirements and guarantees. The output of this step is agreements about privacy and security.
- *Participation*: Once all previous processes are completed and agreements have been made, the manufacturer is approved and enabled to participate in the CDA scheme and it can start to share data, either online or offline, protected according to the security agreements. In return, the manufacturer will receive the agreed decision support, which will decide how to act upon and use it independently.

It is apparent that the more participants there are in the CDA scheme, the more effective it can be, so there will be a high focus on gathering participants. This is a very challenging problem, which will need to involve educational steps that clearly describe the benefits of joining such CDA schemes. We discuss these challenges in Section 4.3, after illustrating the concept of CDA in the following Section 4.2.

4.2 Case study

In the following we provide a case study for the customers data analytics aspect, illustrating how our proposed CDA framework would facilitate the relevant processes.

Customers are the central point of each enterprise, and therefore, learning more about them, in correlation to their other purchasing habits, will be of immense meaning to developing accurate and comprehensive customer profiles. Customer profiles that capture (almost) complete purchase habits can be utilised to better stipulate products' performances, as well as to better serve customers, and, thus, enhance customers' satisfaction levels and repurchase likelihoods. We use Fig. 3 to illustrate the case study for the customers' CDA.

For CDA to be enabled and successful in the customer analytics case, data on customers need to be shared. Customers' data in the CDA processes will need to be matched based on customers' identities. Customers' identities might not always be obvious and easy to derive, as there are often different customers with identical names, as well as customers that use multiple credit cards, etc. Once this data is matched and collaboratively analysed, customers and their purchasing habits can be more accurately modelled, resulting in rules and recommendations that can be extracted and sent back to manufacturers. Example of a recommendation would be the manners in which certain customers can be most effectively approached for certain products by providing hints on certain product or service characteristics that certain customers value. This characteristic, as expected, might be independent of the manufacturer and/or product, so having the information about customers' preference of being approached about products would be valuable to any participant in the scheme.

The problems and challenges in this data analytics aspect, applying also to the example of the manner of approaching customers, can be linked to the fact that certain manufacturers compete for the same customers. Therefore, the main problem would be phrased as 'why should I help my competitors in gaining my customers?', but we anticipate that later this challenge can be transformed and yield a different type of competition, i.e. a competition in delivering the most optimal action for the provided recommendations. For instance, manufacturers can receive a recommendation in the form of 'this customer highly values a certain characteristic in this kind of products', so it means that the competing manufacturers who pay more attention to the specified valued characteristic or find the best way to address it, would be in a better position to gain the customer. The clear winner in this scenario would be the customers, as their needs would be addressed more directly.

Often, the suggested recommendation from the CDA scheme might be on a subtle purchase habit that even the customer is not expressive about and it has been discovered only through their previous purchases. An example for this would be that a customer prefers certain materials or a set of materials, without having a proper way of expressing this. Clearly, the resulting beneficial feature of the CDA on customer analytics is that the competition will be more open, and in the end, benefiting customers, as their preferences will be addressed in a more open and concrete manner.
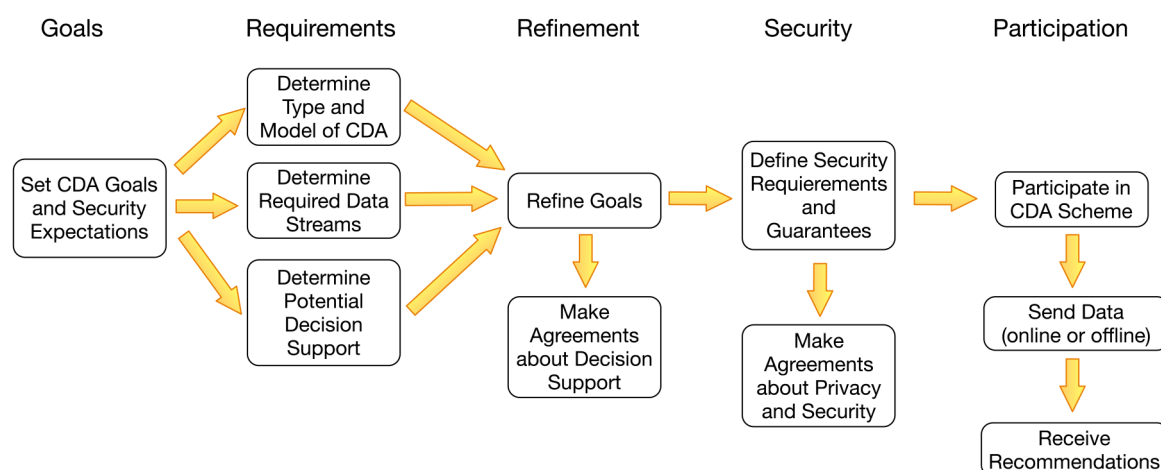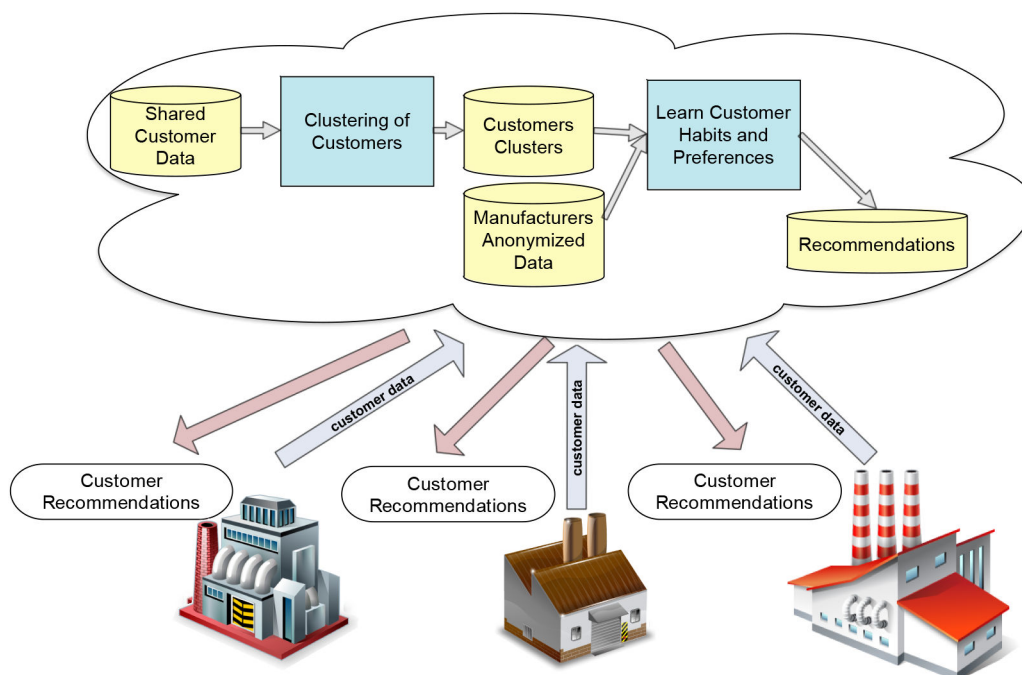

Fig. 2 *CDA participation framework*

Fig. 3 *Illustration of the Customer Analytics case study example*

Table 3 CDA framework challenges

CDA steps \aspect	Customer analytics	MOP analytics	PHM analytics	LO analytics
goals/ requirements/ refinement	*- goals may be overambitious and thus undergo many refinement cycles*	*- no significant challenges, also needed data is relatively automatically collected*	*- setting goals should be unproblematic*	*- setting goals should be unproblematic*
security	*- data protection legislations will need to be considered*	*- competitors might be afraid to share operational data* *- more extensive anonymisation mechanisms might be needed*	*- no special protection mechanisms are needed*	*- due to collaborative nature of logistics, much of this data is already shared*
participation	*- extensive agreements might be necessary to enable participation due to data sensitivity*	*- no significant challenges*	*- intrusive fault and failure event logging will be necessary for participation*	*- should be unproblematic*

4.3 Benefits and challenges of the CDA framework

The presented CDA framework has apparent benefits as more data results in higher accuracy. Furthermore, CDA benefits target both manufacturers and customers. These benefits will, however, come at a price and an adjustment period that will be most likely difficult for manufacturers and it will need adequate supporting actions to enable the CDA. In line with this, in Table 3 we present the anticipated challenges for implementing the CDA framework, by relating the steps with the four data analytics aspects. In the table we have coded the anticipated level of challenges for each data analytics aspect and steps (bold-italic: highest, bold: medium and italic: low), observing that the CDA for customer analytics is the most challenging aspect and the LO is the least challenging one, followed by the PHM.

We, furthermore, throughout literature identified four main performance metrics for cyber-physical production systems that we can see being positively affected and enhanced by utilising the proposed CDA framework [35–37], detailed as follows:

- *Cost reduction*: Being able to quickly build accurate fault models of machines would imply better maintenance schedules and correspondingly lower downtime of machines, implying increased utilisation and, therefore, production.

- *Operation efficiency*: Again, the accurate fault models will provide production and maintenance managers to be better equipped in scheduling operations.
- *Product quality improvement*: Shared customer data, as illustrated by our case study, can only imply shifting of the focus towards the customer needs and along with the improved operations efficiency, this will allow for wider focus span on the product quality.
- *Improved customer experience*: As illustrated by our case study, CDA on customer analytics will imply a greater focus on customer needs and instead of competing for information on what customers need, the competition will be geared towards who can address those needs in a better way. In this way, the customer experience can be significantly enhanced.

To sum up on the challenges and benefits, in the following Section 5, we provide a theoretical analysis of the potential benefits that CDA can provide.

5 Analysis of CDA for Industry 4.0

While Industry 4.0 and CDA can provide many advantages and benefits for the manufacturing sector, it is important to analyse these advantages and benefits to know the potential gains of utilising Industry 4.0 and CDA features for different manufacturing

Table 4 Variables used in the benefit analysis

Variable	Description
m	number of manufacturing enterprises that use the same manufacturing machines.
n	average number of reconfigurable manufacturing systems in each manufacturer
E	additional energy cost per manufacturing machine per year due to a fault
q	possibility of each manufacturing machine to have a continuing defect per year
d	number of defect incidences required to notice a new defect type
t_1	predictable number of years required by the utilised smart system to find a new defect in an individual manufacturer enterprise
E_1	expected additional consumed energy cost of each factory from a fault before it is noticed
E_{wc}	anticipated total added used energy cost of all manufacturers from this defect type before it noticed
E_m	expected energy consumption percentage because of defects with the CDA approach with m manufacturers weight against the total extra energy consumption due to defects without using CDA
S_m	saving factor of the CDA approach in the added energy expenditure
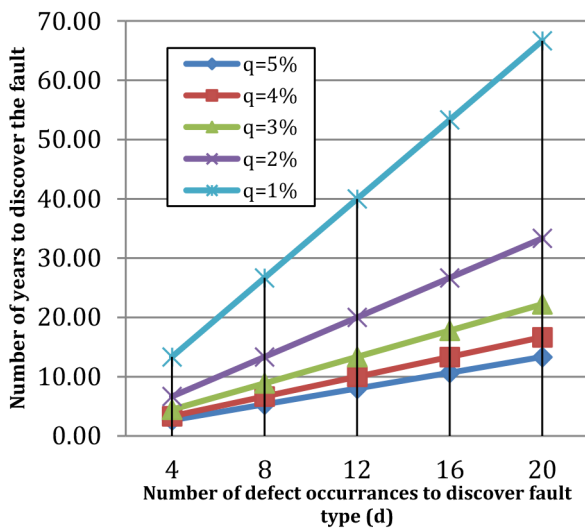
Fig. 4 *Required years to detect a new defect type*

enterprises. In this section, we provide some analyses for estimating the benefits of employing the Industry 4.0 and CDA capabilities for a case study. This case study is to discover a new unknown fault type in some production machines that makes the manufacturing process inefficient [38]. We will compare the benefits of finding such fault in two different situations. The first situation is when an individual solution is used to discover this fault by a single manufacturing enterprise. The second situation is when an Industry 4.0 and CDA solution is employed to find this fault among several manufacturing enterprises that use the same manufacturing machines. For this case analysis, the variables in Table 4 will be used. In this case study, assume there are m manufacturing enterprises employing the same production machines that make some selected products. There are n production machines on average in each of these enterprises. There is an unidentified defect that may appear in these manufacturing machines. This defect will lead them to spend more energy compared to normal operations making the needed products. For instance, more energy is needed to drill a hole in a certain material plate if an old drill bit is used compared to a recently installed drill bit. This is because there is a need for more pressure and more rotations when using the old one to reach to the same required

depth. Nonetheless, it is generally problematic to notice such alteration straightforwardly until the drill bit is physically checked. Let us assume that this costs an added energy expenses, e Euros per year. The possibility that this defect appears annually in each of these manufacturing machines is q.

For each manufacturing enterprise to discover a new common defect type in their production machines, a smart diagnostic system can be used by each manufacturing enterprise. This smart system monitors the production machines and conducts data analysis to find a new defect type after the occurrence of this defect type d times in the available machines. Let us consider that d is less than n ($d < n$). Let t_1 be the predictable number of years required for the smart system to observe this defect type in a manufacturing enterprise. We have

$$t_1 = \frac{d}{(n \times q)} \text{ years} \qquad (1)$$

To make this analysis easier, we will assume that this defect happens to the machine in uniform time distribution. Accordingly, the total estimated added used energy of each manufacturing enterprise from this defect type before it is uncovered costs E_1 Euros, where

$$E_1 = \frac{(d \times e \times t_1)}{2} \text{ Euros} \qquad (2)$$

If we replace t_1 with (1), we have

$$E_1 = \frac{(d^2 \times e)}{(2 \times n \times q)} \text{ Euros} \qquad (3)$$

Thus, the total cost of the added energy for all manufacturing enterprises without using any collaboration is E_{wc} Euros, where

$$E_{wc} = m \times E_1 \text{ Euros} \qquad (4)$$

For instance, to assess the added energy expenditure with a single manufacturing enterprise, we will assume that each manufacturing enterprise has in average 30 comparable production machines. The smart diagnostic system is employed to find new common defects. We will also consider that the extra yearly used energy cost due to the cited defect is 5000 Euros per production machine. Fig. 4 displays the expected years for the employed smart system to distinguish the new defect type estimated based on different levels of diagnostics quality and defect rates per year. The diagnostics quality is stated by the number of defects needed to recognise this defect type.

Fig. 5 displays the total added energy cost before exposure under different diagnostic quality with 4, 8, 12, 16 and 20 defects of a comparable kind. We can evidently conclude from both figures that as defect rate is not very frequent, it will take lengthier period to be observed and it will cost more energy expenditure for the enterprise. Additionally, both figures prove that as we employ a better-quality smart diagnostic system, this will shorten the phase taken to notice the new defect. This is an important advantage that contributes to dropping the energy and the total production costs.

The period required to find new defects can be significantly reduced using CDA. If we apply the same smart diagnosis system from all production machines in all manufacturing enterprises, the defect can be observed after t_m years, where

$$t_m = \frac{d}{(m \times n \times q)} \text{ years} \qquad (5)$$

Consequently, the anticipated cost of the total added used energy of all manufacturing enterprises from this kind of defect before observing it is E_m Euros, where

$$E_m = \frac{(m \times d \times e \times t_m)}{2} \text{ Euros} \qquad (6)$$
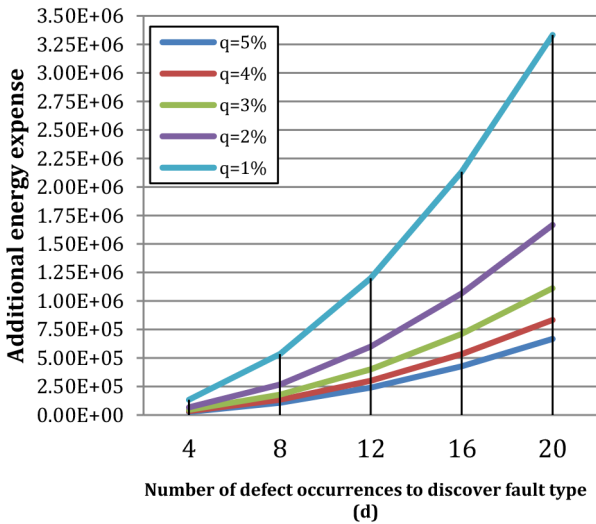
Fig. 5 *Extra energy expenses due to undiscovered defects*

If we replace t_m with (5), we have

$$E_m = \frac{(d^2 \times e)}{(2 \times n \times q)} \text{ Euros} \qquad (7)$$

This is comparable to the added cost of a single manufacturing enterprise regardless of the number of enterprises. Now, to acquire the savings factor of using the CDA method, we need to divide E_m over E_{wc}, and we obtain $1/m$. This means dropping the overhead of extra used energy due to this type of defect to only $1/m$. Besides, the savings factor of the CDA in the additional energy spending will be S_m percent of the additional expenses without CDA, where

$$S_m = 1 - \frac{1}{m} = \frac{(m-1)}{m} \qquad (8)$$

The additional energy-saving factor of the CDA process is shown in Fig. 6. As shown in the figure, this saving factor will be improved with the increase in the number of contributing manufacturers. The saving factor can be 95% using CDA with 20 manufactures. This is a significant saving factor that can participate well in reducing overall manufacturing costs.

6 Discussion

In an era of transforming the industry, large enterprises are usually in better position compared to SMEs. SMEs are typically not well armed with the ability tools, and, thus, require to be more inventive in determining approaches to beat the aggressive competition. The recent challenge is something that not everyone is up to, as the winner is the one who has the largest valuable data that can be utilised for different business and operational improvements. Consequently, small manufacturing enterprises will certainly not have adequate data to compete. As an alternative, if they cooperate by sharing data and data analytics tools and services, the benefits can be substantial.

CDA can form an important tool for survival of SMEs in the new technological revolution. This indicates that this approach is an expected approach for SMEs. Nevertheless, for big enterprises, CDA can also be highly valuable by indicating important savings and improved decision support. The challenge is, although, how to achieve CDA in the best possible manner, such as to guarantee fairness and privacy, as well as keep a healthy competition. In this sense, we expect the emergence of third-party providers of such services who would certify privacy and deliver decision supports for a given cost. For some manufacturing enterprises, if privacy is not of critical worry and they are in a close business relationship, the need for a third-party service could be eliminated, and thus enterprises could save on this. This could be the case also for some aspects of the data analytics that are not sensitive to data sharing.
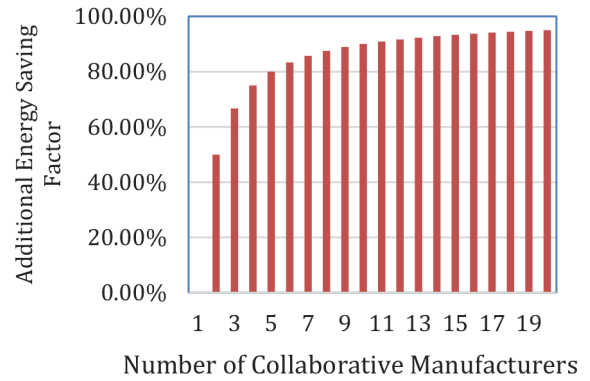

Fig. 6 *CDA saving factor in energy consumption*

Therefore, different CDA approaches for manufacturing will require to be advanced that consider both the data analytics aspect and the business relationships among manufacturing enterprises. This forms a plan for our future work.

However, as previously discussed, for a large number of enterprises, security and privacy are major concerns for sharing their data [39, 40]. Moreover, preventing data sharing can have high risks for some enterprises. To that end, Stefansson [41] investigated >20 cases in logistics operations for both production and transportation and risks associated with preventing proper data sharing with others. Data sharing with ensuring security and privacy for reducing manufacturing costs or increasing profitability can be acceptable for many SMEs. In some cases, reducing costs and increasing profitability are essential objectives to survive and thrive in highly competitive markets. Different technologies can be employed to enforce acceptable levels of security and privacy for these enterprises. Every enterprise can establish clear security and privacy policies and impose regulations for enforcing and applying them.

Different advanced technologies can be used to establish the security and privacy policies of an enterprise and ensure that they are rigorously applied. For example, filtering software programs can be used to remove sensitive information from the shared data. In many collaborative Industry 4.0 applications, complete information about events or transactions are not needed to realise the benefits of the collaboration. As a result, data aggregation and filtering will provide adequate methods to safeguard private information, while still contributing in the collaborative efforts and benefiting from it in their applications. Filtering and data aggregation software can be used for both static and dynamic CDAs. For example, a filtering program can be applied before sharing any data with static collaborative analytics. In the case of dynamic CDAs which will require more technology to be implemented such as cloud computing, fog computing and middleware [42], a local fog node can be used by each enterprise to function as a firewall that has the filtering and aggregation capabilities to control what to share as the data is streamed in real time. This function can be managed by the enterprise itself to apply their sharing rules and ensure their security and privacy.

7 Conclusions

Typical performance metrics for manufacturing facilities are customer satisfaction, productivity, profitability, etc. The data that is easily obtainable is a game changer in addressing the performance criteria and automating the decision-making processes to learn from historical data. The name of the game that captures the ease of gathering data and utilising it for enhancing manufacturing processes, is Industry 4.0. Even though Industry 4.0 is still at the level of a concept, it is happening through the multitude of tools that utilise data to automatically optimise on predefined manufacturing performance metrics. This growing set of tools, however, is still a luxury for the SMEs. To keep up with the competition and the new technological advances, SMEs ought to help each other by joining in their data analytics efforts. Motivated by this, we studied the needs, challenges, and repercussions of instigating CDAs for SMEs, as well as its general

advantages, including advantages for large corporations. We, furthermore, introduced models for performing CDAs, based on the circumstances and limitations of the manufacturer. Finally, our findings were captured in a framework for CDA that identifies and describes the processes that are necessary for its implementation. One of the main challenges that we have identified is the data protection, and such security mechanisms will need to be in place to enable the potential that CDAs has to enhance on many of the performance metrics of manufacturers. This presents a major process in our framework and forms part of our future work, along with the deployment of the framework and its extensive testing.

8 References

[1] Lasi, H., Fettke, P., Kemper, H.-G., *et al.*: 'Industry 4.0', *Business Inf. Syst. Eng.*, 2014, **6**, (4), pp. 239–242

[2] Weyer, S., Schmitt, M., Ohmer, M., *et al.*: 'Towards industry 4.0-standardization as the crucial challenge for highly modular, multi-vendor production systems', *IFAC-Papers online*, 2015, **48**, (3), pp. 579–584

[3] Sommer, L.: 'Industrial revolution-industry 4.0: are German manufacturing SMES the first victims of this revolution?', *J. Ind. Eng. Manag.*, 2015, **8**, (5), p. 1512

[4] Faller, C., Feldmüller, D.: 'Industry 4.0 learning factory for regional SMES', *Proc. CIRP*, 2015, **32**, pp. 88–91

[5] Andulkar, M., Le, D.T., Berger, U.: 'A multi-case study on industry 4.0 for SME'S in Brandenburg, Germany'. Proc. 51st Hawaii Int. Conf. on System Sciences, Hawaii, USA, 2018

[6] Mittal, S., Khan, M.A., Romero, D., *et al.*: 'A critical review of smart manufacturing & industry 4.0 maturity models: implications for small and medium-sized enterprises (SMES)', *J. Manuf. Syst.*, 2018, **49**, pp. 194–214

[7] Radanliev, P., De Roure, D.C., Nurse, J.R., *et al.*: 'The industrial internet-of-things in the industry 4.0 supply chains of small and medium sized enterprises', University of Oxford, 2019

[8] Moeuf, A., Pellerin, R., Lamouri, S., *et al.*: 'The industrial management of SMES in the era of industry 4.0', *Int. J. Prod. Res.*, 2018, **56**, (3), pp. 1118–1136

[9] Lee, J., Lapira, E., Bagheri, B., *et al.*: 'Recent advances and trends in predictive manufacturing systems in big data environment', *Manuf. Lett.*, 2013, **1**, (1), pp. 38–41

[10] Müller, J.M., Maier, L., Veile, J., *et al.*: 'Cooperation strategies among SMES for implementing industry 4.0'. Proc. of the Hamburg Int. Conf. of Logistics (HICL), Hamburg, Germany, 2017

[11] Lazarova-Molnar, S., Mohamed, N.: 'Collaborative data analytics for smart buildings: opportunities and models', *Cluster Comput.*, 2019, **22**, (1), pp. 1065–1077

[12] Lazarova-Molnar, S., Mohamed, N.: 'Towards collaborative data analytics for smart buildings'. Int. Conf. on Information Science and Applications, Macau, China, 2017

[13] Lazarova-Molnar, S., Mohamed, N., Al-Jaroodi, J.: 'Collaborative data analytics for industry 4.0: challenges, opportunities and models'. 2018 6th Int. Conf. on Enterprise Systems (ES), Limassol, Cyprus, 2018

[14] Cheng, G.-J., Liu, L.-T., Qiang, X.-J., *et al.*: 'Industry 4.0 development and application of intelligent manufacturing'. 2016 Int. Conf. on Information System and Artificial Intelligence (ISAI), Hong Kong, China, 2016

[15] Chen, B., Wan, J., Shu, L., *et al.*: 'Smart factory of industry 4.0: key technologies, application case, and challenges', *IEEE. Access.*, 2018, **6**, pp. 6505–6519

[16] Hecklau, F., Galeitzke, M., Flachs, S., *et al.*: 'Holistic approach for human resource management in industry 4.0', *Procedia CIRP*, 2016, **54**, pp. 1–6

[17] Yadegaridehkordi, E., Hourmand, M., Nilashi, M., *et al.*: 'Influence of big data adoption on manufacturing companies' performance: an integrated dematel-anfis approach', *Technol. Forecast. Soc. Change*, 2018, **137**, pp. 199–210

[18] Ren, S., Zhang, Y., Liu, Y., *et al.*: 'A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions', *J. Clean Prod.*, 2019, **210**, pp. 1343–1365

[19] Nedelcu, B.: 'About big data and its challenges and benefits in manufacturing', *Database Syst. J.*, 2013, **4**, (3), pp. 10–19

[20] Zhong, R.Y., Newman, S.T., Huang, G.Q., *et al.*: 'Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives', *Comput. Ind. Eng.*, 2016, **101**, pp. 572–591

[21] Available at https://www.forbes.com/sites/louiscolumbus/2016/12/03/industrial-analytics-based-on-internet-of-things-will-revolutionize-manufacturing/#15caa7836c03, Accessed 2019

[22] Pollard, D., Chuo, S., Lee, B.: 'Strategies for mass customization', *J. Business Economics Res.*, 2016, **14**, (3), p. 101

[23] Kwon, D., Hodkiewicz, M.R., Fan, J., *et al.*: 'IOT-based prognostics and systems health management for industrial applications', *IEEE. Access.*, 2016, **4**, pp. 3659–3670

[24] Wang, G., Gunasekaran, A., Ngai, E.W., *et al.*: 'Big data analytics in logistics and supply chain management: certain investigations for research and applications', *Int. J. Prod. Econ.*, 2016, **176**, pp. 98–110

[25] Wilkinson, D.M., Huberman, B.A.: 'Cooperation and quality in wikipedia'. Proc. 2007 Int. Symp. on Wikis, Montreal, Quebec, Canada, 2007

[26] Schuh, G., Potente, T., Wesch-Potente, C., *et al.*: 'Collaboration mechanisms to increase productivity in the context of industrie 4.0', *Proc. CIRP*, 2014, **19**, pp. 51–56

[27] Yue, X., Cai, H., Yan, H., *et al.*: 'Cloud-assisted industrial cyber-physical systems: an insight', *Microprocess. Microsyst.*, 2015, **39**, (8), pp. 1262–1270

[28] Morgan, J., O'Donnell, G.E.: 'The cyber physical implementation of cloud manufactuirng monitoring systems', *Proc. CIRP*, 2015, **33**, pp. 29–34

[29] Lazarova-Molnar, S., Shaker, H.R., Mohamed, N.: 'Reliability of cyber physical systems with focus on building management systems'. IEEE Int Workshop on Communication, Computing, and Networking in Cyber Physical Systems (CCN-CPS) in association with IEEE Int'l Performance Computing and Communications Conf. (IPCCC 2016), Las Vegas, USA, 2016

[30] Lazarova-Molnar, S., Mohamed, N., Shaker, H.R.: 'Reliability modeling of cyber-physical systems: a holistic overview and challenges'. 2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), Pittsburgh, USA, 2017

[31] Ming, X.G., Yan, J.Q., Wang, X., *et al.*: 'Collaborative process planning and manufacturing in product lifecycle management', *Comput. Ind.*, 2008, **59**, (2–3), pp. 154–166

[32] Trkman, P., McCormack, K., De Oliveira, M.P.V., *et al.*: 'The impact of business analytics on supply chain performance', *Decis. Support Syst.*, 2010, **49**, (3), pp. 318–327

[33] Souza, G.C.: 'Supply chain analytics', *Bus. Horiz.*, 2014, **57**, (5), pp. 595–605

[34] Anitha, P., Patil, M.M.: 'A review on data analytics for supply chain management: a case study', *Int. J. Inf. Eng. Electron. Business*, 2018, **10**, (5), p. 30

[35] Lee, J., Lapira, E., Yang, S., *et al.*: 'Predictive manufacturing system-trends of next-generation production systems', *IFAC Proc. Vol.*, 2013, **46**, (7), pp. 150–156

[36] Li, H.-C., Rong, C.-H., Song, D.-X.: 'Applicability and methods of lean production in railway transportation organization: a case study of urumqi railway bureau in China', *Int. J. Railway*, 2008, **1**, (2), pp. 45–58

[37] Li, D.: 'Perspective for smart factory in petrochemical industry', *Comput. Chem. Eng.*, 2016, **91**, pp. 136–148

[38] Mohamed, N., Al-Jaroodi, J., Lazarova-Molnar, S.: 'Leveraging the capabilities of industry 4.0 for improving energy efficiency in smart factories', *IEEE. Access.*, 2019, **7**, pp. 18008–18020

[39] Alani, M.M., Alloghani, M.: 'Security challenges in the industry 4.0 era', in '*Industry 4.0 and engineering for a sustainable future*' (Springer, Cham, Switzerland, 2019)

[40] Ervural, B.C., Ervural, B.: 'Overview of cyber security in the industry 4.0 era', in '*Industry 4.0: managing the digital transformation*' (Springer, Cham, Switzerland, 2018)

[41] Stefansson, G.: 'Business-to-business data sharing: a source for integration of supply chains', *Int. J. Prod. Econ.*, 2002, **75**, (1–2), pp. 135–146

[42] Al-Jaroodi, J., Mohamed, N., Jawhar, I.: 'A service-oriented middleware framework for manufacturing industry 4.0', *ACM SIGBED Rev.*, 2018, **15**, (5), pp. 29–36