The Digital Manufacturing and Design (DiManD)

Grant agreement No 814078– H2020-MSCA-ITN European Training Network Grant

Deliverable 5.3

Using Big Data in Shop-floor – Challenges and Approaches

Jun 2022

Lead parties for Deliverable : UNINOVA
Deliverable due date : M38
Actual submission date : M38
Dissemination level : Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Table of Content

Summary

Big data can be defined as potentially large dataset that can either be structured and unstructured. In manufacturing, shop-floor big data incorporates data collected at every stage of the production process. This includes data from machines, the connecting devices and even the operators that perform the manufacturing operation. The large size of the data available on manufacturing shop-floor presents a need for establishment of tools and techniques along with associated best practices to leverage the advantage of data-driven performance improvement and optimisation. In work carried out, the data life-cycle in manufacturing is studied with focus in each of the component.

The main driver is to get an insight on the aspects of correlation among data gathered through multiple data sources and relate it to certain causality. Various applications are studied to better this understanding, i.e, establishing a certainty aspect. The grouping of data points, forming a trend or classification, is extremely vital to study the impact of variations on production performance. This grouping of data assists in identification of outliers that have a significant impact on the process.

Machine learning and reinforcement learning techniques form the building blocks of the intelligent manufacturing paradigm. The research assist in identifying possible application and challenges of these techniques in manufacturing, along with the area where they might have most effect (in data life-cycle). Application and challenges of data sourcing, collecting, data transmission, storage processing, and data visualisation are discussed that will help in better understanding of data-cycle elements on shop-floor in manufacturing. This understanding will be beneficial to attain data-driven objective for the production process by application of intelligent manufacturing techniques.

Team involved in deliverable writing:

ESR 1 : Fan Mo, The University of Nottingham
ESR 2 : Agajan Torayev, The University of Nottingham
ESR 3 : Angela Carrera, Mondragon Unibertsitatea
ESR 10 : Luis Estrada-Jimenez, UNINOVA
ESR 11 : Terrin Babu Pulikottil, UNINOVA
ESR 13 : Joaquin Peralta, Mondragon Unibertsitatea
ESR 14 : Hamood Ur Rehman, TQC Ltd.

Supervisors : Prof. Jose Barata & Sanaz Nikghadam, UNINOVA

Chapter 1

Introduction

The evolution of data storing and analyzing has been a key factor in the development of manufacturing processes. During the pre-industrial revolution, low amounts of data were stored and were mostly transmitted verbally, which led to low production volumes and low quality products. Thereafter, during the first industrial revolution, two kinds of data were being recorded, machine and worker data. Worker data (attendance and performance) helped improve productivity and machine data helped better in maintenance. Mass production model introduced in the second industrial revolution also shifted the job of data processing to educated managers. Scientific methods and statistical model helped in all stages of manufacturing from production planning to inventory management [1]. With the introduction of Information Technology in manufacturing, computer systems like CAM, FEA etc. and Information systems like MES,ERP etc. helped in product creation, process optimization and management. The merge between data and manufacturing in the information age has helped in the shift from dedicated production to flexible production. The extension of Information Technology with unified communication (ICT) further enhanced the role of data in manufacturing.

The concept of Smart Manufacturing (SM) emerges as a new paradigm focused on responding in real time to the ever-changing demands and conditions in factories, supply networks and customer needs [2]. The key Smart Manufacturing (SM) technologies are Cyber-Physical System (CPS), that integrate physical assets with their computational capabilities; Internet of things (IoT), highly connected devices embedded with sensors; and big data [3]. The big data age arises with the massive use of mobile and smart devices, the great availability of IoT devices and cloud computing when traditional methods were not enough for efficient information processing [4]. In general, the term big data refers to the storage and analysis of data sets that are characterized by their large *volume* and *variety* of sources; the high *velocity* at which they are generated and must be processed; and the *value* generated by its analysis [5].

In the age of big data technologies, various data sources are present, and data are collected from connected devices, software solutions, sensors, and Internet of things (IoT). Manufacturing data can be categorized into management, equipment, user, product, operational, and process data on a high level [1, 6]. On a low-level, manufacturing data is categorized into structured, semi-structured, and unstructured data [7]. Structured data have a clear relationship between its attributes, and it is the easiest data type to store and organize. Structured data are usually represented using tables. Unstructured data comprises most manufacturing data, has no associated data model, and cannot be organized using tables or spreadsheets. Examples of unstructured data are images, audio, text, video. Semi-structured data do not reside in relational databases but have an organizational structure that makes them easier to analyze. Examples of semi-structured data are XML,JSON and

HTML.

The collection and processing of the data in the shop floor is critical as most of manufacturing operations are carried out here. The advent of IoT and new industrial protocols have supported the acquisition of the in formation from manufacturing cells, products, transport systems and even people [8]. Thus, many data-driven manufacturing applications have emerged recently e.g. Smart: "design", "planning and process optimization", "Material distribution and tracking", "Process monitoring", "Quality Control" and "equipment maintenance" [1]. Those applications rely on the transformation of primary data to information use to make the process more intelligent. Examples of shop floor data are energy consumption, quality test, equipment status, equipment parameters, resource loading, delivery time, material data, etc [1]. Despite the benefits foreseen by the usage and processing of data in the shop-floor, there are challenges that need to be considered.

The 5Vs characteristics of big data are widely acknowledged as the challenges of big data in manufacturing, including volume (level of data size), velocity (ingesting or processing big data in streams or batches, in real time or non-real time), variety (dealing with complex big data formats, schemas, semantic models and information), value (analysing data to deliver added-value to some events), and veracity (validate data consistency and trustworthiness) [9]. In addition, there is the issue of cyber security; because the big data platform connects the physical space and the cyber space so intimately, the danger of cybersecurity might swiftly spread to the manufacturing system's physical system [7].

The increasing size of the data on shop-floor promotes a need for accurately classifying the data for reliable decision making . Influx of huge amounts of big data generated from multitude of production systems on shop floor, make this decision making very complicated and strenuous . Combined with multiple data sources, different transmission protocols and storage requirements for production systems on shop-floor make it a very difficult task . This research contribution aims to develop a homogeneous approach to gathering and utilizing data on shop-floor in manufacturing environments taking influence and insight from the literature review. It targets the complete data cycle involving "Data Sources" and making effective use of them for achieving the desired data required for objective completion. This approach also discusses the needs, requirements, and methods for "Data Collection" and "Data Transmission". A manner of homogenising the data acquired is needed on shop-floor as it contains multiple production systems operating on different protocols and other technical requirements . This approach also discusses the "Data Storage", "Data Processing", and "Data Visualisation" applied to shop-floor in manufacturing to achieve the daily production objectives. The contributed approach builds on these aspects of data cycle to elaborate on "Data Application" .

This contribution leverages the data cycle widely used to capture data in Big Data paradigm and leverages it to shop-floor in manufacturing. The suitability and adaptation of data-based manufacturing is the main goal in this contribution. This research work , addressing this need for Big Data on shop-floor, establishes the approach for data acquisition, processing and utilisation for decision making. The challenges faced towards real-time data-based manufacturing is also elaborated.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Chapter 2

Data Life Cycle

Big data, and data in general, requires to be structured into specific content formats and context to be useful for users [10]. Structured data is useful for automating processes in manufacturing, as it enables machines to be able to communicate among themselves and enables users to extract knowledge. Nevertheless, data sources have different formats and may be structured, semi-structured and unstructured. Therefore, research has given focus to the data life cycle and how to extract knowledge from varied, heterogeneous data sources, enabling informed decision making. In this context, the data life cycle in manufacturing for decision making has been presented as consisting of seven stages [1], which are listed in the following. Furthermore, Figure 2.1 presents a visual representation of the seven stages of the data life cycle.

1. **Data sources:** Data sources generate data across all the manufacturing value chain and product life cycle. The data may come from machines and tools, products, users, ICT systems, and networks.

2. **Data collection:** After data sources generate data, data collection is performed. The collection is performed by IoT technologies, by means of smart sensor nodes equipped with sensors, such as accelerometers and temperature sensors, allowing measurement and monitoring of the manufacturing processes and the product life cycle in the following stages. Data collection may be performed at different frequencies, referred to as sampling frequency or sampling rate, based on the processing power of the sensor node and the requirements of the variable being measured. In addition to the shop floor data sources, other data collection sources, such as third-party application program interfaces or web crawling, may be used to collect data, further enriching and expanding the context of the data collected during the process.

3. **Data transmission:** Data transmission maintains the communications between the elements involved in the data life cycle, e.g. manufacturing systems and manufacturing resources. Defining standardized means of transmission, communication and application protocols define how the elements communicate data among each other, allowing real-time, secure, and scalable data transmission. As with data collection, data may be transmitted at different frequencies, based on the requirements of the monitoring strategy, such as real-time data transmission or batch data transmission.

4. **Data storage:** Data obtained during data collection must be stored securely and integrally. Considering that data sources may be structured, semi-structured and unstructured, several

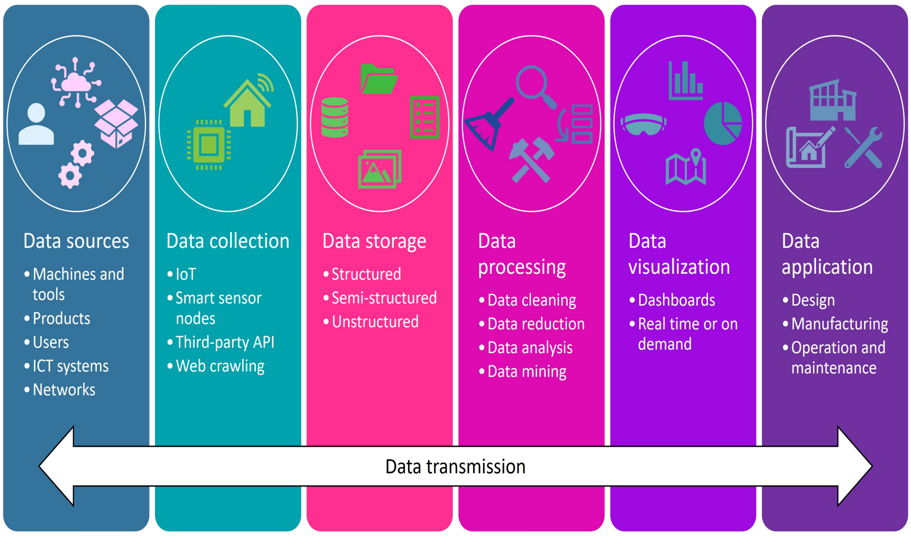

Figure 2.1: Example of data life cycle

different storage types may be considered. To this end, besides structured data storage, object-based storage provides a flexible solution for storing semi-structured and unstructured data, thus covering the integrity requirement of data storage. In addition, by means of cloud computing, data storage may achieve cost effectiveness and high-processing power, as well as security, scalability and heterogeneity.

5. **Data processing:** Data processing builds upon data storage and refers to the operations required to extract information, i.e. knowledge from heterogeneous data sources. By processing raw data, hidden information and patterns may be revealed, providing stakeholders with valuable information for decision making. Data is processed by means of data cleaning, data reduction, data analysis, and data mining techniques. Furthermore, data processing has become more efficient recently, owing to advances in artificial intelligence, cloud computing and IoT.

6. **Data visualization:** Data visualization provides the means to visually understand the information extracted during data processing. Data may be visualized in dashboards, including statements, charts, graphs and augmented reality. In addition, data may be queried in real time or on demand, based on the users needs, enabling decision making based on historical or real-time data.

7. **Data application:** Data application refers to data analytics performed during the entire product life cycle, providing stakeholders with tools for decision making. Data analytics may be applied during the design phase, translating customer needs into product features and quality requirements. Thereafter, during production, data analytics monitor the production process and lead to informed decision making regarding the manufacturing process, improving product quality and reducing production costs. Finally, during product operation and maintenance, data analytics may be used to predict possible faults and to provide preventive maintenance, elongating the life cycle of the product and improving relationships with costumers.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Chapter 3

Methodology

The work focuses on understanding the approaches and challenges in implementing big data in the shop-floor. The restriction on just manufacturing shop-floor provides focus to the work. Focus on just relevant application of the shop-floor provides another filter before defining the research question.

A general research question is defined for the work - "What are the recent trends and challenges in big data life cycle in shop-floor ?". The research question would be addressed by considering the 6 data life cycle i.e., Data sources, Data collection, Data storage, Data processing, Data transmission and Data Visualization. Considering the wide scope (address all 6 data life cycle) of the work, a narrative review is adopted with certain criteria.

After literature collection and considering the research question, relevant information was extracted from the paper for each data life cycle. The results of the research question helped in providing interesting insights and derive the current challenges faced in adopting big data in the manufacturing shop-floor. The methodology adopted for the review is represented in figure 3.1.

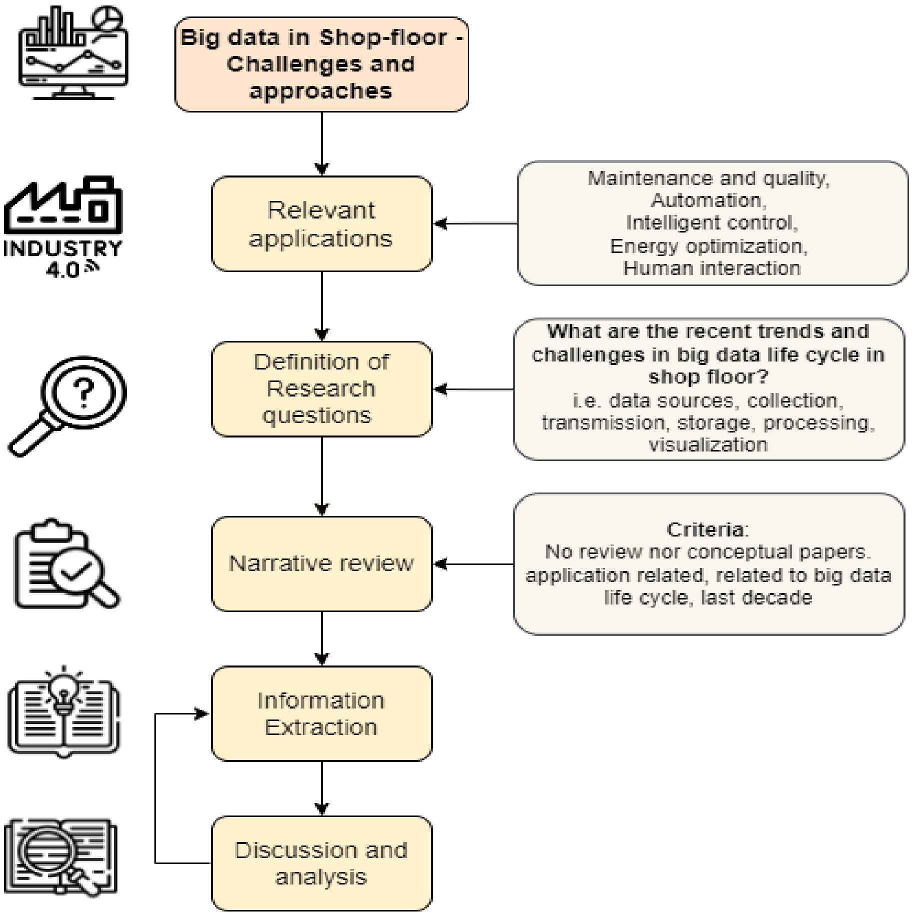

Figure 3.1: Adopted Methodology for the work

Chapter 4

Results

This section explains the results of the narrative review of publications related to big data life cycle. The section is divided into different stages of data life cycle. The results presented are a collective overview of the publications presented in the last decade on each of these stages related to big data in manufacturing shop-floor.

4.1 Data Sources

Different applications in the context of smart manufacturing require different data sources. They are mostly based on the utilization of internet of things devices i.e. sensors that take relevant information of machines, shop-floor, products, people and environmental variables. Another important source of data are the ones provided by heterogeneous product requirements, specially in product driven manufacturing applications.

For decision making activities, one example of data sources are customer requirement documents, datasets or CAD models. These sources are multi-modal with different forms and hence require separate processing methods. Another example are information embedded in CAD models. In this case, Collada can be used as the data format to describe CAD models. If the CAD is modeled in CATIA V5, then the converter from CATIA V5 to Collada can be used to get the Collada model[11].

For monitoring energy in a shop floor are smart meters, current and voltage clamps, or machine-integrated devices that provide out-of-the-box instantaneous power consumption [12]. Industrial robots, for example, can provide the power consumption for each joint of the robot directly from a robot controller [13]. Experimental data regarding actuation torques and servo drive voltages, directly used to derive the plant input power, can be captured using energy sensors such as clamps[14]. Alternatively, single-phase and 3-phase smart plugs are becoming popular for monitoring the energy consumption of manufacturing equipment on a shop floor [15].

Human data can also provide additional context information to the current shop-floor situation. This data provide a better user experience for operators, improving productivity and decision quality. Human data can be divided into human attribute data and state data. Human attribute data comprises demographic and characteristic information that will not change or change sporadically (i.e. age, profession, education status, skills). This data can be later used for "user modelling" to deliver information or services according, for instance, to the proficiency, skills or interest of the user. In the other hand, human state data refers to a collection of all kinds of data that eventually allow the modelling of abstract human characteristics, such as behaviour, comfort, etc [16]. Traditional IoT devices can be used to acquire data about operators' state (current position, vital functions, etc.). For

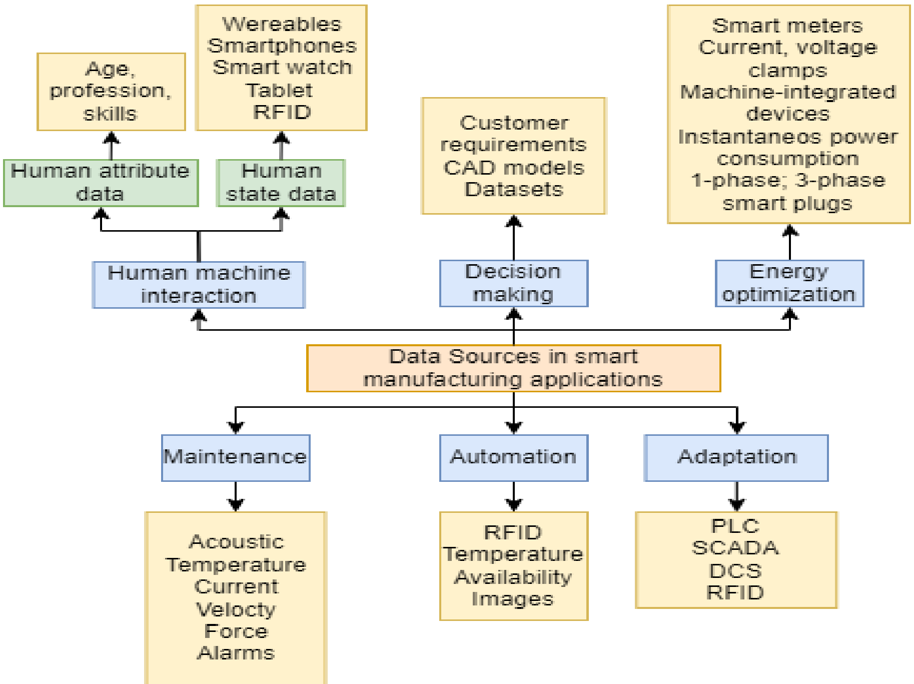

Figure 4.1: Data sources in smart manufacturing applications

instance, wearable trackers can measure human performance under stressful or difficult conditions analyzing them and sending warnings if needed [17]. Furthermore, operators can use portable smart devices (smartphone, smartwatch, tablet) with NFC (Near Field Communication) readers to check into a location and receive information about relevant parts of the production system equipped with NFC or RFID tag.[17]. The behaviour can also be inferred through the interactions that users have with machines or applications by capturing these with the use of plugins or applications such as Google Analytics or Matomo. All of the acquired data can be sent to the cloud through IoT services, which can be processed and analyzed to deliver personalized information to operators and supervisors or inform them about issues.

Most of data driven automation applications rely on optimal decision making considering status of machines or conveyors(availability) [18], by using smart sensors to track equipment and people e.g. RFID tags [19, 20, 21, 22], monitoring best conditions of machines in terms of temperature e.g. [23] or by using information of images (quality control) that works as a decision factor for the autonomous reconfiguration and adaptation process [24].

Data-based maintenance sensors include vibration [25, 26], acoustic emission [25, 26], temperature [21, 25], current [25, 26], velocity [21], pressure [21], and forces [27] from various parts of the machine. These sensor either exist in the machine [28] or are add-on sensors dependent on the application. The PLC controllers provide process related data like cutting speed, feed, depth of cut and so on [25]. Certain application specific data sources also aid in monitoring and maintenance activities. For example, 3D laser scanner to evaluate the tool flank wear [26]. Another source includes device status (alarms, logs etc.) [28] and historical failure data [29] logged after quality inspection which aid in identifying product failure patterns. RFID tags also aid in identifying the defective products to compare with the failure data [21].

Accuracy and quality of data play a vital role in successful implementation of intelligent systems. This depends on the effectiveness of data sources. Utilisation of this accumulated data may result in data gaps and incompatibility in system applications, to overcome which proper calibration needs to be carried out. Data sources consist of automation system resources (like sensors, actuators, PLC, SCADA, DCS and CNC systems) and identification systems (like RFID, AutoID, barcodes and vision systems etc.), communication standards between production resources (like fieldbus, wired and wireless communication) along with accompanying data exchange standards (like OPCUA, MTConnect, MQTT etc.). Automation technologies allow a significant reduction of human participation on the shop-floor during production operation. There can be processes, however, that are not automated, mainly due to infeasibility of economic outcome. Specific production processes may involve manual work to be carried out in different manners. The employee carrying out the work may enter the information to a management support system. As per research, the information accumulated from employees through this approach is highly unreliable and cannot be used for machine adaptation. Automated production systems assist in automated data acquisition without human intervention. Data accumulated in this manner can be used simultaneously for purposes but certain pre-processing and appropriate interfaces may be necessary. Most common data sources in automated production systems for machine adaptation can be control and measurement devices, measurement instruments like sensors and transducers, PLCs (and other control mechanisms) and robots etc.

4.2 Data Collection

The data collection techniques for decision-making are dependent on the data sources. In case of customer requirement, natural language processing techniques such as: named entity recognition [30], relation extraction [31], and attribute extraction [32] are utilised. If the data comes from the dataset, some deep learning techniques and sampling techniques can be used to collect the data[33].

There are mainly two types of data collection techniques; manual data acquisition and automatic data acquisition. Manual data acquisition techniques are employee dependent and are gathered through a manufacturing support system, but are highly inconsistent and unreliable [34]. Automated data collection is through automated systems like sensors, measuring, and control devices that correspond to changes in physical process [35].

Data is sourced through primary and secondary data sources and collected sequentially through different physical events. This accumulated data is of low value density when treated individually, but together form great value. This value is extracted by evaluations, simulations, and predictions.

Data collection in production system, depends on the nature of the gathered data that may be structured or unstructured [36]. Multiple frameworks are in-place that incorporate data collection strategies for structured and unstructured data [iteazad2020role. Simply, the data collection for machine adaptation is a six-step process involving Initialisation, Configuration, Capturing, Analysing and Focusing [37].

[7] stated that almost half of big data collection applications were distributed in monitoring (25%) and predictive applications (24%), characterized for real-time process and non-real-time process respectively. Real-time process data analysis in manufacturing refers to methods where data from production lines is acquired, processed and delivered to operators, in order to timely detect anomalies or to quickly know the status of the shop floor, production, machines, and personnel[38]. This is one of the basic needs for operators on the shop floor, who require a synthesized and centralized view of multiple sources of data that can even be highly dispersed. On the other hand, predictive applications do not necessarily require a real-time process, and focuses on extracting patterns and trends based on historical process data for optimization and management innovation [38].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

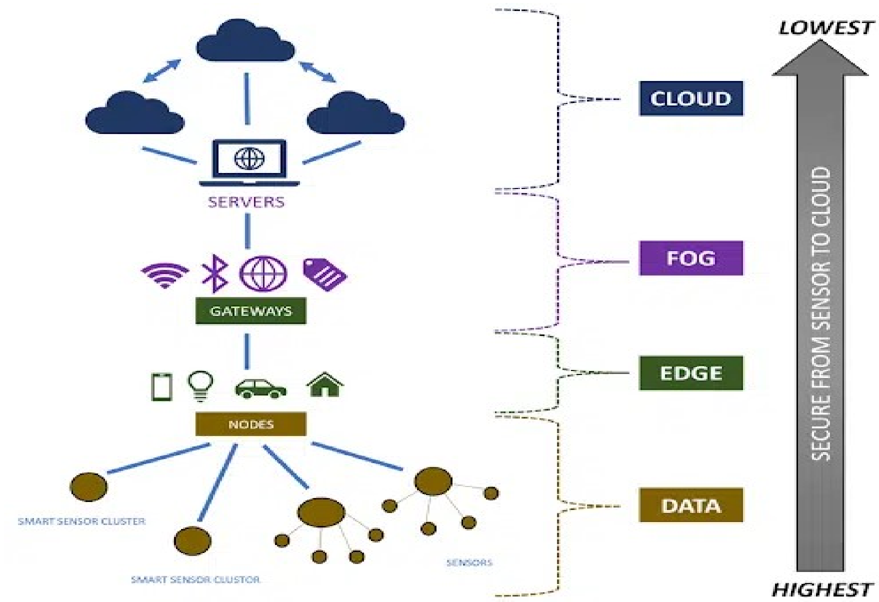

Figure 4.2: Data Collection in Manufacturing applications

Even though real time data collection is preferred, In practice it is seldom the case for maintenance related data. Add-on sensors like temperature, vibration, pressure & forces and PLC controllers for process data like cutting speed, feed, depth of cut provide near real timely information (every min). device status and logs [28] are periodically collected and stored. Wear information is collected after a predefined amount of time to accurately analysis the wear (e.g. tool wear is measured every 20min in [25]).Some process parameters and performance metrics (Non-real time data) are provided after each production run/shift [21] like maintenance history, failure record [29], OEE, resource consumption etc. Almost all the data relevant to monitoring or maintenance are time series data and have a time stamp while collected. Data collection techniques include support to Restful / configurable application layer protocol, OPC unified architecture, distributed data acquisition (e.g. Flume [28]).

Automation activities rely on event driven data collection techniques e.g. time driven, quantity driven, operation driven [39]. Event driven approaches allow the storage of manufacturing information after a specific time interval. These techniques are also useful to query manufacturing services for process automation purposes. Optimal decision making usually require storage of historical data and and the comparison with a real time monitoring data extraction [21].

For time driven data collection, energy data for manufacturing equipment can be studied. Energy is usually monitored every given time interval for monitoring total energy consumption, such as every 15 minutes. However, some applications, such as profiling the robotic motions and understanding the parameters affecting the energy consumption, requires real-time energy data sampled every few milliseconds [40].

4.3 Data Transmission

The data transmission can be sockets, OPC-UA, MQTT, TCP/IP (such as PLC simulator), or other communication protocols (Figure 4.3) depending on the application domain and can be dynamically

This project has received funding from the European Union's
Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie
grant No. 814078

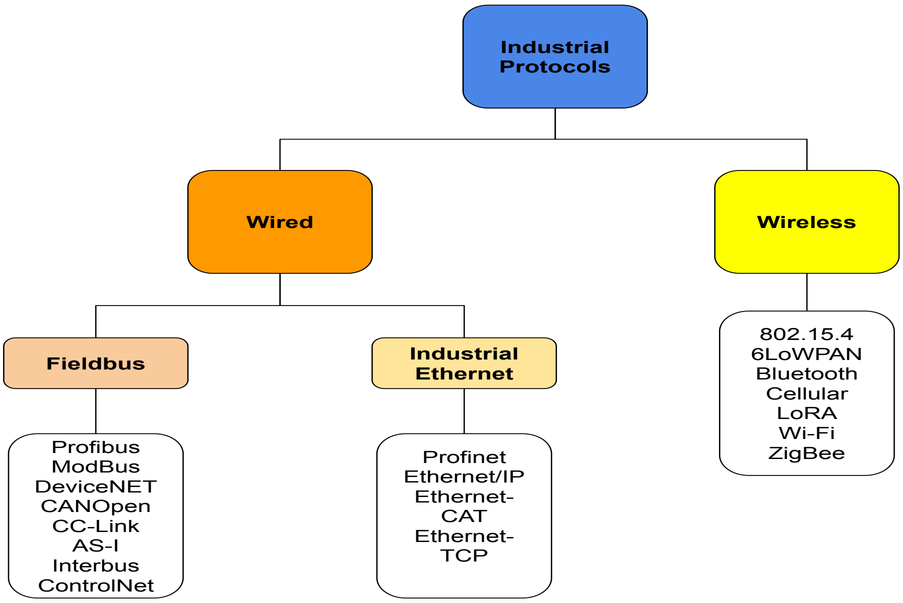

Figure 4.3: Industrial protocols

chosen. This communication method will be suggested to be used as the middle-ware between the digital twin and the optimization environment. And it can also be used as the communication channel between different devices in the digital twin and the real devices. If the workstations in the manufacturing system use different operating systems, then OPC UA is a better solution. Cloud-based optimization is recommended in this system because it promotes modularity among components of the pipeline [41].

The transmission of energy data for further processing depends on the logging frequency of energy data. Usually, an ethernet connection is used for transmitting the data. However, high-frequency energy data is first stored in an external memory device of an energy monitoring solution. After some time, all the collected data is transmitted manually to the processing computer. Some energy monitoring solutions offer transmitting data via WiFi which is an advantage and disadvantage at the same time. Transmitting energy data via WiFi is transport flexibility and high transmission distance, but WiFi comes with shortcomings such as high latency and transmission unreliability. Hence, industrial standards such as Modbus and Profinet are used for mission-critical applications[42, 40].

Process automation require the connection of manufacturing resources to the internet. Generally it is either by using Ethernet [18] or wireless communication [19]. Decision making, negotiation and data acquisition can be implemented using some industrial standards with higher reliability: OPC-UA, Modbus, Profibus [43]. Internet of things communication can be used to perform data transmission in a public subscribe manner e.g. MQTT protocol [24] for event driven process automation purposes.

Real time data could be transmitted through internet, WiFi, Zigbee, 4G & VPN and non-real time data are transmitted through various technologies or application like Sqoop Apache or Data/X [21]. Production data and some sensory data (high performance sensors with very high frequency) are transmitted through Ethernet to a local server and then after feature extraction are sent to cloud servers through WiFi protocol[25]. The introduction of IoT in the shop-floor has increased the transmission of low frequency sensor information directly from the source through WiFi from various sources. This has also had an impact on the latency of the system's response. Industrial wireless network include Industrial switch, Industrial routing and Wireless AP.

Data transmission for machine adaptation is carried out through supported data transmission protocols like OPCUA , MQTTT, MT Connect etc. Data transmission rates play a vital role that depend on the manufacturing application. In order to incorporate multiple data formats, standards, and needs for machine adaptation, a combination of technologies is proposed that assists in data transmission. A framework in this regard is necessary that aid in data transmission across production domain.

4.4 Data Storage

Common data formats to store machine information are XML and JSON files [19]. Different data types include Structured (data presented in tables), Semi-structured (XML, JSON, HTML) and unstructured data(documents, images, audio, video, text, emails)[39]. Unstructured data are first processed to extract relevant information internally before being stored in the database. For example, Tool Wear information are extracted from the wear images using a image processing software available with the machine and converted into flank/crater wear values along with their time stamps [26]. Depending on the type of data, they are stored using different techniques.

Traditionally, Relational DataBase Managers (RDBMS) and DDBS are used to store structured data. RDBMS are characterized by well-defined schemas and relationships. Basic user information can be stored in traditional databases such as Mysql, SQL, Postgre, and SQL Lite. RDBMS have been used for interaction data storage. For instance, Matomo, an user analytics platform, captures the user's interaction stream (i.e. clicks, page views) in a Mysql or MariaDB database. These kinds of databases offer limited scalability.

NO SQL databases (i.e. Mongo DB, Cassandra) are a better approach for semi-structured (JSON, XML) and unstructured data (audio,video etc.). XML is also used exchange between structured to semi-structured [21]). Hadoop Distributed File System (HDFS) could also be used for dealing with unstructured data. Some examples of these kind of databases include,

- Cassandra to store event data of automation controller
- MogoDB (document NoSQL database) to store machine data
- Time sensitive DBs (TSDBS) like OpenTSDB and InfluxDB to store and access sensor data
- Influx DB and DalmatinerDB for time-stamped or time-series data

Data models are another way of represent the manufacturing data. Data models include two parts - run time conditions (process knowledge & time-sensitive dimension) and process model (product's production requirement). Once the data models are defined, knowledge graph can be used to store the datas. There are two main types of storage for knowledge graphs. One is RDF-based storage; the other is graph database based storage. An important design principle of RDF is the ease of data distribution and sharing, while graph databases focus on efficient graph queries and search. The Neo4j system is a widely used graph database [44]. It has an active community, and the system itself is efficient in querying, but the only shortcoming is that it does not support quasi-distribution.

Smart manufacturing applications tend to use distributed file systems (for data-at-rest) and databases (for data-at-motion) for processing and storage [18]. Historical data are ingested to and from databases, to predict the production planning performance, safety critical aspects and network

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Table 4.1: Data storage types and technologies used in manufacturing shop-floor

Data Storage Type	Data Storage Technologies
Relational database (RDBMS)	MySQL, SQLlite, Oracle DB, SQL server, ProgresSQL
NoSQL database	Column-based: HBase,Cassandra Document-based: MongoDB Key value-based: Redis Graph-based: Neo4j
NewSQL database	VoltDB
Other data storage types	Time series data base : OpenTSDB Search engine : Solr, Elasticsearch, SparkSQL Data warehouse : Hive, Kylin ETL (Extract, Transform and Load) : Pig Others : HDFS, Clustrix, NuoDB

designs. To reduce the amount of space needed for storing some tools used are Hadoop and Map reduce.

Production and Sensor data from the machines are initially stored in industrial computers connected to each machines, which are then processed internally for feature extractions and understanding the machine state [25]. Its later sent to cloud servers where the data are managed and stored in the database. The cloud acts as a remote server for data storage.

Automation applications are relying in the storage of manufacturing information as well as services, increases the responsiveness and interoperability of the shop floor and thus the capacity of automation. The choice of storage solutions greatly affects the application. High-frequency big data files require special solutions such as Hadoop and Spark that can deal with the high volume property of Big Data. The energy data is usually recorded in regular timestamps, which results in time-series data [45]. There are special database solutions for storing time-series data, such as InfluxDB. Also, relational database methods are used in energy data for their reliability. Some monitoring solutions store the collected energy in device memory using comma-separated values (CSV) files.

4.5 Data Processing

When information is gathered and transformed into usable form, data processing takes place. Data processing must be done appropriately in order to avoid having a detrimental impact on the final product, or data output, and is typically carried out by a data scientist or team of data scientists. Different techniques can be used for data processing. (Figure 4.4)

In energy consumption area, data processing is a computationally intensive task. First, the energy should be resampled to match the recorded timestamps. Resampling methods such as averaging, forward filling, or backward filling are usually used in the literature [46]. The averaging method takes an average value within a pre-defined time interval and replaces the missing values with average energy consumption. In forward- and backward filling methods missing timestamps are filled with values before or after the missing timestamp, respectively. Once the data is processed, it is fed into application-dependent algorithms such as ARIMA, SARIMA, Bayesian Optimization, clustering, neural networks [47], genetic algorithms [48] and parameter identification methods [49].

As for processing the data in decison making, there are different approaches. First, in a large number of process text documents, a method based on multi-neural collaboration is used to extract

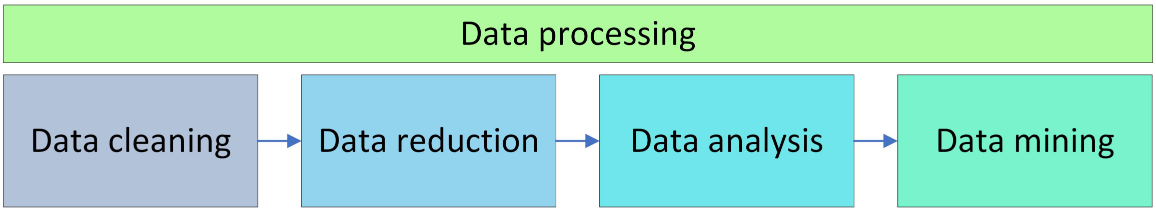

Figure 4.4: Data processing process

knowledge, and the extracted knowledge is classified accordingly through tags. At this level, ontology model and schema layer of the knowledge graph should be defined. Then the knowledge should be represented based on fuzzy comprehensive evaluation [50]. Some knowledge can be directly described as the production rules [51], some knowledge is more suitable to be described as knowledge graph [52]. At last, due to the wide range of knowledge sources, the knowledge base constructed according to the two steps above has high redundancy, so it is necessary to use latent semantic analysis, similarity calculations and attribute weighting to eliminate redundancy in the knowledge. First, the entity triples in the preliminary knowledge base are mapped with the Protege ontology library, and then the semantic web rule language (SWRL) is used to represent the empirical rule knowledge. Finally, the data layer is instantiated to construct the final knowledge base [53].

As for the data processing in HMI, in addition to the use of several data mining and machine learning techniques, the development of analytic solutions will require selecting the right strategy according to diverse scenarios. Streaming, small-batch, and large-batch analytics are the three main processing strategies for big data [54]. Batch processing is the most traditional form of processing where big volumes of data are collected, that can represent a large period of time (i.e. hours, day, week) and analyzed over very complex machine learning models. Here, real-time is not a priority. Streaming is a processing technique for real-time analysis of data streams, particularly necessary when data arrives at high velocity. Small-batch processing (also known as micro-batch) is the process of small cumulus of data on a small time window (i.e. minutes).

Data processing can be also used for automation.Intelligent decision making for process automation and self-organization requires the analysis of machine status and energy consumption. This makes necessary the use of machine learning techniques. Some examples for process automation include: Neural networks, Support vector machine, K-Nearest Neighbours [23]. Negotiation based approaches with machine learning can be found when choosing proper routing or transportation of products e.g. for storing or scrapping them[24]. Genetic algorithms are also used under the scope of ML. For process automation genetic algorithms allow the finding of optimal production resources e.g. the ones with minimum energy consumption or the ones that require less production time. In general, classical machine learning techniques are enough for this type applications.

Besides, in maintenance sector, feature extraction of the time series data from sensors like vibration/forces include both time-domain and frequency domain feature extraction. Time domain features include RMS, peak, mean, standard deviation, skewness, Kurtosis, Crest factor and so on [25]. Frequency domain include Main frequency, harmonics, freq. band energy% etc. It is relevant only for high frequency data to be considered in frequency domain after noise reduction in the signal. Data and pattern mining models for maintenance (e.g. Apriori [21] or FPGrowth [29]) could be used for knowledge and rules generation. Generated knowlege along with the production data could aid in fault diagnosis & prediction. Correlation analysis provides internal relationships between device and faults [28].

Traditional and Deep machine Learning techniques are used for data analytic. Clustering is

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

by far the most common machine learning techniques used for the preliminary grouping of the sensor information and to create labels according to their process state [25, 29]. This is followed by classification techniques based on traditional (e.g. K-Means in [29]) or deep learning (e.g. CNN in [27]). Some technologies used for data analysis in maintenance include STORM [21] (distributed computing), STORM cluster [28] (resource scheduling), Hadoop [21] (offline prediction - considering both current status & historical information).

The collected data needs to be processed to generate insights. Primary steps in data processing involve cleaning the data to remove noisy and incorrect format issues. Streamlink(Flink, Storm), micro-batching (Spark)and batching data processing (MapReduce) provide technologies to clean and process big data volumes. Manufacturing applications like complex event processing by Storm, and detecting deviations by Flink, prediction and quality control by MapReduce are some examples where these technologies are used to process manufacturing data. Knowledge can be generated by harvesting big data technologies on the generated big data. Apache Hive-Mind based platforms has aided to this knowledge generation for predictive maintenance. Hadoop and OWL technologies can manage knowledge of intelligent applications for smart manufacturing applications.

4.6 Data Visualization

Data visualization is an integral part of data analysis which concentrates on the use of tables and graphs for presenting data, quantitative and qualitative information to the user and for the user to communicate with the data [55]. Few state-of-the-art works describe methods for data visualization in the context of smart manufacturing automation and big data. Usually, it is implemented to create dashboards to monitor and access production status or in some cases as a direct interface between the customer and the shop floor. Thus, dashboards enable the presentation of a grand amount of data such as sensor data, manufacturing planning, and operational and maintenance data.

Dashboards are often interactive and users can filter and query data, zoom in/out and scroll. Many of the visualizations show changes over time and are updated as new data is released. Furthermore, dashboards can display real-time data that is updated every few seconds or minutes. In general, data visualization can include [56]:

- Different types of charts and graphs, tables, time trends, etc.
- Interactive widgets (i.e. knobs, dimers, key pads, etc.) used to interact with CPS, IoT devices and applications, based on current data analysis.
- Visualization of geo-referenced data (machines in different locations, operators location tracking, external sensors)

From the technological perspective, in research, scholars prefer the use of Python programming language to develop machine learning models. Therefore, for data visualization, Python libraries such as Seaborn or Matplotlib are chosen to develop charts and graphs. [57] used matplotlib to visualize a heat map o to find the correlation between the variables involved on milling tool wear (Figure 4.5.**a**). Depending on the tools and technology used (e.g. SQL databases, graph databases), visualisation methods integrated into the development environment can be used[44].

However, these options are not intuitive or designed for end-users. At the moment, multiple platforms and frameworks have come out to produce analytics applications and visualizations in a simple manner with very aesthetically pleasing results. Grafana is one of the most popular open-source platforms for interactive data visualization. [58] used Grafana to create a dashboard for visualising energy data at the workstation level to show operational KPI and power consumption

Figure 4.5: Data visualization. a) [57] Seaborn visualization b) [58] Grafana visualization c)[54] Grafana and Quick-Sight d) [28] web and mobile apps e) [59] HTML5, CSS, JavaScript web application

trends (Figure 4.5.**b**). Similarly, [54] developed dashboards using Grafana and Amazon QuickSight for its compatibility with Spark to display the results of small-batch processing for the detection of anomalies on CNC Machines (Figure 4.5.**c**). Other similar products include Qlikview, Tableau, Kibana and Splunk.

Even when these platforms are claimed for their ease of use; the target users are data scientists and engineers or business analysts or DevOps engineers. For end-users (i.e customers, operators, supervisors) customized applications accessible through mobile devices or web interfaces using browsers[43] are the best option. In [25], a Web and iOS-based user interface is used in real-time for decision making on the assessment of health. In [28], the manufacturing data processed is sent to backstage supporters and the diagnosis or prognosis reports are visualized on large screens through a web application (Single View integrated failure map pattern and cause [29]) or sent to mobile devices of the maintenance personnel (Figure 4.5.**d**). These kinds of applications will require some sort of software development. Javascript is the ultimate web standard for reactive applications, with multiple frameworks such as React, AngularJs, NodeJs, etc. There are specific Javascript libraries that allow the development of interactive visualizations such as CanvasJS or ChartJS. [59] developed a web application for historical analysis and real-time tracking of assembly line performance. The web is created with a combination of HTML5, CSS, JavaScript, the JavaScript Data-Driven Documents (D3) library, the Three.js and several JavaScript framework & utility libraries including Underscore.js, Backbone.js and JQuery (Figure 4.5.**e**).

From the user perspective, it is important to consider that manufacturing processes involve different types of users where multiple variables intervene (i.e. expertise, role, age, etc). Therefore, users will have different perceptions to visual data presentation and interactive data analysis [38]. User-centred design as a methodology can help to understand the requirements and needs of determined roles in industry.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Chapter 5

Discussion

Different manufacturing applications require different data sources. Data sources comprise mostly smart sensors and IoT devices that convert physical variables into digitized measurable units. Smart decision making in product driven manufacturing applications rely on specifications of production requirements. Manufacturing automation concepts are based on logic-based or negotiation based approaches. Data-driven automation has been less considered, making this as an opportunity for future research. Other applications rely fundamentally in the data acquisition and a number of sensors placed in shop-floor machines and resources. Two examples are maintenance and energy optimization. In one hand maintenance rely on acoustics, temperature, velocity, pressure and other variables, to understand health status of machines. On the other hand, energy optimization application rely mostly on the measurement of electrical variables i.e. mart meters, current and voltage clamps, single-phase and 3-phase smart plugs. With the advent of human-centre manufacturing applications, the acquisition of data from operators is coming a trend of current research, specially data that can be used to model human characteristics, such as behaviour and comfort. Wearable trackers can measure human performance under stressful or difficult conditions. Important considerations regarding data sources are privacy in the collection of data that in some cases should be can not even be used because of various regulations i.e General Data Protection Regulation.

The data collected from source can be accumulated through with either manual data acquisition or automatic data acquisition. The trade-off happens in form or consistency and reliability of the data. Data collection is majorly dependent on the type of data source and can come to sources such as evaluations, simulations, and predictions. Data collected can be structured or unstructured. This data collection is usually accompanied by an underlying framework that leverages step-wise process to gather desired data for decision-making. Certain applications like predictive maintenance, monitoring, energy consumption and event-driven automation require data to be collected as per specific requirements. These requirements can be real-time, time-driven, periodic or fulfilling any other application-specific criterion.

The data transmission can be sockets, OPC-UA, MQTT, TCP/IP (such as PLC simulator), or other communication protocols depending on the application domain and can be dynamically chosen. This communication method will be suggested as the middle-ware between the digital twin and the optimization environment. Moreover, it can also be used as the communication channel between different devices in the digital twin and the real devices. If the workstations in the manufacturing system use different operating systems, then OPC UA is a better solution. The transmission of data for further processing depends on the logging frequency of energy data. Usually, an ethernet connection is used for transmitting the data. However, high-frequency energy data is first stored in an external memory device of an energy monitoring solution. After some time, all the collected data is transmitted manually to the processing computer. Some monitoring solutions offer transmitting data via WiFi which is an advantage and disadvantage at the same time. Transmitting energy data via WiFi is transport flexibility and high transmission distance, but WiFi has shortcomings such as high latency and transmission unreliability. Hence, industrial standards such as Modbus and Profinet are used for mission-critical applications. The introduction of IoT on the shop floor has increased the transmission of low-frequency sensor information directly from the source through WiFi from various sources. This has also impacted the latency of the system's response. Industrial wireless networks include Industrial switches, Industrial routing, and Wireless AP.

As manufacturers becomes increasingly reliant on sensors and various data sources, data storage will become an increasingly important concern, especially the ability to storage large amount of data. There is a trend in the manufacturers from moving from traditional RDBMS database to NoSQL and NewSQL database considering the scalabilty. There is a need to develop techniques to not just store the data in a structured manner but also filter the redundant data and delete the data which is no longer relevant. This would greatly reduce the storage cost and complexity. There are very few work dealing with this aspect.

Data processing techniques has been widely used in manufacturing. With the development of Internet of Things (IoT), 5G, and cloud computing technologies, the amount of data from manufacturing systems has been increasing rapidly. With massive industrial data, achievements beyond expectations have been made in the product design, manufacturing, and maintain process. Data processing have been a core technology to empower intelligent manufacturing systems.

Finally, visualization is usually a neglected aspect of research. As demonstrated, multiple scholars prefer python libraries for easy static visualizations. However, for providing a proper commercial implementation of big data applications, visualization is as essential as other stages. The ability of applications to further exploit data from user behaviour to improve the visualization aspect in manufacturing is something that needs further research. Furthermore, there is a lack of standardization that requires researchers and engineers to identify generic abstractions for industrial data and understand different users groups to develop new frameworks for visualization applications.

Challenges

In this section, we compiled the challenges found in the literature. Although some of the challenges below are application-specific, they were found quite often in the reviewed literature.

- Data measurement solutions usually come with inherent measurement errors. Although these errors are relatively small, they affect transferability [60]. For instance, the same sensor for the same equipment performing the same application can yield different energy consumption values. These noisy and non-deterministic measurement values challenge data-processing and decision-making algorithms.

- The frequency of collected data [40] is another challenge in the literature. A sampling at a high rate produces much data that is difficult to transmit and process in real-time. However, certain applications require high-frequency data, such as energy parameter profiling applications. Therefore, careful tradeoff should be considered energy data from manufacturing in a shop floor

- To capture lifecycle data, discover knowledge and share it among all lifecycle stakeholders [21], a data acquisition system is needed that incorporates all information gathered during the production process.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

- Real-time processing, analysis, production reporting, and monitoring of data-driven sources must be in place for real-time analysis of sensor data[28], [28].
- Lack of reliable data and valuable knowledge that can be employed to support the optimized decision-making of product life cycle management.
- The heterogeneity of data is another challenge. For example, fault prediction considering multi-source heterogeneous data and complex processes [26, 29, 27]
- The heterogeneity requires novel data processing techniques. Utilizing traditional signal processing techniques considering the 5V challenges posed by industrial big data [26], fault prediction considering multi-source heterogeneous data and complex processes [26, 29, 27].
- The design for data visualization to improve human interaction is a complex task. [61] listed different aspects to consider for data visualization, especially *Visual and task complexity*, referring that complex infographics and large amounts of data unorganized or ungrouped, can cause distress to users. Also, the increase in the number of steps to realize a task can produce mistakes and reduce the operator's performance.
- Data issues are fundamental challenges to smart manufacturing, which extract actionable information from good quality of data. In order to prepare the suitable data for smart applications, amount of cost and time is consumed to address the data issues.
- Cybersecurity will continuously challenge manufacturing since security standards are still not available in some system [62].
- The security of big data analytics in manufacturing systems is another major concern in the application.
- Governance of big data handles data integrity, quality, provenance, retention, processing, and analysis in full data lifecycle [63]. The governance of industrial big data considers the issues of security and privacy [64].

This project has received funding from the European Union's
Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie
grant No. 814078

Chapter 6

Conclusion

In this research, a basis for the development of an homogeneous approach to gather and use data on shop-floor in manufacturing environments has been presented. A literature review of research regarding big data in manufacturing has been performed, targeting the complete data life cycle. In this regard, the needs, requirements and methods for the seven stages of the data life cycle of big data in manufacturing have been presented and discussed. Therefore, approaches for data acquisition, processing and utilisation for decision making in shop-floor in manufacturing have been established and challenges in each stage have been elaborated.

As results of this study, approaches have been identified in each stage of the big-data life cycle in manufacturing, focusing on maintenance, automation, quality, decision making, energy optimization, user interaction, and adaptability. Data sources, such as sensors, documents and models, have been identified and elaborated, detailing their usage and benefits, as well as possible drawbacks. Thereupon, data collection techniques have been presented, i.e. manual data acquisition and automatic data acquisition, describing the benefits and drawbacks of each. Furthermore, a separation between monitoring and predictive applications has been described, highlighting the effect that the intended application has in data collection. Having presented data collection techniques, data transmission protocols and techniques have been studied. Techniques and protocols for data transmission have been presented, as well as the cases in which each may be used. Following, data storage possibilities have been presented. Since data may be structured, semi-structured and unstructured, storage options have been discussed for each type of data structure, as well as the methods to integrate data in different formats and from different sources. In the context of data processing, several approaches towards data processing have been presented, as well as leading technologies for big data processing. In general, artificial intelligence and statistical approaches have been identified as the main contributors in this stage. Finally, data visualization methods, an integral part of data analysis, have been described in the context of smart manufacturing automation and big data. Several platforms and frameworks for data visualization have been reviewed and programming languages suitable for creating dashboards and visualization applications have been described.

Have been presented the results of the literature review, a discussion of the trends and insights from the review process has been presented. It has been identified that the primary data sources include smart sensors and IoT devices. Nevertheless, human-centered manufacturing applications have included data acquisition from operators, allowing modelling of behaviour and comfort. An important consideration that has been highlighted, regardless of the source of the data, is data privacy and restrictions that may apply due to regulations. Regarding data transmission, several protocols have been identified and their usage will depend on the technologies being used and the application. Data format, data size, transmission distance and transmission rates have a determining effect on which protocols to use and how to integrate the data being sent. Thereafter, in data storage, moving from traditional structured data storage, such as RDBMS, to unstructured and semi-structured data storage, such as NoSQL and NewSQL, has been identified as the leading trend. In addition, it has been identified that there is a lack of focus on irrelevant data filtering and deletion, which might help to reduce cost and processing power in applications where there are economical or storage constraints. In general, this research has identified several challenges in literature. Challenges involve possible errors in the collected data, which may lead to inaccurate measurements, as well as the challenges regarding the handling of varied sampling frequencies and the impact on the transmission technologies used. Furthermore, challenges regarding heterogeneity of data have been identified, where the integration of varied data sources could represent a challenge during data storage, processing, and visualization, deriving in incorrect analysis of data or complexity in understanding the data obtained during the data life cycle. Finally, cybersecurity has been identified as an important challenge, as several studies have lacked attention in this regard.

Having reviewed and discussed the state of the art of big-data life cycle in shop-floor in manufacturing, a consolidated framework and methodology for the big-data life cycle, based on the findings of this review, is to be presented in the upcoming sections of this work package, as well as test cases for validation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

Bibliography

[1] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Journal of Manufacturing Systems*, vol. 48, pp. 157–169, 2018.

[2] B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, and T. Freiheit, "Smart manufacturing and intelligent manufacturing: A comparative review," *Engineering*, vol. 7, no. 6, pp. 738–757, 2021.

[3] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. Do Noh, "Smart manufacturing: Past research, present findings, and future directions," *International journal of precision engineering and manufacturing-green technology*, vol. 3, no. 1, pp. 111–128, 2016.

[4] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, 2018.

[5] B. Furht and F. Villanustre, "Introduction to big data," in *Big data technologies and applications.* Springer, 2016, pp. 3–11.

[6] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, p. 114060, 2021.

[7] Y. Cui, S. Kara, and K. C. Chan, "Manufacturing big data ecosystem: A systematic literature review," *Robotics and computer-integrated Manufacturing*, vol. 62, p. 101861, 2020.

[8] H. Yang, S. Kumara, S. T. Bukkapatnam, and F. Tsung, "The internet of things for smart manufacturing: A review," *IISE Transactions*, vol. 51, no. 11, pp. 1190–1216, 2019.

[9] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *2013 International conference on collaboration technologies and systems (CTS).* IEEE, 2013, pp. 48–55.

[10] J. Lee, E. Lapira, B. Bagheri, and H.-a. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manufacturing letters*, vol. 1, no. 1, pp. 38–41, 2013.

[11] M. Milivojević, I. Antolović, and D. Rančić, "Using collada and x3d for webgl based 3d data visualization."

[12] M. Yao, Z. Shao, and Y. Zhao, "Review on Energy Consumption Optimization Methods of Typical Discrete Manufacturing Equipment," in *Intelligent Robotics and Applications*, X.-J. Liu, Z. Nie, J. Yu, F. Xie, and R. Song, Eds. Cham: Springer International Publishing, 2021, pp. 48–58.

[13] Paryanto, M. Brossog, M. Bornschlegl, and J. Franke, "Reducing the energy consumption of industrial robots in manufacturing systems," *The International Journal of Advanced Manufacturing Technology*, vol. 78, no. 5, pp. 1315–1328, May 2015. [Online]. Available: https://doi.org/10.1007/s00170-014-6737-z

[14] D. Meike, M. Pellicciari, and G. Berselli, "Energy Efficient Use of Multirobot Production Lines in the Automotive Industry: Detailed System Modeling and Optimization," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 3, pp. 798–809, Jul. 2014, conference Name: IEEE Transactions on Automation Science and Engineering.

[15] K. Ding, Y. Zhang, F. T. Chan, C. Zhang, J. Lv, Q. Liu, J. Leng, and H. Fu, "A cyber-physical production monitoring service system for energy-aware collaborative production monitoring in a smart shop floor," *Journal of Cleaner Production*, vol. 297, p. 126599, 2021.

[16] R. J. Machchhar, C. N. K. Toller, A. Bertoni, and M. Bertoni, "Data-driven value creation in smart product-service system design: State-of-the-art and research directions," *Computers in Industry*, vol. 137, p. 103606, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016636152200001X

[17] I. Zolotová, P. Papcun, E. Kajáti, M. Miškuf, and J. Mocnej, "Smart and cognitive solutions for operator 4.0: Laboratory h-cpps case studies," *Computers & Industrial Engineering*, vol. 139, p. 105471, 2020.

[18] S. Wang, J. Wan, M. Imran, D. Li, and C. Zhang, "Cloud-based smart manufacturing for personalized candy packing application," *The Journal of Supercomputing*, vol. 74, no. 9, pp. 4339–4357, 2018.

[19] Y.-C. Lin, M.-H. Hung, H.-C. Huang, C.-C. Chen, H.-C. Yang, Y.-S. Hsieh, and F.-T. Cheng, "Development of advanced manufacturing cloud of things (amcot)—a smart manufacturing platform," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1809–1816, 2017.

[20] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Dai, X. Chen, and T. Zhang, "A big data approach for logistics trajectory discovery from rfid-enabled production data," *International Journal of Production Economics*, vol. 165, pp. 260–272, 2015.

[21] Y. Zhang, S. Ren, Y. Liu, and S. Si, "A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products," *Journal of cleaner production*, vol. 142, pp. 626–641, 2017.

[22] Y. Zhang, S. Ren, Y. Liu, T. Sakao, and D. Huisingh, "A framework for big data driven product lifecycle management," *Journal of Cleaner Production*, vol. 159, pp. 229–240, 2017.

[23] M. Kohlert and A. König, "Large, high-dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization in polymer film industry," *Neural Computing and Applications*, vol. 26, no. 3, pp. 581–588, 2015.

[24] H. U. Rehman, T. Pulikottil, L. A. Estrada-Jimenez, F. Mo, J. C. Chaplin, J. Barata, and S. Ratchev, "Cloud based decision making for multi-agent production systems," in *EPIA Conference on Artificial Intelligence*. Springer, 2021, pp. 673–686.

[25] J. Lee, H. D. Ardakani, S. Yang, and B. Bagheri, "Industrial big data analytics and cyber-physical systems for future maintenance & service innovation," *Procedia cirp*, vol. 38, pp. 3–7, 2015.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

[26] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," *IEEE Access*, vol. 5, pp. 23 484–23 491, 2017.

[27] G. Martínez-Arellano, G. Terrazas, and S. Ratchev, "Tool wear classification using time series imaging and deep learning," *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 9, pp. 3647–3662, 2019.

[28] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2039–2047, 2017.

[29] K. Nakata, R. Orihara, Y. Mizuoka, and K. Takagi, "A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 4, pp. 339–344, 2017.

[30] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," *arXiv preprint arXiv:1910.11470*, 2019.

[31] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-t. Yih, "Cross-sentence n-ary relation extraction with graph lstms," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 101–115, 2017.

[32] L. Sun, "Research on product attribute extraction and classification method for online review," in *2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*. IEEE, 2017, pp. 117–121.

[33] S. Durga, R. Nag, and E. Daniel, "Survey on machine learning and deep learning algorithms used in internet of things (iot) healthcare," in *2019 3rd international conference on computing methodologies and communication (ICCMC)*. IEEE, 2019, pp. 1018–1022.

[34] C. Reuter and F. Brambring, "Improving data consistency in production control," *Procedia CIRP*, vol. 41, pp. 51–56, 2016.

[35] N. Robertson and T. Perera, "Automated data collection for simulation?" *Simulation Practice and Theory*, vol. 9, no. 6-8, pp. 349–364, 2002.

[36] P. Azad, N. J. Navimipour, A. M. Rahmani, and A. Sharifi, "The role of structured and unstructured data managing mechanisms in the internet of things," *Cluster computing*, vol. 23, no. 2, pp. 1185–1198, 2020.

[37] K. Duh, G. Neubig, K. Sudoh, and H. Tsukada, "Adaptation data selection using neural language models: Experiments in machine translation," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2013, pp. 678–683.

[38] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren, "A survey of visualization for smart manufacturing," *Journal of Visualization*, vol. 22, no. 2, pp. 419–435, 2019.

[39] Y. Lu and X. Xu, "Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 92–102, 2019.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

[40] M. Gadaleta, G. Berselli, M. Pellicciari, and F. Grassia, "Extensive experimental investigation for the optimization of the energy consumption of a high payload industrial robot with open research dataset," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102046, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S073658452030257X

[41] S.-J. Shin, "An opc ua-compliant interface of data analytics models for interoperable manufacturing intelligence," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3588–3598, 2020.

[42] M. Gadaleta, M. Pellicciari, and G. Berselli, "Optimization of the energy consumption of industrial robots for automatic code generation," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 452–464, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736584518301856

[43] R. Bonnard, K. M. M. Vieira, S. ARANTES, R. Lorbieski, C. NUNES, and A. P. Mattei, "A big data/analytics platform for industry 4.0 implementation in smes," *CIGI QUALITA*, 2019.

[44] D. Fernandes and J. Bernardino, "Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb." in *Data*, 2018, pp. 373–380.

[45] C.-A. Chen, M. Won, R. Stoleru, and G. G. Xie, "Energy-efficient fault-tolerant data storage and processing in mobile cloud," *IEEE Transactions on cloud computing*, vol. 3, no. 1, pp. 28–41, 2014.

[46] M. Zhang, Y. Zuo, and F. Tao, "Equipment energy consumption management in digital twin shop-floor: A framework and potential applications," in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE, 2018, pp. 1–5.

[47] M. Zhang and J. Yan, "A data-driven method for optimizing the energy consumption of industrial robots," *Journal of Cleaner Production*, vol. 285, p. 124862, Feb. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652620349064

[48] J. Yan and M. Zhang, "A transfer-learning based energy consumption modeling method for industrial robots," *Journal of Cleaner Production*, vol. 325, p. 129299, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652621034843

[49] A. Liu, H. Liu, B. Yao, W. Xu, and M. Yang, "Energy consumption modeling of industrial robot based on simulated power data and parameter identification," *Advances in Mechanical Engineering*, vol. 10, no. 5, p. 1687814018773852, May 2018, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/1687814018773852

[50] L. Guo, F. Yan, T. Li, T. Yang, and Y. Lu, "An automatic method for constructing machining process knowledge base from knowledge graph," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102222, 2022.

[51] Y. Zhang, X. Luo, H. Zhang, and J. W. Sutherland, "A knowledge representation for unit manufacturing processes," *The International Journal of Advanced Manufacturing Technology*, vol. 73, no. 5, pp. 1011–1031, 2014.

[52] B. Zhou, J. Bao, J. Li, Y. Lu, T. Liu, and Q. Zhang, "A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops," *Robotics and Computer-Integrated Manufacturing*, vol. 71, p. 102160, 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078

[53] H. L. Nguyen, D. T. Vu, and J. J. Jung, "Knowledge graph fusion for smart systems: A survey," *Information Fusion*, vol. 61, pp. 56–70, 2020.

[54] M. Saez, S. Lengieza, F. Maturana, K. Barton, and D. Tilbury, "A data transformation adapter for smart manufacturing systems with edge and cloud computing capabilities," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE, 2018, pp. 0519–0524.

[55] M. Mani and S. Fei, "Effective big data visualization," in *Proceedings of the 21st International Database Engineering & Applications Symposium*, 2017, pp. 298–303.

[56] G. W. Young and R. Kitchin, "Creating design guidelines for building city dashboards from a user's perspectives," *International Journal of Human-Computer Studies*, vol. 140, p. 102429, 2020.

[57] S. Vijay, B. Kuraichen *et al.*, "Data driven prognostics of milling tool wear: A machine learning approach," in *2021 International Conference on Computational Performance Evaluation (ComPE)*. IEEE, 2021, pp. 002–007.

[58] S. Kahveci, B. Alkan, A. Mus'ab H, B. Ahmad, and R. Harrison, "An end-to-end big data analytics platform for iot-enabled smart factories: A case study of battery module assembly system for electric vehicles," *Journal of Manufacturing Systems*, vol. 63, pp. 214–223, 2022.

[59] P. Xu, H. Mei, L. Ren, and W. Chen, "Vidx: Visual diagnostics of assembly line performance in smart factories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 291–300, 2017.

[60] L. Li, H. Huang, F. Zhao, M. Gao, Z. Liu, and J. W. Sutherland, "Power measurement in energy efficient manufacturing: Accuracy analysis, challenges, and perspectives for improvement," *Procedia CIRP*, vol. 80, pp. 138–143, 2019.

[61] N. Kumar and S. Prajapati, "Challenges for interface designers in designing sensor dashboards in the context of industry 4.0," *Open Ind. Manuf. Eng. J*, vol. 13, pp. 539–542, 2019.

[62] Y. Lu, K. C. Morris, S. Frechette *et al.*, "Current standards landscape for smart manufacturing systems," *National Institute of Standards and Technology, NISTIR*, vol. 8107, p. 39, 2016.

[63] P. P. Tallon, "Corporate governance of big data: Perspectives on value, risk, and cost," *Computer*, vol. 46, no. 6, pp. 32–38, 2013.

[64] M. Usman, A. Jolfaei, and M. A. Jan, "Rasec: an intelligent framework for reliable and secure multilevel edge computing in industrial environments," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 4543–4551, 2020.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No. 814078