

Entry

Knowledge Integration in Smart Factories

Johannes Zenkert , Christian Weber , Mareike Dornhöfer , Hasan Abu-Rasheed  and Madjid Fathi 

Department of Electrical Engineering and Computer Science, Institute of Knowledge Based Systems and Knowledge Management, University of Siegen, 57076 Siegen, Germany; christian.weber@uni-siegen.de (C.W.); m.dornhoefer@uni-siegen.de (M.D.); hasan.abu.rasheed@uni-siegen.de (H.A.-R.); fathi@informatik.uni-siegen.de (M.F.)

* Correspondence: johannes.zenkert@uni-siegen.de

Definition: Knowledge integration is well explained by the human–organization–technology (HOT) approach known from knowledge management. This approach contains the horizontal and vertical interaction and communication between employees, human-to-machine, but also machine-to-machine. Different organizational structures and processes are supported with the help of appropriate technologies and suitable data processing and integration techniques. In a Smart Factory, manufacturing systems act largely autonomously on the basis of continuously collected data. The technical design concerns the networking of machines, their connectivity and the interaction between human and machine as well as machine-to-machine. Within a Smart Factory, machines can be considered as intelligent manufacturing systems. Such manufacturing systems can autonomously adapt to events through the ability to intelligently analyze data and act as adaptive manufacturing systems that consider changes in production, the supply chain and customer requirements. Inter-connected physical devices, sensors, actuators, and controllers form the building block of the Smart Factory, which is called the Internet of Things (IoT). IoT uses different data processing solutions, such as cloud computing, fog computing, or edge computing, to fuse and process data. This is accomplished in an integrated and cross-device manner.

Keywords: smart factory; cloud computing; fog computing; edge computing; knowledge integration; knowledge management; data analytics; text analytics; knowledge graph



Citation: Zenkert, J.; Weber, C.; Dornhöfer, M.; Abu-Rasheed, H.; Fathi, M. Knowledge Integration in Smart Factories. *Encyclopedia* **2021**, *1*, 792–811. <https://doi.org/10.3390/encyclopedia1030061>

Academic Editors: Raffaele Barretta, Ramesh Agarwal, Krzysztof Kamil Żur and Giuseppe Ruta

Received: 15 July 2021

Accepted: 12 August 2021

Published: 16 August 2021

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



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

1. Introduction

In the wake of the Industry 4.0 development, the concept of Smart Factories and related technologies such as Cyber–Physical Systems (CPS) or the application of Internet of Things (IoT) in an industrial context emerged in the span of just ten years. Cyber–Physical Systems combine the analogue or physical production world with the digital world in a newfound complexity. Consequently, data and knowledge are playing an increasingly bigger role, supporting and leading to data-driven manufacturing (e.g., [1,2]).

Industry 4.0 has first been published on a larger scale as a (marketing) concept in 2011 at the Hannover fair in Germany. What followed was the backwards view on how to define the previous epochs of Industry 1.0 to 3.0 and their respective historical focus (e.g., [3]). Industry 4.0 presents a forward view of how the concept may be used to transform the current production environment and integrate digital solutions to improve aspects such as performance, maintenance, manufacturing of individualized products or to generate transparency over the whole production process or value chain of a company. Zhong et al. (2017) conclude that “*Industry 4.0 combines embedded production system technologies with intelligent production processes to pave the way for a new technological age that will fundamentally transform industry value chains, production value chains, and business models*” [4]. The technological advance also requires interaction with and integration of skilled workforces, even though this is often not addressed [5]. In this light, Industry 4.0 can be further defined as a network of humans and machines, covering the whole value chain, while supporting digitization

and fostering real-time analysis of data to make the manufacturing processes more transparent and simultaneously more efficient to tailor intelligent products and services to the customer [6]. Depending on the type of realization and number of data sources, there might be a requirement for Big Data analysis [1,2].

Intensive research has been conducted on how to make the existing factories “smarter”. In this context, the term “smart” refers to making manufacturing processes more autonomous, self-configured and data-driven. Such capabilities enable, for example, gathering and utilizing of knowledge about machine failures, to enable predictive maintenance actions or ad hoc process adaptations. In addition, the products and services which are manufactured often are aimed to be “smart” too, meaning they contain the means to gather data which may be used to improve functionalities or services through continuous data feedback to the manufacturer.

A generic definition of the term Smart Factory is still difficult, as many authors provide definitions based on their specific research area [7]. It can be concluded from this that the Smart Factory concept is targeting a multi-dimensional transformation of the manufacturing sector that is still continuing. Based on their analysis, Shi et al. (2020) conclude on four main features of a Smart Factory: (1) sensors for data gathering and environment detection with the goal of analysis and self-organization, (2) interconnectivity, interoperability and real-time control leading to flexibility, (3) application of artificial intelligence (AI) technologies such as robots, analysis algorithms as well as (4) virtual reality to enhance “*human-machine integration*” [7]. To target the diversity of topics related to the term Smart Factory, Strozzi et al. (2017) conducted a literature survey of publications between 2007 and 2016 and concluded from more than 400 publications direct relations between smart factories and the topics of real-time processing, wireless communication, (multi)-agent solutions, RFID, intelligent, smart, flexible and real-time manufacturing, ontologies, cloud computing, sustainability and optimization [8], identifying main areas but also enablers for a Smart Factory.

The overall question this work tries to answer is “How do Industry 4.0 environments or Smart Factory plants of the future look like and what role does data and knowledge play in this development?” Tao et al., (2019), referencing Zhong et al. (2017) [4], summarize that “*Manufacturing is shifting from knowledge-based intelligent manufacturing to data-driven and knowledge-enabled smart manufacturing, in which the term “smart” refers to the creation and use of data*” [9]. This shift has to be considered with the help of concepts known from the disciplines of data analytics, knowledge management (KM) and knowledge integration, machine learning and artificial intelligence. It presents a change from “knowledge-based”, explicitly represented, qualitative data to the consideration of quantitative data in which meaningful patterns trigger manufacturing decisions, while being informed by supporting knowledge representations, such as ontologies. Especially the pronounced roles of data and knowledge are key aspects of future manufacturing environments and products.

Before talking more about this change the terms data, information and knowledge as well as knowledge management will be introduced briefly, giving a better background of understanding [10]: From a knowledge management perspective, the three terms are closely related, whereas data are the basis being formed out of a specific alphabet and grammar/syntax and may be structured, semi-structured or unstructured. Information builds on top of data which are used and interpreted in a certain (semantical) context, while knowledge is interconnected, applied or integrated information and oftentimes relates to a specific application area or an individual. That is why the terms of individual and collective knowledge are important factors for knowledge management, a discipline supporting, e.g., the acquisition, development, distribution, application or storage of knowledge within an organization. Different KM models or processes may be established and manage targeting human, organizational and technological aspects. Frey-Luxemberger gives an overview of the KM field [10].

The changes and role of data in manufacturing detailed above, is motivated or required by the rising customer demands of customized or tailored orders [11]. From an outside

perspective the change of market demands requires hybrid solutions which not only focus on the manufacturing of physical devices or products but an accompanying (smart) service [12], which is only possible if the product generates data to be analyzed and used for offering said service. At the same time, the interconnected technologies require a change in knowledge management. Bettiol et al. (2020) conclude: “On the one hand advanced, interconnected technologies generate new knowledge autonomously, but on the other hand, in order to really deploy the value connected to data produced by such technologies, firm should also rely on the social dimension of knowledge management dynamics” [13]. The social dimension will be discussed later when reflecting on the changing role of employees in Smart Factories.

To meet a lot size of one, while offering extensive automated configuration abilities throughout the production process, the Smart Factory has to offer configuration and adaptation possibilities in a scalable way. At the same time these have to be manageable by the human workers, as well as being aligned to the underlying business processes. The realization is only possible by collecting and using data and knowledge throughout the manufacturing and documentation process, as well as by deploying automated data analytics and visualization tools to enable real time management and reconfiguration. It is expected that in the future workers inside Smart Factories will have to fulfill different roles or tasks in different processes or together with (intelligent) machines (e.g., [14,15]). Furthermore, instead of only administering one isolated machine, they will be supporting overarching tasks as the surveying and monitoring of interconnected production machines or plants, as well as flexible automation solutions. This again requires knowledge about inter-dependencies in the production process as well as about consequences for multiple production queues, e.g., in case of a failure of an intermediate machine. In this context the topic of predictive maintenance (e.g., [16]) is another major issue, as the gathered data inside the Smart Factory can and has to be used to minimize the times of failures in the more complex manufacturing environment, deploying analytics strategies (e.g., [16]) or machine learning algorithms to detect potential failures or maintenance measures. The concept of a digital twin (DT) (e.g., [9]) might be used here to fuse data and simulation models to create a real-time digital simulation and forecast the real environment, supporting the early detection of potential problems and real-time reconfiguration.

In the following, the different aspects of a Smart Factory including computing, analytics and knowledge integration perspectives will be discussed in more detail.

2. Smart Factory

2.1. Smart Factory Environment

Before the different levels of integration and computing in *Smart Factories* are explicitly addressed, a conceptual overview of a Smart Factory environment is given and explained. Figure 1 summarizes the integration levels and technologies related to a Smart Factory. The icons indicate the interconnectivity and communication between them and are detailed in the following sub-sections. In this setting, the horizontal perspective is focused on the shop floor level of the manufacturing facilities where products are manufactured, and related data are collected for later integration and analysis. The vertical perspective is a perspective of knowledge integration where gathered data from the manufacturing environment are fused, aggregated and integrated with the knowledge about the underlying business processes. The goal is to derive at the lowest level a real-time perspective on the state of manufacturing, while enabling at the top level a predictive business perspective. These alignments are reflected both horizontally and vertically in the RAMI 4.0 architecture of the i40-platform [17] and vertically in the 5C model [18].

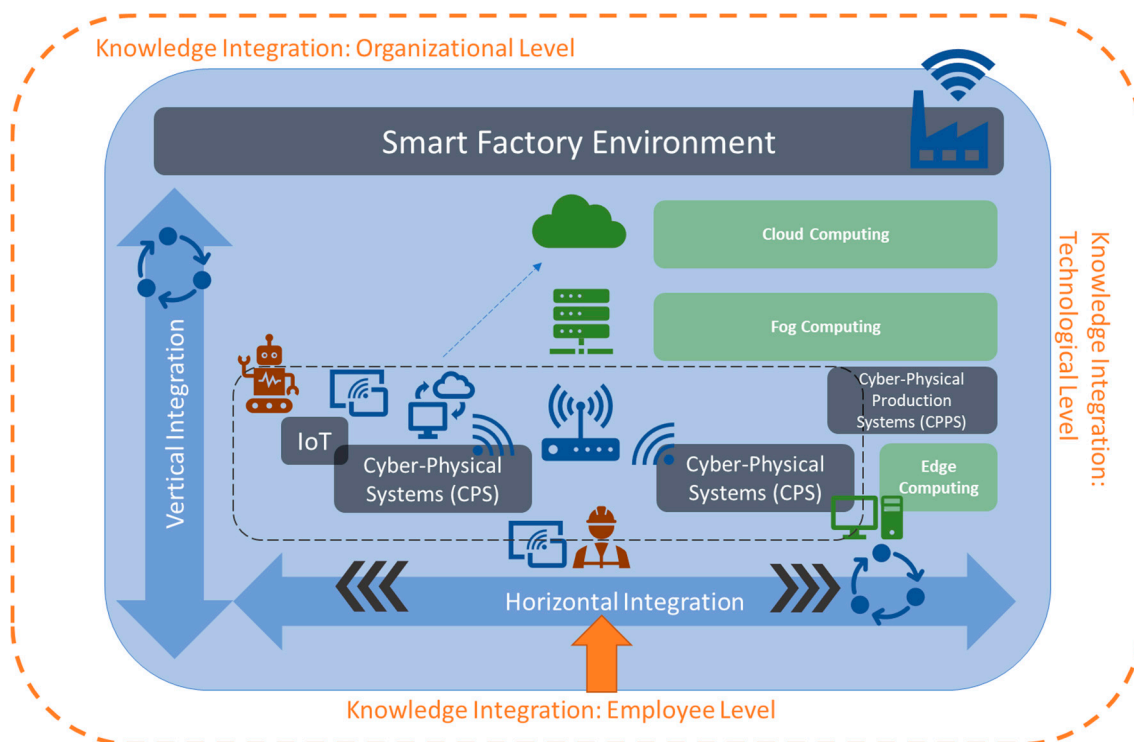


Figure 1. Smart Factory Environment and Knowledge Integration.

Depending on the author, either one or both terms of Cyber-Physical Systems (CPS) or Cyber-Physical Production Systems (CPPS) are used as the main building blocks of a Smart Factory. While the term CPS is sometimes used synonymously with CPPS, the term CPPS is also used to refer to a higher-level system that consists of multiple singular CPS [2,6]. In this paper, we follow the second interpretation. We use the term CPPS to refer to a wider scope, in which interconnected CPSs build the CPPS. In a practical scenario, these may be different production machines connected to one production process, where each machine resembles a unit as an integration of the “cyber”, meaning the computation and networking component, as well as the “physical” electrical and mechanical components. Karnouskos et al. (2019) named autonomy, integrability and convertibility as the main CPPS pillars [19].

Next to CPS or CPPS, other main building blocks are IoT devices, which may be any devices that are able to gather data through embedded sensors as an interconnected entity within a wider network, and subsequently integrating the collected data into the Smart Factory network. This way, previously closed manufacturing environments or passive objects may be transformed into an active role inside the network and, as such, the manufacturing or monitoring processes.

Different communication protocols, technologies and standards may be used for the identification of objects, realization of connectivity or transmission of data, e.g., RFID, WLAN, GPS, OPC UA or specific industrial communication protocols (e.g., [6]). The identification of individual objects inside the production process is essential for individualization of products and automation of processes [20]. Soic et al. (2020) reflect in their work about context-awareness in smart environments that it requires the interconnection of “*physical and virtual nodes*”, whereas the nodes relate to “*an arbitrary number of sensors and actuators or external systems*” [21]. Fei et al. (2019) conclude that data gathered from “*interconnected twin cybernetics digital system[s]*” supports prediction and decision making [22]. Gorodetsky et al. (2019) view “*digital twin-based CPPS*” as a basis for “*real-time data-driven operation control*” [23].

IIoT architecture models such as the Industrial Internet of Things (IIoT) architecture published by the Industrial Internet Consortium (IIC) [24] or the RAMI 4.0 architecture of

the i40-platform [17] are approaches to standardize the implementation of a Smart Factory from the physical layer up to the application layer similar to the OSI 7 layer model.

Each CPS (e.g., through embedded sensors), or IoT device, produces or gathers data which need to be processed and analyzed to generate immediate actions or further derive knowledge about the status of the devices or the manufacturing process in focus. This may happen directly inside the device or with the help of a local computing device attached to the machine, called edge component and *Edge Computing*, respectively, on a local network level, called *Fog Computing*, or with the help of a central *Cloud Computing* platform. The Cloud Computing solution is beneficial when there are different production plants and the data need to be gathered at one common, yet transparently distributed, place to be analyzed. Edge or Fog Computing are better in cases of immediate processing or processing inside a local plant. In this context, it is important to define if data need to be analyzed immediately, e.g., for monitoring purposes or for cases where an analysis of a specific time frame is necessary. In the first case, *Stream Processing* or *Stream Analytics* can be used where the data are gathered and immediately analyzed “in place”, meaning while being streamed, while in the second case *Batch Processing* or *Batch analysis* takes place, where data are collected and processed together with the option of aggregating over time and features. In Section 3, the different aspects are discussed in more detail.

2.2. Multi-Dimensional Knowledge Integration in Smart Factories

Theorem 1. *From the perspective of Knowledge Management and considering the three associated main pillars of (1) human workforce, (2) organizational structures or processes as well as (3) technology, the establishment of a Smart Factory requires a Multi-dimensional Knowledge Integration Perspective targeting the aforementioned three pillars, while considering the knowledge associated with, as well as the interconnected manufacturing processes and automation requirements to support a “Smart” solution.*

Based on Theorem 1, the following sub-sections detail the different aspects of how to create such a multi-dimensional knowledge integration perspective. We consider the human–organization–technology (HOT) (e.g., [10], or [25]) perspectives the target for this multi-dimensional knowledge integration [26]. Examples of the human (employee) perspective are roles like involved knowledge experts, lifelong learning, human–machine interaction, etc. Examples on the organizational perspective are the transformation of a hierarchical into a network-based structure, organizational learning and a data-driven business process integration; and finally, from the technological perspective we have the variance of knowledge assets (e.g., textual content, sensor data) as well as analysis processes, which will be detailed in Section 3.3. Figure 1 summarizes the perspectives around the Smart Factory environment.

3. Knowledge Integration

3.1. Knowledge Integration on Organizational Level: Horizontal and Vertical Integration

The Smart Factory integration or transformation from a regular factory into a Smart Factory bases on a horizontal and a vertical integration perspective. The horizontal perspective targets the production and transportation processes along the value chain, while the vertical perspective has a look at how the “new” or “transformed” production environment fits and interrelates with the other organizational areas such as production planning, logistics, sales or marketing, etc. The platform i4.0 visualized this change between Industry 3.0 [27] towards Industry 4.0 [28] with the change from a production pyramid towards a production network. Hierarchies are no longer as important in Industry 4.0 as the concept transforms the organization into an interconnected network, where hierarchical or departmental boundaries are resolved or less significant. The important aspect is the value chain and all departments working towards customer-oriented goals. This of course requires a change in the way the employees work and interact with each other. While

before employees in the same hierarchy level or working on the same subject were mostly communicating with each other, this organizational change also leads to a social change inside the company. In addition, job profiles are changing towards a “knowledge worker”. This will be discussed in Section 3.3.2 as it requires knowledge management aspects to support this change process.

Going back to the horizontal and vertical integration, both aspects are not only related to the organizational changes necessary to establish a Smart Factory. They also target the data perspective. Data play a central role for the Smart Factory, as it is needed to automate processes or exchange information between different manufacturing machines or between divisions as production and logistics, etc. These data are gathered mostly with the help of sensors, embedded into IoT devices which are part of the CPS. A sensor might be attached directly at a specific manufacturing machine or at different gates where the products or pre-products come along during their way through the manufacturing environment. RFID tags may be used to automatically scan a product, check its identity and update the status or next manufacturing step. If the data are gathered and analyzed along the manufacturing process, then they are part of the horizontal integration, as the results of analysis might directly influence the next production steps and focus on a real-time intervention. Data analytics [16] from a vertical perspective have a broader focus and gather and integrate data from and at different hierarchical levels and different IT systems, as, e.g., the enterprise resources planning (ERP) system, as well as over a longer time period to, e.g., generate reports about the Smart Factory or a specific production development for a specified time.

3.2. Knowledge Integration: Employee Level

If we consider the organizational changes leading to a Smart Factory, it is important to consider the role of the current and future employees in this environment. This applies especially to those employees whose production environments have been analogue or not interconnected before and who now need to be part of the digitized production process. As such, an essential factor in this transformation process is lifelong learning and needs to be considered in all knowledge management activities of the Smart Factory [29].

From the perspective of knowledge integration, it is recommended to involve knowledge engineers or managers to support the transformation process but to also consider the concerns or problems the employees have and to acquire and provide the training they need. Another aspect is to establish the knowledge exchange between different involved engineering disciplines, as well as computer science for the design and understanding of complex systems such as CPS.

The employees in Smart Factories may be concerned about their future role and tasks as there will be continuous shifts in human–machine interaction (HMI), where the mechanized counterparts of a human worker in the future may be different CPS, a robot or an AI application. The roles that workers execute may be the controller of a machine, peer or teammate up to their replacement by an intelligent machine [30]. Ansari, Erol and Shin (2018) differentiate HMI into human–machine cooperation and human–machine collaboration in the Smart Factory environment: “A human labor on the one side is assisted by smart devices and machines (human–machine cooperation) and on the other should interact and exchange information with intelligent machines (human–machine collaboration)” [14]. This differentiation indicates that the human worker has benefits which will make their work process easier due to assisted technical and smart solutions, but also pose challenges as the worker needs to learn to work together with this new technical and digitized work environment. Vogel-Heuser et al. (2017) highlight that the human worker is now interconnected with CPS with the help of multi-modal human–machine interfaces [29]. The automation and application of AI are factors which question (a) the role of the human worker in the work process as well as (b) their abilities in decision making as it might be influenced by or contrary to the recommendations or actions of the AI application. An AI “would always base its decision-making on optimizing its objectives, rather than incorporating social or emotional factors” [15], posing a challenge of who is the main decision maker in the Smart

Factory and if the explicit and implicit knowledge and experience of the human worker are more valuable than the programmed AI logic, e.g., based on data and the execution of machine learning algorithms. Seeber et al. (2019) resume that “*The optimal conditions for humans and truly intelligent AI to coexist and work together have not yet been adequately analyzed*”, leading to future challenges and requirements to create recommendations or standards for said coexistence and collaboration between human and machine teammates [15]. One approach is a “mutual learning” between human and machine, supported by *human acquisition, machine acquisition, human participation and machine participation* leading to the execution of shared tasks between human and machine [14]. North, Maier and Haas (2018) envision that “*in the future expertise will be defined as human (expert) plus intelligent machine*”, with the challenge being how they learn and work together [30]. A possible system, showcasing this synergetic collaboration is the implementation of cobots, passive robots that are tailored for collaboratively inhabiting a shared space for the purpose of operating processes together with humans [31].

3.3. Knowledge Integration: Technological Level

In the introduction, the central role of data and knowledge in modern manufacturing has been motivated. Tao et al. are even speaking of “data-driven smart manufacturing” [1]. This means a continuous generation of data streams, leading to big data which require processing and analysis [2]. The following Table 1 summarizes how these data may be used or processed to generate knowledge inside the Smart Factory.

Table 1. Overview of Knowledge Integration on a Technological Level.

Knowledge Processing	Realization in Smart Factory	Benefits	Exemplary Technologies ¹
Data Computing/ Processing	Cloud Computing	Software as a Service (SaaS), central software applications without local installation	Amazon AWS, Microsoft Azure
	Fog Computing	Enhance low-latency, mobility, network bandwidth, security and privacy	Cisco IOx
	Edge Computing	Offload network bandwidth, shorter delay time (latency)	Cisco IOx, Intel IOT solutions, Nvidia EGX
Data Analytics	Descriptive Analytics	Status and usage monitoring, reporting, anomaly detection and diagnosis, modeling, or training	RapidMiner, RStudio Server, Tableau
		Stream Analytics: Real-Time Analysis	Apache Kafka/Flink, Elasticsearch and Kibana
		Batch Analytics: Monitoring/Reporting	Apache Spark/Zeppelin, Cassandra, Tensorflow, Keras
	Predictive Analytics	Predicting capacity needs and utilization, material and energy consumption, predicting component and system wear and failures	RapidMiner, RStudio Server, Tensorflow, Keras, AutoKeras, Google AutoML
	Prescriptive Analytics	Guidance to recommend operational changes to optimize processes, avoid failures and associated downtime	RapidMiner, RStudio Server, Tensorflow, Keras, AutoKeras, Google AutoML
Simulation	Digital Twin Concept	Real-time analysis, simulation of scalable products and product changes, wear and tear projection	MATLAB Simulink, Azure Digital Twins, Ansys Twin Builder

Table 1. Cont.

Knowledge Processing	Realization in Smart Factory	Benefits	Exemplary Technologies ¹
Semantic Knowledge Representation	Knowledge Graphs	Contextualization of multi-source data, Semantic relational learning	Neo4j, Grakn, Ontotext GraphDB, Eccenca corporate memory, Protégé
Text Mining	Intelligent Documents	Integration of lessons learned from reporting and failure logs in new design or production cycles	tm (R), nltk (Python), RapidMiner, Text Analytics Toolbox (MATLAB), Apache OpenNLP, Stanford CoreNLP

¹ Technologies and products are listed here exemplary and do not represent an exhaustive market overview.

Based on the table above, the mentioned technologies may be brought into context from a data to a knowledge integration perspective. The interaction and integration of these technologies are shown in Figure 2. It is assumed that the processing from data to information to knowledge, an integral concept of knowledge management, is applied here as well. Data might be in the form of structured streaming data, mostly generated from sensors or in logging units, or unstructured in form, such as texts, process documentation or reports created by the experts inside the Smart Factory environment. The structured data are mostly processed as incoming data streams which are stored in a database (locally or inside a cloud) and may either be analyzed in real-time or at a later stage in a batch. Machine learning methods might be applied for the task of learning and reasoning, especially if the application requires tasks such as predictive maintenance, decision support, or providing recommendations. Unstructured data can be utilized by pre-processing with the help of text mining methods such as entity and relationship recognition or semantic enrichments before being composed in a knowledge graph. A knowledge graph allows an enrichment with results from the analyzed structured data stream as well. All results can contribute to different applications in the context of smart manufacturing, such as monitoring, digital twin simulations or decision support.

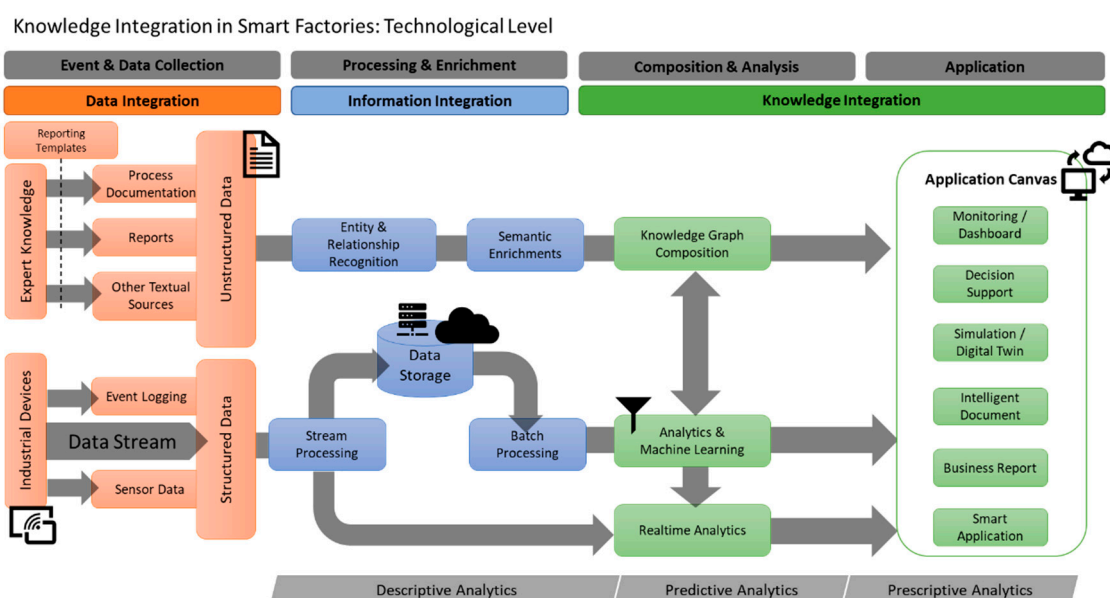


Figure 2. Smart Factory Environment—Technological Level.

In the following sub-sections, the technological level will be discussed in more detail.

3.3.1. Data Computing/Processing in Smart Factories

Smart Factories apply different kinds of computing levels to meet requirements regarding, e.g., real-time processing or big data analysis. The deciding factor for how to gather and process data streams inside the Smart Factory is oftentimes the available time or the amount of data. As a rule of thumb, the faster an analysis should be executed, the closer the necessary computing unit needs to be to the production machine. On the other hand, large amounts of data oftentimes need a central storage to be fused and aggregated before they are processed. Therefore, Smart Factories consider three computing tiers or levels: Cloud, Fog and Edge Computing, e.g., [32]. In the case of a Smart Factory, the application of these computing approaches means that the previously closed environments, using only operation technology (OT), are now opened towards information technology (IT) thus providing opportunities regarding computing power but also challenges, e.g., from a security perspective.

Cloud Computing: Cloud Computing covers a computer science and hardware concept where central computing, network or software services are bought on service platforms. These are known as Platform as a Service (PaaS), if an organization is unable to establish its own data center or local data storage. Central indicates a single point of access to the service, which in turn may be within the cloud distributed transparently across multiple processing nodes. In the wake of Cloud Computing, different services emerged such as Software as a Service (SaaS), providing central, scalable software applications, which need not be installed locally but are accessible universally. Within the literature, Cloud Computing is strongly associated with Big Data as it offers the most benefits when huge amounts of data need to be analyzed with extensive, parallelized computing power which a local single-server unit may not be able to provide. In the context of Industry 4.0, Zhong et al. (2017) even speak of “*cloud manufacturing*” as a new model for an “*intelligent manufacturing system*” [4]. In this case, the company or the machines generating the data need to have a network connection to the cloud services to store their data inside the cloud environment and use the available processing power and tools for analysis. The results will be transmitted back to the company. One aspect discussed in this context is data privacy and intellectual property, following the externalization of company data into a booked cloud service. To meet these concerns, there are possibilities for using a public cloud, private cloud or a hybrid version (e.g., [3]). In the case of a public cloud, the data will not be public for all, but it indicates that different user groups or organizations are using this cloud or virtual environment together. The private cloud environment is available only for the organization itself. Hybrid clouds blend both concepts seamlessly.

Fog Computing: Next to Cloud Computing, Fog Computing, while not as prominent, still plays an important role in the implementation of computing resources in the Smart Factory. The OpenFog reference architecture of the OpenFog Consortium (2017) defines it as: “*A horizontal, system-level architecture that distributes computing, storage, control and networking functions closer to the users along a cloud-to-thing continuum*” [33]. Oftentimes it is also defined in relation to the other two types of computing: Cloud and Edge. Hu et al. (2017) summarize the concept of fog computing as “*Fog computing extends the cloud services to the edge of network, and makes computation, communication and storage closer to edge devices and end-users, which aims to enhance low-latency, mobility, network bandwidth, security and privacy*” [32]. The term “*edge of the network*” defines devices such as routers or gateways available for the edge devices to connect to the network. These devices are called fog nodes and offer the first central computing unit from the perspective of the edge device. Most of the time, fog nodes themselves are again connected to the cloud layer thus covering the computing unit in between [32]. Fei et al. (2019) pronounce the main difference towards Cloud Computing is that the available resources are more limited and require “*optimal management*”, while on the other hand fog nodes are much closer to the edge devices, reducing network delays [22].

Edge Computing: Edge Computing targets a computing and storage unit that is directly attached to or embedded into an edge device (e.g., IoT Device, embedded sen-

sor in a production unit). Edge computing describes a decentralized cloud computing architecture. The data generated, for example, from local computers or IoT devices, are processed immediately at or within the device itself or in the vicinity, i.e., at the outermost edge of the network of an IT infrastructure (e.g., [34]). This aspect has direct relations to stream processing, which will be discussed later in this section. This helps to offload network bandwidth and releases “an important part of the computational load from the Cloud servers” to a data center or cloud [34]. The most important feature of edge computing is the extension and trade-off of cloud services to the edge of the network. As a result, computation, communication, and storage are much closer to the end user and the vicinity of the creation of the data. These edge nodes may include, e.g., smart sensors, smartphones, smart vehicles, or dedicated edge servers [32]. Furthermore, the computing at the edge may introduce concepts of data separation where data are filtered, compacted, aggregated, and analyzed in the manufacturing environment while only resulting data with a potentially enriched information density is transmitted to other systems and services. Edge nodes may be used for different steps of the data analytics process [16]. This way it can also be scaled what information needs to be revealed to external services (e.g., public clouds) to receive a desired analytical feedback. The Edge Computing Task Group of the IIC concludes a vertical integration of edge computing over the whole technology stack of the *Industrial Internet Reference Architecture* (IIRA), while horizontally it may be used for *peer-to-peer networking, edge-device collaboration, distributed queries, distributes data management or data governance* [35].

3.3.2. Types of Data Analytics in Smart Factories

Data Analytics is a central factor for the implementation of automated or semi-automated Smart Factories. Depending on the literature reference, it may also be more specifically called *Industrial Analytics* (e.g., [35]). The analyzed data may be structured (e.g., sensor data), semi-structured (e.g., meta data or semi-structured feedback data) or even unstructured data in the form of varying texts or images. The ways of extracting knowledge from these data are diverse. “*Extracting practical knowledge from heterogeneous data is challenging and thus, determining the right methodologies and tools for querying and aggregating sensor data is crucial . . .*” [36].

Smart Factories emphasize the analysis of sensor data, due to the distributed sensor data streams created from singular or interconnected CPS or IoT devices inside the production plant or Smart Factory environment. “*Data is a key enabler for smart manufacturing*” [1]. While more data allow more insights, this holds true only if the fitting data analytics, statistics or machine learning methods are applied considering the purpose of why certain data streams or batches are being analyzed. From a knowledge integration perspective, it is important to define the use case (e.g., predictive maintenance of a machine, monitoring of the current CPS status, etc.) and fitting data analytics concepts before choosing a concrete implementation technology.

Data analytics in Smart Factories follow general lifecycle management models and dedicated analytics methods adapting them to the type of data available in the environment and the given task at hand. It is like any other data analytics application dependent on its lifecycle “*of data collection, transmission, storage, processing, visualization, and application*” [1]. The IIC (2017) summarizes the applicable data analytics method categories of *descriptive analytics, predictive analytics* and *prescriptive analytics*: For the descriptive category, both batch and stream processing may be included to create monitoring, diagnosis or reporting applications, but to support also the training of a machine learning model. Predictive analytics applies “*statistical and machine learning techniques*”, which may be used for, e.g., predictive maintenance or material consumption, while “*prescriptive analytics uses the results from predictive analytics*” to develop recommendations, process optimization and prevention of failures [16].

Next to the use case and envisioned category of analytics task, it is important to define how the specific data need to be prepared and what kind of pre-analysis steps should be

executed (e.g., [1]), such as filtering or removal of invalid or irrelevant data [37], handling of data gaps [38] where, e.g., the sensor did not work, what patterns should be detected or which parameters are needed for the visualization of results, to name only a few. The definition and application of data quality criteria are essential at this step.

While talking about descriptive analytics, the terms stream and batch processing have been mentioned. A major deciding factor for the type of analysis is time. Is there a requirement for real-time analysis or not? Depending on the answer, stream analytics/real-time analytics or batch analytics may be implemented:

Stream Analytics: Stream Analytics or Stream Processing indicates that an incoming data stream, e.g., generated from one or more sensors attached to a CPS, is being analyzed immediately, utilizing “in place” algorithms which need a limited data window to operate and are computational sparse. Aggarwal (2013) highlights, “*with increasing volume of the data, it is no longer possible to process the data efficiently by using multiple passes. Rather, one can process a data item at most once*” [39]. To meet the requirements of real-time processing, the incoming data are being separated into small frames and immediately forwarded into an “analyzer module” to generate knowledge for reactive measures or monitoring inside the production plant. Training models may be created and trained at different computing points, e.g., edge nodes [16]. Turaga and van der Schaar (2017) summarize different forms of data streaming analytics features, such as *Streaming and In-Motion Analysis* for on the fly analytics, *Distributed Data Analysis*, *High-Performance and Scalable Analysis*, *Multi-Modal Analysis*, *Loss-Tolerant Analysis* or *Adaptive and Time-varying Analysis* [40]. Stream analytics require local processing units, which is why the concepts of fog or edge computing are prevalent for this type of processing. A transfer of stream data into a cloud environment before it is being analyzed might lead to too much latency and a delay in reaction time as the results oftentimes lead to an immediate action inside the production plant. Summarizing, Fei et al. (2019) proclaim “*data stream analytics [as] one of the core components of CPS*” thus being motivated to evaluate different machine learning (ML) algorithms regarding their applicability for analyzing CPS data streams [22].

Batch Analytics: Batch Analytics follow the principle of gathering data for some time before running a fitting analysis or ML method. Depending on the batch size, this may lead to big data analytics (e.g., [1]) and the necessity of applying technologies for handling huge amounts of data. A well explored and integrated, yet continuously developing set of methods for batch processing are deep learning networks, enabling the batch-wise processing of large volumes of data with a large input-feature space [41]. Production plants oftentimes run logging or monitoring applications, building up historians of accumulated data which might also be part of the analysis process. The purpose of data accumulation instead of immediate analysis is mostly due to the need to create reports (e.g., monthly) or visualize timeline developments or machine workload. Analyzed trends or what is part of predictive analytics also depends on past data instead of only ad hoc available data streams.

Overall, it is important to decide which data may be part of the stream processing and analysis and which ones are accumulated, e.g., with the help of cloud computing. This is not an either/or decision. Some data may also be analyzed immediately, while afterwards being stored in the batch/cloud to be part of another analysis, e.g., timeline analyses, or being used to offline-update deployed algorithms. These types of data are especially valuable as they offer a dual benefit and knowledge for ad hoc and long-term decisions.

3.3.3. Simulation and Decision Making in Smart Factories Using Digital Twins

Simulation in Smart Factories is directly associated with the concept of the “*Digital Twin*” (DT). Alternatively, the terms “*Digital Shadow*” (DS) or “*Digital Model*” (DM) are being used in the literature as well, but they do not necessarily have the same meaning [42]. The term digital twin indicates that the physical device, system or entity is being represented in a digital form which has identical characteristics and allows to interact, test or manipulate it in the digital space. Before explaining more about the DT, it is important to understand

the differences to CPS or IoT as they are core building blocks of the Smart Factory and also bridge a gap between physical and cyber world. Tao et al. (2019) resume that “CPS and DTs share the same essential concepts of an intensive cyber–physical connection, real-time interaction, organization integration, and in-depth collaboration” but they are not the same concept [9]. “CPS are multidimensional and complex systems that integrate the cyber world and the dynamic physical world”, while DTs focus more on the virtual models, the matching of their behavior to the physical counterpart and feedback flow of data between both [9]. Harrison, Vera and Ahmad (2021) envision that DTs may support the whole lifecycle and engineering of CPPS [43]. Jacoby and Usländer compared the concepts of DT and IoT and related standards or representation models [44]. They resume that while they are similar and both “center on resources”, DTs aim for optimization and automation, applying computer science concepts such as AI or machine learning in an integrative manner [44].

DTs are being used to simulate situations or decision alternatives which may not be executed easily in the physical world during operation. The Industrial Internet Consortium and Plattform Industrie 4.0 (2020) summarized in a joint whitepaper, that the DT is “adequate for communication, storage, interpretation, process and analysis of data pertaining to the entity in order to monitor and predict its states and behaviors within a certain context” [45]. DT has advantages in the case of decision making, analysis of gathered data or optimization of the physical counterpart. It is important to define the scope and purpose as the DT may address various aspects next to simulation [46]. Martinez et al. (2021) recently conducted research where they reflected that DTs can be implemented on all levels of the classical automation pyramid while applying AI concepts, thus also indicating that a DT implementation is not only focused on one level but is a holistic approach [47]. Harrison, Vera and Ahmad (2021) also see an evolution of the automation pyramid due to “data-driven manufacturing capabilities” [43].

The creation of a DT requires key aspects such as the availability of underlying models, service interfaces for connecting to the DT, and data collected from the physical entity it is related to [45]. Kritzinger et al. (2017) conducted a literature study regarding DM, DS and DT and tried to differentiate between the terms based on the data flow between the physical and digital object. In the case of a Digital Model, they see only manual data flows; the Digital Shadow, however, also supports automated data flow from physical to digital object, while the Digital Twin supports a bidirectional automated data flow between the physical and digital object [42]. Schluse et al. (2018) combine these aspects and summarize that a DT “integrates all knowledge resulting from modeling activities in engineering (digital model) and from working data captured during real-world operation (digital shadow)” [46]. Hänel et al. (2021) reflect from a data and information perspective on the application possibilities of DTs “for High-Tech Machining Applications” and present a model targeting a “shadow-twin transformation loop” based on the acquired feedback data [48].

The models of a DT may cover different details such as geometrical, material or physical properties, behavior or rules as well as data models [9]. The complexity and possibilities for simulation applications in the Smart Factory based on DT depend highly on the level of detail of the DT and the available data gathered by or from the physical device and its integration and analysis in the DT. DTs may be independent/discrete or they are part of a composition of DTs thus allowing, e.g., the representation of a production line of different physical entities [45]. Tao et al. (2019) summarize this as the unit, system and system of system level of DTs [9]. From a practical perspective, Autiosalo et al. (2021) implemented different DTs with varying update times on the asset, fleet and strategic level for an overhead crane scenario, giving special focus to API standards and data flows [49].

From a knowledge integration perspective, the data flow is essential for the concept of the DT as it allows the creation of knowledge with the help of data analysis or machine learning methods, thus feeding the DT with new knowledge from its physical counterpart. At this stage there is an interrelation to the section of stream analytics (see Section 3.3.2) as the real-time simulation of the DT would require the application of real-time data processing and stream analytics as well. While real-time data may be a factor for the DT

to show the current state of the physical entity with the help of a visualized DT, historical or accumulated data during, e.g., the operation process of the physical entity, allow more complex simulations for optimization, failure studies or predictive maintenance, thus building a bridge to batch analytics (see Section 3.3.2). To realize these complex applications, authors research how concepts from AI, machine learning or other data science aspects such as big data analytics, may be applied in an efficient manner, or for which use cases of the Smart Factory they are applicable (e.g., [50]).

3.3.4. Semi- and Unstructured Data Integration

In a Smart Factory, knowledge can be mined from a range of data types. Sensor readings, for example, represent a structured data type, where the meaning of each value is definite. However, process documentations in modern industries also hold a great potential to capture expert knowledge in Smart Factories. Such documentations include verbal descriptions of procedures, failures, maintenance, etc. and it reflects inherently the experience of the human who is authoring the document. These user-specific inputs from the experts, like textual formats in general, are only available in a semi-structured or unstructured form. The latter type represents a completely free user input, whose meaning is only determined by analyzing the written text. An example of this type can be a customer complaint report. The former type, semi-structured data, features some information about the free text that can help in determining its meaning. For example, a failure report can take the form of a table, in which certain fields are filled with free textual input. Here, although the cell content is free text, the cell itself is predefined by the table structure to contain information about a certain aspect of the failure, e.g., the description of a failure reason, which makes the overall data in the table semi-structured.

Extracting knowledge from semi- and unstructured data requires methods that can automatically analyze the textual content to mine the information and represent its meaning. For this reason, we will highlight in this section the role of text mining and semantic knowledge representation for extracting and modeling knowledge from these two types of data, in order to integrate this knowledge into the overall decision-making process of a Smart Factory.

Application of Text Mining and Text Generation in Smart Factories

Text mining is a collection of methods and algorithms intended for information extraction and knowledge discovery from structured, semi-structured and unstructured textual data [51,52]. The utilization of text mining methods and approaches has a wide spectrum in the state of the art. Its applications include:

- Information extraction and retrieval: The information extraction process can be described as turning the unstructured textual data into structured and usable information. Since textual data are prone to including inconsistencies and human errors or potentially irrelevant data, a pre-processing phase takes place before the information extraction. After the raw text is pre-processed, several approaches have been addressed in the literature for the extraction of knowledge from the text, including Named Entity Recognition (NER) [53] and Relation Extraction (RE) [54]. Both methods are directly relevant for a semantic representation of data, since entities and relations can be identified in a meaningful way to define and populate the nodes and edges, in graph based representations of texts, as, e.g., knowledge graphs, while utilizing additional pre-existing domain knowledge [55].
- Classification, clustering and topic modelling: Classification is one of the important methods that allow creating a connection between a document and a search query. It is utilized in several applications, including the medical, commercial and industrial design processes. In the latter, Jiang et al. (2017) have developed an algorithm that predicts the importance of a product feature based on written user reviews [56]. Traditional text classification methods include Naïve Bayes, Support Vector Machines (SVM), and k-Nearest Neighbor. All those algorithms depend on the features of a

document and the predefined labels that classify this document into a certain category. Unlike traditional approaches, which use Term Frequency–Inverse Document Frequency (TF-IDF), cosine similarities or probability functions, machine and deep learning algorithms learn to map document features to the available labels through creating a matching function that enables the system of categorizing a new document to one or more of the predefined classes based on its features.

In contrast to classification, clustering and topic modelling are unsupervised learning approaches. Clustering and topic modelling have been applied in the design and engineering fields. They, e.g., enable the analysis of previous textual data, collected from various process steps and activities, such as emails, change logs and regular reports, for knowledge extraction and reuse. Grieco et al. (2017) have applied clustering on free text written on the Engineering Change Request (ECR) [57]. This is a type of engineering design log record of frequently required changes during the design phase, which are collected for products and their components. This approach has been observed to significantly help to summarize the main features and changes to the product during the development process of previous projects.

Smart factories require architectures that integrate knowledge and focus on textual analysis of unstructured data. One approach to actively utilize unstructured data is the creation of an Intelligent Document and therefore the generation of texts, e.g., for reporting purposes, which are generated on demand and may include interactive elements [26]. An intelligent document collects and contains references to the identified process documentation and relevant extracted knowledge sources. This includes expert identifiers as explicit entities to indicate the source creator and therefore potentially more implicit process knowledge. Such linked resources can be either followed by the user through generated links or be imported into the intelligent document to be composed into one homogenous text. Besides reporting purposes, additional applications can be the analysis and integration of results from past incident analyses as a so-called “lessons learned” documentation. In the case of subsequent changes to product generations and processes, identified errors resulting from incident analyses (e.g., log analysis) and error causes during manufacturing (machine and production reports) can be considered. Further text generation techniques can be used to improve the blending of imported texts [58]. For the development of intelligent documents, Zenkert et al. (2018a) propose a four-step process [26]:

- Acquisition and pre-indexing: In the acquisition and indexing phase, all relevant knowledge assets—including experts and extracted content from the knowledge base—are considered for extraction. Domain specific information such as textual information from design and process documentation is considered and subsequently identified through text analysis.
- Initialization: In the initialization phase, relevant knowledge sources are collected on the basis of the acquisition and indexing results, and metadata is assigned to documents. To sustain the extraction context and metadata of extraction algorithms, such results are stored using a graph-based knowledge representation concept, called Multidimensional Knowledge Representation (MKR) [59]. Based on the graph-based MKR of each document, other entities and semantically related content from other documents can be referenced [59]. Each extracted document is processed using text mining techniques to feed into the meta-structure of the MKR.
- Process Enrichment: In the process enrichment phase, relevant content is collected from the MKR structure, based on the relationships between elements and the underlying texts, and put into the format of an intelligent document. Each generated, intelligent document is a composition of identified knowledge assets. Based on the interaction of users, links to further resources can be followed and the document can be modified on demand and further enriched by additional asset recommendations.
- Automated Update Cycle: The automatic enrichment of the intelligent documents runs iteratively and continuously and modifies documents or adds content to new assets

when new or changed source texts or extraction results are available (e.g., new meta-data such as keywords, entity links, related process contexts, new incident reports).

Semantic Knowledge Representation in Smart Factories

Knowledge graphs (KG), or knowledge maps, are graphical representations of a knowledge base. Whether ontologies or other semantic representations that are based on relations between graph entities, such networks are well suitable for knowledge representation and decision support in complex environments. As a knowledge representation tool, knowledge graphs are constructed through analyzing available information and representing it and its relevant semantic relationships as a graph through using rule-based integration mechanisms or adhering to a formal representation structure. The added value of such semantic representations is rooted in their ability to describe the relationships between pieces of information. This interconnectivity and contextualization of information enables it to be used as a medium of stored knowledge. To extract information from a KG, knowledge querying procedures are used.

Yahya et al. (2021) surveyed the use of semantic technologies in the scope of industry 4.0 and smart manufacturing [60]. They point out the multiple domains, in which KGs are involved to solve production challenges, such as predictive maintenance, resource optimization, and information management. They also propose an enhanced and generalized model, namely the reference generalized ontological model (RGOM), based on RAMI 4.0. Another study by Beden et al. (2021) analyzes the role of semantic technologies in the asset administration shells in RAMI 4.0 [61]. The authors show that semantic approaches enhance the way physical assets are digitally represented through adding an element of formalism to the knowledge representation.

Interaction with the knowledge graph, in terms of their construction and querying, is essential for their utilization. Several approaches for this process have been investigated, including Triple Pattern Fragments [62] and Example Entity Tuples [63]. Knowledge graphs have been exploited in several applications. Their semantic structure was used alongside text mining in order to enrich a graph with extracted textual data. The result of this combination is also addressed by Dietz et al. (2018), who highlighted in their summary on the utilization of knowledge graphs with text mining, the rising role of knowledge graphs in text-centric retrieval, especially for search system applications [64]. This further highlights the interaction between knowledge graphs and natural language processing (NLP). Knowledge graphs have provided the opportunity to bridge the gap between the data-driven and symbolic approaches, which considerably influenced the research in this field [65–67]. On the other hand, knowledge graphs can be constructed in different ways, being tailored to the input data and the graph's domain of application. As a consequence, the information extracted from graphs with different semantic representations and tailoring can also have different relevance in the specific targeted application [26].

The ability of knowledge graphs to provide the foundation for search and matching tasks can be supported by several data mining techniques. For example, Tiwari et al. (2015) utilize RDF triples to better represent large, heterogeneous, and distributed databases [68]. Their approach relies on RDFs being a machine-understandable format, which can be integrated with human-understandable sentences using natural language processing and generation. Abu-Rasheed et al. (2022) construct a multidimensional knowledge graph from multiple documentation sources in the electronic chip design process [69]. They develop a domain-specific text mining workflow, in which expert knowledge of the domain vocabulary complements and enhances the intelligent model's predictions. The results from the text mining are then used to define the nodes and relations of the knowledge graph for different entity types, where each type is tailored to a certain information source. The knowledge graph is then used as a source and graphical representation for an explainable information retrieval task, which is based on a transitive graph-based search and relation-based reasoning [69].

Although the nature of knowledge graphs depends on describing the information in the form of subject–predicate–object triples, labeled graphs can also be represented in the form of adjacency tensors [70,71]. In this form, machine learning and deep neural networks can be utilized for the analysis process and thus for prediction and decision support.

Contemporary research suggests that not only machine learning and deep learning techniques are enhancing the construction of knowledge graphs, but also the graphs themselves are utilized for enhancing the performance of those techniques. Together, the combination of both approaches is investigated for empowering prediction processes. This is observed in the work of Qin et al. (2018), who suggest an approach for the anomaly detection problem based on a deep learning model that has been supported by a knowledge graph [72].

The previous approaches of semi- and unstructured data analysis and integration complement the other procedures in a Smart Factory to accomplish knowledge integration and enhance processes with decision support. Text mining and semantic technologies play an important role in this integration process. They provide tailored solutions to utilize semi- and unstructured data in addressing the needs and requirements of the factory's domain of application.

4. Conclusions

In this article, the relevant concepts of knowledge integration in a Smart Factory were presented. Different aspects of the Smart Factory environment and its architecture as well as the building blocks such as the Internet-of-Things (IoT), Cyber–Physical Systems (CPS) and Cyber–Physical Production Systems (CPPS) were explained and set into context. Horizontal and vertical knowledge integration in a Smart Factory were presented from the organizational, employee, and technology viewpoint as overarching perspectives of knowledge management in Smart factories.

From the organizational perspective, the horizontal integration of production and transportation processes along the value chain was explained. The vertical perspective shows how the Smart Factory manufacturing environment is changing towards a data-driven, decentralized environment and how this has to be connected and synchronized with other organizational units such as planning, logistics, sales and marketing. A key aspect for both integration perspectives is the data gathered along the value chain and the extracted and related knowledge to support offline and real-time decision-making.

From the employee perspective, the role of employees in Smart Factories was discussed and concepts such as the human–machine interaction (HMI) and the increasing need to interact intelligently with machines, to learn and work jointly were highlighted.

From the technology perspective, this article focuses on the fundamental concepts of Cloud Computing, Fog Computing, and Edge Computing, as well as intelligent methods to support and facilitate the implementation of and knowledge integration within a Smart Factory. In this context, the article addresses the processing variations of structured data in terms of stream analysis and batch analysis. Furthermore, the concept of the Digital Twin is explained and the possibilities for simulation and decision support are shown. In addition to the processing of structured data, the high potential of unstructured data in smart factories is increasingly being recognized but yet not fully unearthed for the Smart Factory of the future. Therefore, this article highlights a range of techniques for unstructured data processing, such as semantic analysis and text mining, with new forms of knowledge representation, such as knowledge graphs and applications such as intelligent documents.

Considering the presented state of the art, Smart Factories are a reality now. Data-driven techniques and the integration of information on all levels of a factory have working solutions in theory and practice. However, the leap towards knowledge integration is scarce and nuanced consideration of information or knowledge has not been made. To do the leap, strategies for a steady contextualization of information across the value chain are needed. Data-driven techniques are foremost quantitatively driven, deriving, and recognizing patterns from large amounts of data. Deriving conclusions from such patterns

and stepping up from “being informed” to “knowing how to act”, context and an actionable representation of the formed knowledge are needed. To do so, it needs contextualizing semantic techniques on the technical side and the integration of the human factor, as an informed and integrated partner, on the organizational side. Both needs do have working solutions independently, but the joint integration into the highly flexible, complex, digitized and rapidly changing environment of a Smart Factory still needs better solutions. Following the perspective of knowledge integration, as envisioned by this article, will create the fundament for this new vision of integration.

Author Contributions: Conceptualization, J.Z., C.W., M.D. and H.A.-R.; methodology, J.Z., C.W., M.D., H.A.-R. and M.F.; validation, J.Z., C.W., M.D. and H.A.-R.; formal analysis, J.Z., C.W., M.D., H.A.-R. and M.F.; investigation, J.Z., C.W., M.D., H.A.-R. and M.F.; resources, M.D.; writing—original draft preparation, J.Z., C.W., M.D. and H.A.-R.; writing—review and editing, J.Z., C.W., M.D., H.A.-R. and M.F.; visualization, M.D.; supervision, J.Z., M.D. and M.F.; project administration, J.Z. and M.F.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank their students Hüseyin Turan and Sefa Colak for their support in the pre-study phase of this entry.

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

Entry Link on the Encyclopedia Platform: <https://encyclopedia.pub/14253>.

References

1. Tao, F.; Qi, Q.; Liu, A.; Kusiak, A. Data-driven smart manufacturing. *J. Manuf. Syst.* **2018**, *48*, 157–169. [CrossRef]
2. Alcácer, V.; Machado, V. Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Eng. Sci. Technol. Int. J.* **2019**, *22*, 899–919. [CrossRef]
3. Oztemel, E.; Gursev, S. Literature review of Industry 4.0 and related technologies. *J. Intell. Manuf.* **2020**, *31*, 127–182. [CrossRef]
4. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [CrossRef]
5. Neumann, W.P.; Winkelhaus, S.; Grosse, E.H.; Glock, C.H. Industry 4.0 and the human factor—A systems framework and analysis methodology for successful development. *Int. J. Prod. Econ.* **2021**, *233*, 107992. [CrossRef]
6. Roth, A. (Ed.) *Einführung und Umsetzung von Industrie 4.0: Grundlagen, Vorgehensmodell und Use Cases aus der Praxis*; Springer Gabler: Berlin, Germany, 2016; ISBN 978-3-662-48504-0.
7. Shi, Z.; Xie, Y.; Xue, W.; Chen, Y.; Fu, L.; Xu, X. Smart factory in Industry 4.0. *Syst. Res. Behav. Sci.* **2020**, *37*, 607–617. [CrossRef]
8. Strozzi, F.; Colicchia, C.; Creazza, A.; Noè, C. Literature review on the ‘Smart Factory’ concept using bibliometric tools. *Int. J. Prod. Res.* **2017**, *55*, 6572–6591. [CrossRef]
9. Tao, F.; Qi, Q.; Wang, L.; Nee, A. Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. *Engineering* **2019**, *5*, 653–661. [CrossRef]
10. Frey-Luxemburger, M. (Ed.) *Wissensmanagement-Grundlagen und Praktische Anwendung: Eine Einführung in das IT-Gestützte Management der Ressource Wissen*, 2nd ed.; IT im Unternehmen; Springer Vieweg: Wiesbaden, Germany, 2014; ISBN 978-3-658-04752-8.
11. Capestro, M.; Kinkel, S. Industry 4.0 and Knowledge Management: A Review of Empirical Studies. In *Knowledge Management and Industry 4.0: New Paradigms for Value Creation*; Bettiol, M., Di Maria, E., Micelli, S., Eds.; Knowledge Management and Organizational Learning; Springer International Publishing: Cham, Switzerland, 2020; pp. 19–52. ISBN 978-3-030-43589-9.
12. Brauckmann, O. *Smart Production: Wertschöpfung durch Geschäftsmodelle*; Springer Vieweg: Berlin, Germany, 2015; ISBN 978-3-662-45301-8.
13. Bettiol, M.; Maria, E.D.; Micelli, S. (Eds.) *Knowledge Management and Industry 4.0: New Paradigms for Value Creation*; Knowledge Management and Organizational Learning; Springer International Publishing: Cham, Switzerland, 2020; ISBN 978-3-030-43588-2.
14. Ansari, F.; Erol, S.; Sihni, W. Rethinking Human-Machine Learning in Industry 4.0: How Does the Paradigm Shift Treat the Role of Human Learning? *Procedia Manuf.* **2018**, *23*, 117–122. [CrossRef]
15. Seeber, I.; Bittner, E.; Briggs, R.O.; de Vreede, T.; de Vreede, G.-J.; Elkins, A.; Maier, R.; Merz, A.B.; Oeste-Reiß, S.; Randrup, N.; et al. Machines as teammates: A research agenda on AI in team collaboration. *Inf. Manag.* **2020**, *57*, 103174. [CrossRef]
16. Industrial Internet Consortium (IIC). The Industrial Internet of Things Volume T3: Analytics Framework. Available online: https://iiconsortium.org/pdf/IIC_Industrial_Analytics_Framework_Oct_2017.pdf (accessed on 6 August 2021).
17. Plattform-i4.0 Referenzarchitekturmodell (RAMI) 4.0 (Reference Architecture Model). Available online: <https://www.plattform-i40.de/PI40/Redaktion/DE/Infografiken/referenzarchitekturmodell-4-0.html> (accessed on 6 August 2021).
18. Bagheri, B.; Yang, S.; Kao, H.-A.; Lee, J. Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment. *IFAC-PapersOnLine* **2015**, *48*, 1622–1627. [CrossRef]

19. Karnouskos, S.; Ribeiro, L.; Leitão, P.; Luder, A.; Vogel-Heuser, B. Key Directions for Industrial Agent Based Cyber-Physical Production Systems. In Proceedings of the 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), Taipei, Taiwan, 6–9 May 2019; pp. 17–22.
20. Dornhöfer, M.; Sack, S.; Zenkert, J.; Fathi, M. Simulation of Smart Factory Processes Applying Multi-Agent-Systems—A Knowledge Management Perspective. *J. Manuf. Mater. Process.* **2020**, *4*, 89. [CrossRef]
21. Soic, R.; Vukovic, M.; Skocir, P.; Jezic, G. Context-Aware Service Orchestration in Smart Environments. In *Agents and Multi-agent Systems: Technologies and Applications 2019*; Jezic, G., Chen-Burger, Y.-H.J., Kusek, M., Šperka, R., Howlett, R.J., Jain, L.C., Eds.; Springer: Singapore, 2020; pp. 35–45.
22. Fei, X.; Shah, N.; Verba, N.; Chao, K.-M.; Sanchez-Anguix, V.; Lewandowski, J.; James, A.; Usman, Z. CPS data streams analytics based on machine learning for Cloud and Fog Computing: A survey. *Futur. Gener. Comput. Syst.* **2019**, *90*, 435–450. [CrossRef]
23. Gorodetsky, V.I.; Kozhevnikov, S.S.; Novichkov, D.; Skobelev, P.O. The Framework for Designing Autonomous Cyber-Physical Multi-Agent Systems for Adaptive Resource Management. In *Industrial Applications of Holonic and Multi-Agent Systems*; Mařík, V., Kadera, P., Rzevski, G., Zoitl, A., Anderst-Kotsis, G., Tjoa, A.M., Khalil, I., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 52–64.
24. Industrial Internet Consortium (IIC). The Industrial Internet of Things Volume G1: Reference Architecture. Available online: <https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf> (accessed on 6 August 2021).
25. Bullinger, H.-J.; Wörner, K.; Prieto, J. Wissensmanagement—Modelle und Strategien für die Praxis. In *Wissensmanagement: Schritte zum Intelligenten Unternehmen*; Bürgel, H.D., Ed.; Edition Alcatel SEL Stiftung; Springer: Berlin/Heidelberg, Germany, 1998; pp. 21–39. ISBN 978-3-642-71995-0.
26. Zenkert, J.; Weber, C.; Klahold, A.; Fathi, M.; Hahn, K. Knowledge-Based Production Documentation Analysis: An Integrated Text Mining Architecture. In Proceedings of the 2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), Windsor, ON, Canada, 5–8 August 2018; pp. 717–720.
27. Plattform-i4.0 Infographic about Hierarchy in Industry 3.0. Available online: <https://www.plattform-i40.de/PI40/Redaktion/DE/Infografiken/hierarchie-in-der-industrie-3-0.html> (accessed on 6 August 2021).
28. Plattform-i4.0 Infographic about Hierarchy in Industry 4.0. Available online: <https://www.plattform-i40.de/PI40/Redaktion/DE/Infografiken/hierarchie-in-der-industrie-4-0.html> (accessed on 6 August 2021).
29. Vogel-Heuser, B.; Bauernhansl, T.; Hompel, M. (Eds.) *Handbuch Industrie 4.0 Bd.4: Allgemeine Grundlagen*, 2nd ed.; VDI Springer Reference; Springer Vieweg: Berlin, Germany, 2017; ISBN 978-3-662-53253-9.
30. North, K.; Maier, R.; Haas, O. Value Creation in the Digitally Enabled Knowledge Economy. In *Progress in IS*; Springer International Publishing: Berlin, Germany, 2018; pp. 1–29.
31. Gillespie, R.B.; Colgate, J.E.; Peshkin, M.A. A General Framework for Cobot Control. *IEEE Trans. Robot. Autom.* **2001**, *17*, 391–401. [CrossRef]
32. Hu, P.; Dhelim, S.; Ning, H.; Qiu, T. Survey on fog computing: Architecture, key technologies, applications and open issues. *J. Netw. Comput. Appl.* **2017**, *98*, 27–42. [CrossRef]
33. OpenFog Consortium OpenFog Reference Architecture for Fog Computing. Available online: https://www.iiconsortium.org/pdf/OpenFog_Reference_Architecture_2_09_17.pdf (accessed on 6 August 2021).
34. Sittón-Candanedo, I.; Alonso, R.S.; Rodríguez-González, S.; García Coria, J.A.; De La Prieta, F. Edge Computing Architectures in Industry 4.0: A General Survey and Comparison. In *International Workshop on Soft Computing Models in Industrial and Environmental Applications*; Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J.A., Quintián, H., Corchado, E., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 121–131.
35. Industrial Internet Consortium Edge Computing Task Group Introduction to Edge Computing. Available online: <https://hub.iiconsortium.org/intro-edge-computing> (accessed on 6 August 2021).
36. Lim, K.Y.H.; Zheng, P.; Chen, C.-H. A state-of-the-art survey of Digital Twin: Techniques, engineering product lifecycle management and business innovation perspectives. *J. Intell. Manuf.* **2020**, *31*, 1313–1337. [CrossRef]
37. Papadimitriou, S.; Sun, J.; Faloutsos, C.; Yu, P.S. Dimensionality Reduction and Filtering on Time Series Sensor Streams. In *Managing and Mining Sensor Data*; Springer: Boston, MA, USA, 2012; pp. 103–141.
38. Enders, C.K. *Applied Missing Data Analysis*, 1st ed.; Guilford Publications: New York, NY, USA, 2010; ISBN 978-1-60623-639-0.
39. Aggarwal, C.C. Mining Sensor Data Streams. In *Managing and Mining Sensor Data*; Springer: Boston, MA, USA, 2012; pp. 143–171.
40. Turaga, D.S.; Van Der Schaar, M. Distributed Online Learning and Stream Processing for a Smarter Planet. In *Fog for 5G and IoT*; Chiang, M., Balasubramanian, B., Bonomi, F., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017; pp. 234–260. ISBN 978-1-119-18720-2.
41. Zhang, Q.; Yang, L.-T.; Chen, Z.; Li, P. A survey on deep learning for big data. *Inf. Fusion* **2018**, *42*, 146–157. [CrossRef]
42. Kritzing, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* **2018**, *51*, 1016–1022. [CrossRef]
43. Harrison, R.; Vera, D.A.; Ahmad, B. A Connective Framework to Support the Lifecycle of Cyber-Physical Production Systems. *Proc. IEEE* **2021**, *109*, 568–581. [CrossRef]
44. Jacoby, M.; Usländer, T. Digital Twin and Internet of Things—Current Standards Landscape. *Appl. Sci.* **2020**, *10*, 6519. [CrossRef]

45. Industrial Internet Consortium and Plattform Industrie 4.0 Digital Twin and Asset Administration Shell Concepts and Application in the Industrial Internet and Industrie 4.0: An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper. Available online: <https://www.iiconsortium.org/pdf/Digital-Twin-and-Asset-Administration-Shell-Concepts-and-Application-Joint-Whitepaper.pdf> (accessed on 6 August 2021).
46. Schluse, M.; Priggemeyer, M.; Atorf, L.; Rossmann, J. Experimentable Digital Twins—Streamlining Simulation-Based Systems Engineering for Industry 4.0. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1722–1731. [\[CrossRef\]](#)
47. Martinez, E.; Ponce, P.; Macias, I.; Molina, A. Automation Pyramid as Constructor for a Complete Digital Twin, Case Study: A Didactic Manufacturing System. *Sensors* **2021**, *21*, 4656. [\[CrossRef\]](#)
48. Hänel, A.; Seidel, A.; Frieß, U.; Teicher, U.; Wiemer, H.; Wang, D.; Wenkler, E.; Penter, L.; Hellmich, A.; Ihlenfeldt, S. Digital Twins for High-Tech Machining Applications—A Model-Based Analytics-Ready Approach. *J. Manuf. Mater. Process.* **2021**, *5*, 80. [\[CrossRef\]](#)
49. Autiosalo, J.; Ala-Laurinaho, R.; Mattila, J.; Valtonen, M.; Peltoranta, V.; Tammi, K. Towards Integrated Digital Twins for Industrial Products: Case Study on an Overhead Crane. *Appl. Sci.* **2021**, *11*, 683. [\[CrossRef\]](#)
50. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [\[CrossRef\]](#)
51. Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E.D.; Gutierrez, J.B.; Kochut, K. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *arXiv* **2017**, arXiv:1707.02919.
52. Shi, F.; Chen, L.; Han, J.; Childs, P. A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval. *J. Mech. Des.* **2017**, *139*, 111402. [\[CrossRef\]](#)
53. Nadeau, D.; Sekine, S. A survey of named entity recognition and classification. *Linguisticae Investig.* **2007**, *30*, 3–26. [\[CrossRef\]](#)
54. Pawar, S.; Palshikar, G.K.; Bhattacharyya, P. Relation Extraction: A Survey. *arXiv* **2017**, arXiv:1712.05191.
55. Abu Rasheed, H.; Weber, C.; Zenkert, J.; Czerner, P.; Krumm, R.; Fathi, M. A Text Extraction-Based Smart Knowledge Graph Composition for Integrating Lessons Learned during the Microchip Design. In *Advances in Intelligent Systems and Computing*; Springer International Publishing: Cham, Switzerland, 2020; pp. 594–610.
56. Jiang, H.; Kwong, C.K.; Yung, K.L. Predicting Future Importance of Product Features Based on Online Customer Reviews. *J. Mech. Des.* **2017**, *139*, 111413. [\[CrossRef\]](#)
57. Grieco, A.; Pacella, M.; Blaco, M. On the Application of Text Clustering in Engineering Change Process. *Procedia CIRP* **2017**, *62*, 187–192. [\[CrossRef\]](#)
58. Klahold, A.; Fathi, M. *Computer Aided Writing*; Springer International Publishing: Berlin, Germany, 2020.
59. Zenkert, J.; Klahold, A.; Fathi, M. Knowledge Discovery in Multidimensional Knowledge Representation Framework: An Integrative Approach for the Visualization of Text Analytics Results. *Iran J. Comput. Sci.* **2018**, *1*, 199–216. [\[CrossRef\]](#)
60. Yahya, M.; Breslin, J.; Ali, M. Semantic Web and Knowledge Graphs for Industry 4.0. *Appl. Sci.* **2021**, *11*, 5110. [\[CrossRef\]](#)
61. Beden, S.; Cao, Q.; Beckmann, A. Semantic Asset Administration Shells in Industry 4.0: A Survey. In Proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada, 10–12 May 2021; pp. 31–38.
62. Heling, L.; Acosta, M.; Maleshkova, M.; Sure-Vetter, Y. Querying Large Knowledge Graphs over Triple Pattern Fragments: An Empirical Study. In *The Semantic Web-ISWC 2018*; Vrandečić, D., Bontcheva, K., Suárez-Figueroa, M.C., Presutti, V., Ce-lino, I., Sabou, M., Kaffee, L.-A., Simperl, E., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11137, pp. 86–102. ISBN 978-3-030-00667-9.
63. Jayaram, N.; Khan, A.; Li, C.; Yan, X.; Elmasri, R. Querying Knowledge Graphs by Example Entity Tuples. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 2797–2811. [\[CrossRef\]](#)
64. Dietz, L.; Kotov, A.; Meij, E. Utilizing Knowledge Graphs for Text-Centric Information Retrieval. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 1387–1390.
65. Li, H.; Xiong, C.; Callan, J. Natural Language Supported Relation Matching for Question Answering with Knowledge Graphs. In *The First Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis (KG4IR 2017)*; CEUR: Tokyo, Japan, 2017.
66. Shekarpour, S.; Marx, E.; Auer, S.; Sheth, A. RQUERY: Rewriting Natural Language Queries on Knowledge Graphs to Alleviate the Vocabulary Mismatch Problem. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
67. Hu, S.; Zou, L.; Yu, J.X.; Wang, H.; Zhao, D. Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 824–837. [\[CrossRef\]](#)
68. Tiwari, A.; Rajesh, K.; Srujana, P. Semantically Enriched Knowledge Extraction With Data Mining. *Int. J. Comput. Appl. Technol. Res.* **2015**, *4*, 7–10. [\[CrossRef\]](#)
69. Abu-Rasheed, H.; Weber, C.; Zenkert, J.; Krumm, R.; Fathi, M. Explainable Graph-Based Search for Lessons-Learned Documents in the Semiconductor Industry. In *Intelligent Computing*; Arai, K., Ed.; Springer International Publishing: Cham, Switzerland, 2022; pp. 1097–1106.
70. Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E. A Review of Relational Machine Learning for Knowledge Graphs. *Proc. IEEE* **2015**, *104*, 11–33. [\[CrossRef\]](#)

-
71. Liu, Z.; Han, X. Deep Learning in Knowledge Graph. In *Deep Learning in Natural Language Processing*; Deng, L., Liu, Y., Eds.; Springer: Singapore, 2018; pp. 117–145. ISBN 978-981-10-5208-8.
 72. Qin, Z.; Cen, C.; Jie, W.; Gee, T.S.; Chandrasekhar, V.R.; Peng, Z.; Zeng, Z. Knowledge-Graph Based Multi-Target Deep-Learning Models for Train Anomaly Detection. In *Proceedings of the 2018 International Conference on Intelligent Rail Transportation (ICIRT)*, Singapore, 12–14 December 2018; pp. 1–5.