Tampere University

WAEL M. MOHAMMED

A Methodology for Architecting Digital Twins in Factory Automation

Tampere University Dissertations 1014

Tampere University Dissertations 1014

WAEL M. MOHAMMED

A Methodology for Architecting Digital Twins in Factory Automation

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Engineering and Natural Sciences
of Tampere University,
for public discussion in the auditorium room FA032
of the Festia building, Korkeakoulunkatu 8, Tampere,
on 10th May 2024, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Engineering and Natural Sciences
Finland

Responsible supervisor and Custos	Professor Jose L. Martinez Lastra Tampere University Finland	
Supervisor	Dr. Rodolfo Haber Guerra Spanish National Research Council (CSIC) Spain	
Pre-examiners	Professor Okyay Kaynak Bogazici University Türkiye	Professor Yang Shi University of Victoria Canada
Opponents	Professor Okyay Kaynak Bogazici University Türkiye	Associate Professor Bilal Ahmad University of Warwick United Kingdom

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-3422-2 (print)
ISBN 978-952-03-3423-9 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-3423-9

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2024

ABSTRACT

The fourth industrial revolution (Industry 4.0) gathered substantial attention form academic communities during the past years. As a multi-disciplinary concept, Industry 4.0 aims at formalizing the integration of digital technologies like Internet of Things (IoT), Artificial Intelligence (AI), and Big Data with factory automation and robotics. This integration aims at improving the overall performance of the manufacturing systems. Even though the Industry 4.0 has arrived as a new concept, it includes topics and concepts that have been introduced previously like the Lean Management, among others. This doctoral research includes a systematic review to position technical trends and concepts that contribute to the foundation of Lean 4.0 concept. Besides other outcomes, this review shows that Digital Twins (DT) acts as a facilitator of the concept of Digital Factories (DF). In general terms, a DT consists of a physical system, a digital replica of the physical system, and a continuous exchange of information between these two systems as initially introduced by Micheal Grieves. Therefore, implementing DTs in factory automation will contribute to digitising the factories, which in return, contributes to implementing a Lean 4.0 concept.

Digital Twins technology can be applied in various domains including manufacturing, healthcare, and urban development. In manufacturing, the concept of a process digital twin requires structured coordination of various systems that operate concurrently and interact within the domain of digital factories. These systems may include shop floor systems, Manufacturing Execution Systems (MES), and Enterprise Resource Planning (ERP). For a DT to interact with these systems, and as demonstrated in this doctoral research, it is important to adopt a generic and flexible approach as these systems may change its interfaces or core algorithms.

Fundamentally, an essential element within the digital twins' system is the information model. Traditionally, these systems relied on databases due to its technical performance. However, to meet their growing demands, a more robust, yet

flexible technology is required. In this regard, knowledge-based systems may function as a catalyst for constructing process digital twins. Furthermore, ontologies can be employed as a backbone for these knowledge systems, as it has been demonstrated as a suitable technology during this doctoral research.

In terms of building DTs for manufacturing process, this research considers DT architecture as a comprehensive set of directives for constructing the system, identifying the components and subcomponents that constitutes the DT, and detailing the blueprint of the components' interactions. In other words, the core objective of this research includes providing guidelines for building process digital twins that features qualitative attributes like modularity, scalability, reusability, interoperability, and composability. These attributes prove to be essential for satisfying the target of providing generic and flexible solution for digitising factories. In addition, and as a result of this research, the presented architecture includes implementation that constructs digital twins using a multi-view, multi-layer, and multi-perspective which asserted the importance of the aforementioned attributes.

Besides the benefits of simulation and monitoring manufacturing processes, an essential additional advantage of the manufacturing processes DT is its capability to serve as a synthetic data generator. As demonstrated in this doctoral research, this practise enables developers of AI systems to collect big amount of data without the disturbing the real process. In fact, this data holds great significance in today's industrial landscape where novel AI techniques can enhance the functionality of physical systems.

This thesis is a compendium dissertation, which includes five peer-reviewed articles.

PREFACE

The research outcome reported in this thesis was performed within the Future Automation Systems and Technology Laboratory (FAST-Lab), Automation Technology and Mechanical Engineering (ATME) unit, Faculty of Engineering and Natural Sciences (ENS), Tampere University, Finland. This research was financially supported by a doctoral school grant covering the period of 2018-2022.

ACKNOWLEDGEMENT

بسم الله الرحمن الرحيم

الحمد لله وحده لا شريك له، له الفضل الخالص والشكر الواصل على ما أنعم عليّ خلال دراستي وعملي.

"دراسة الدكتوراه هي واحدة من العديد من السباقات التي تواجهها في حياتك. حاول أن تنهي وتتقدم، حتى تتمكن من خوض سباقات أخرى أكثر إثارة!!" بهذه الكلمات، كان البروفيسور لاسترا يشجعني على إنهاء رحلة الدكتوراه. أنا ممتن له بعمق على إرشاده وخبرته ودعمه. إرشادك لم يشكل فقط اتجاه هذه الأطروحة، ولكنه ساهم بشكل كبير أيضًا في نموي كباحث.

أود أن أشكر أيضا بروفيسور كيناك، بروفيسور بلال وبروفيسور يانج شي على والوقت الذي قدموه لمراجعة ومناقشة الاطروحة.

ثم، أود أن أعبر عن تقديري الصادق لكل زملائي في الفاست لاب الذين شاركوا معي هذه الرحلة. شكرًا لآن، لويس وماتي على المساعدة والدعم خلال الرحلة الطويلة معًا. أود أن أوصل هذا الشكر إلى صامويل وأمير على الصداقة الطويلة التي بدأت مع درجة الماجستير في عام 2014. كما أود أن أشكر "العصابة الإسبانية" (أيتور، بابلو وأنجيلا) على الضحك والعديد من جلسات التفكير التي جمعتنا سويًا.

أود أن أشكر أصدقائي المقربين على الدعم والتفهم الذي قدموه لي وأفضل الأمنيات. أولاً، صديقي العزيز بورخا وزوجته أماليا. هناك أصدقاء وهناك أصدقاء، وبعد ذلك أنت. كذلك أود أن أشكر أصدقائي المقربين أسامة، إبراهيم وحرز الله. ما زلت أتذكر تجمعاتنا (والمرشح المفضل لدي) بالقرب من مقهى كلية الهندسة في درجة البكالوريوس.

أود أن أشكر عائلتي وأقاربي على حبهم ودعمهم وتشجيعهم وإيمانهم اللامحدود. أود أن أشكر والديّ محمد ورشيدة. أنا لا شيء بدونكما. أود أن أشكر أختي بسمة وزوجها محمد وأخوتي حاتم وزوجته هند، غانم وزوجته اعتدال، وإياد. لا شيء يساوي محادثاتنا وضحكاتنا في ساعات متأخرة من الليل. كما أود أن أشكر والديّ زوجتي ديمتري وسفيتلانا وابنتهما داريا وزوجها أدريان.

أخيرًا وليس آخرًا، أود أن أكرس هذه الأطروحة لزوجتي، شمسي، قمري وسعادتي، صوفيا. شكرًا لك لأنك كنت معي في السراء والضراء في هذه الرحلة. لكانت رحلتي بدون حبك، صداقتك، ضحكاتك، ومساعدتك شاقه وطويله.

In the name of Allah, the Most Gracious, the Most Merciful

All praise is due to Allah alone, with no partner. He alone deserves pure and endless gratitude for what He has blessed me with during my studies and work.

"The doctoral study is one of the many races that you will face in your life. Try to finish and move on, so you can do other races which are more interesting !!". With these words, Professor Lastra was encouraging me to finish my doctoral journey. I am deeply grateful to him for his guidance, expertise, and support. Your mentorship has not only shaped the direction of this thesis but has also contributed significantly to my growth as a researcher.

I would also like to thank Professor Kaynak, Professor Bilal, and Professor Yang Shi for the time they spent reviewing and examining the dissertation.

Then, I would like to express my sincere appreciation to all my colleagues at Fast Lab who shared this journey with me. Thank you, Ann, Louis, and Matt for the assistance and support throughout our long journey together. I'd like to extend this gratitude to Samuel and Amir for the long-lasting friendship that began with our master's degree in 2014. I'd also like to thank "The Spanish Gang" (Aitor, Pablo, and Angela) for the laughter and many brainstorming sessions we had together.

I want to thank my close friends for the support and understanding they've provided me with. Firstly, my close and bromance Borja and his wife Amalia. Hay amigos y amigos, y luego estas tu. I also want to thank my close friends Osama, Ibrahim, and Hirzalla. I still remember our gatherings (and my favorite candidate) near the Engineering College café during our bachelor's degree.

I would like to thank my family and relatives for their unlimited love, support, encouragement, and belief. I would like to thank my parents Mohammed and Rachida. I am nothing without you. I would like to thank my sister Basma and her husband Mohammed and my brothers Hatem and his wife Hend, Ghanem and his wife Etidal, and Eyad. Nothing comes close to our late-night talks and laughs. Also, I would like to thank my parents-in-law Demitri and Svetlana and sister-in-law Daria and her husband Adrien.

Last but not least, I want to dedicate this dissertation to my wife, my sunshine, my moon, and my happiness, Sofia. Thank you for being with me through thick and thin on this journey. My journey would have been tough and long without your love, friendship, laughter, and help.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AIDEAS	AI Driven industrial Equipment product life cycle boosting Agility, Sustainability and resilience
C2NET	Cloud Collaborative Manufacturing Networks
CPS	Cyber-physical System
CS	Computer Science
DIF	Digital Interconnected Factories
DT	Digital Twin
DF	Digital Factory
ERP	Enterprise Resource Planning
EU	European Union
HCD	Human-Centric Design
HDT	Human Digital Twin
HRA	Human Realistic Avatar
HRC	Human-Robot Collaboration
HRI	Human-Robot Interaction

I4.0	The fourth Industrial revolution
I5.0	The fifth industrial revolution
ICPS	Industrial Cyber-physical System
ICT	Information and Communication Technology
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
ISA	International Society of Automation
ISO	International Organization for Standardization
MES	Manufacturing Execution System
MR	Mixed Reality
PLC	Programmable Logic Controller
RDF	Resource Description Framework
RQ	Research Question
SoCPS	System of Cyber-physical System
SoS	System of Systems
SWRL	Semantic Web Rule Language

LIST OF ORIGINAL PUBLICATIONS

Ejsmont, K.; Gladysz, B.; Corti, D.; Castaño, F.; **Mohammed, W.M.**; Martinez Lastra, J.L. Towards 'Lean Industry 4.0′ – Current Trends and Future Perspectives. Cogent Business & Management 2020, 7, 1781995, doi:10.1080/23311975.2020.1781995. (JCR: Q3, JUFO: 1)

Mohammed, W.M.; Ferrer, B.R.; Iarovyi, S.; Negri, E.; Fumagalli, L.; Lobov, A.; Lastra, J.L.M. Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems. International Journal of Computer Integrated Manufacturing 2018, 1–13, doi:10.1080/0951192X.2017.1407874. (JCR[1]: Q2, JUFO[2]: 1)

Ramis Ferrer, B.; **Mohammed, W.M.**; Ahmad, M.; Iarovyi, S.; Zhang, J.; Harrison, R.; Martinez Lastra, J.L. Comparing Ontologies and Databases: A Critical Review of Lifecycle Engineering Models in Manufacturing. Knowl Inf Syst 2021, 63, 1271–1304, doi:10.1007/s10115-021-01558-4. (JCR: Q2, JUFO: 2)

Mohammed, W.M.; Haber, R.E.; Martinez Lastra, J.L. Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications. Machines 2022, 10, 861, doi:10.3390/machines10100861. (JCR: Q2, JUFO: 1)

Mohammed, W.M.; Martinez Lastra, J.L. FASTory Digital Twin Data. Data in Brief 2021, 35, 106912, doi:10.1016/j.dib.2021.106912. (JCR: Q3, JUFO: 1)

[1] Journal Citation Report by Web of Science (https://jcr.clarivate.com/jcr/home)

[2] Publication forum by the Finnish scientific community (https://www.julkaisufoorumi.fi/en)

AUTHORS' CONTRIBUTION ON REFEREED PUBLICATIONS

Publication I: Towards 'Lean Industry 4.0' – Current Trends and Future Perspectives

This publication is a collaboration work between Warsaw University, The University of Applied Sciences and Arts of Southern Switzerland, Centre for Automation and Robotics, Polytechnic University of Madrid, and Tampere University. As a systematic review paper, the data collection, the organization and the leadership were managed by the Krzysztof Ejsmont and Bartlomiej Gladysz. Donatella Corti, Fernando Castaño and the doctoral candidate contributed to the review to the paper and the writing as instructed by and alongside the corresponding author. Jose L. Martinez Lastra contributed to the supervision work of the doctoral candidate.

Publication II: Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems

The doctoral candidate contributed to the outcome of this publication with other researchers from Tampere University and Politecnico di Milano. The Doctoral student led the conceptualization, the development of the approach, the writing, and the review. Borja Ramis Ferrer and Sergii Iarovyi contributed to the supervision, the writing, and the review. Elisa Negri, Luca Fumagalli and Andrei Lobov contributed to the review of the publication. Jose L. Martinez Lastra Contributed to the review and editing and the supervision.

Publication III: Comparing Ontologies and Databases: A Critical Review of Lifecycle Engineering Models in Manufacturing

The publication is a collaboration between Tampere University, University of Warwick and Kalmar Global. Borja Ramis Ferrer, the doctoral candidate and Mussawar Ahmad contributed to the conceptualisation and the development of the paper approach. Borja Ramis Ferrer and Mussawar Ahmad contributed to the technology and research review. The doctoral candidate, Borja Ramis Ferrer and Sergii Iarovyi Contributed to the development of the testing methods and applications. The doctoral candidate contributed to the creation of the synthetic data using the FASTory digital twin. Jiayi Zhang and Robert Harrison contributed to the review of the paper. Jose Luis Martinez Lastra contributed to the review and supervision of the research work.

Publication IV: Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications

The doctoral candidate and Jose L. Martinez Lastra contributed to the conceptualisation in the creation of the DT paradigm. The doctoral candidate contributed to the technical review, the development of approach, the testing and validation. Rodolfo E Haber and Jose L. Martinez Lastra contributed to the supervision of the work.

Publication V: FASTory Digital Twin Data

The publication work is conducted by the doctoral candidate. Jose L. Martinez Lastra contributed to the supervision of the work and the review of the published article.

1 INTRODUCTION

1.1 Motivation and justification

Factories are highly connected entities from the shop floor level up to the Enterprise Resource Planning (ERP) level nowadays. This high connectivity contributes to advances in distributed production systems on both the information and physical aspects [1]. This distribution introduces challenges to production planning activities, such as scheduling production work, material procurement, logistics planning, production resource allocation, and maintenance. To support in easing these challenges, engineers and production planners use digital twins. A Digital Twin (DT) is a digital replica of an existing physical system, where these two twins (physical and digital) can exchange information continuously [2,3]. DTs can be used for simulation, data generation, testing and validation, reverse engineering, among other applications. In fact, the concept of using simulations can be traced back to the Middle Ages, where replicas of the woman's womb were used to train midwives [4]. Currently, and with advances in Computer Science (CS) and Information and Communication Technologies (ICT), the twinning of a physical system with a digital counterpart is possible [5].

Besides gaining significant attention from researchers in the academic communities, the concept of digital twins has gained attention in the commercial sector as well. In this regard, Attaran and Celik highlighted the growing market size for digital twins, estimated to reach 47 billion dollars by the year 2030 [6]. This momentum of DT technology is driven by advances in technologies such as Artificial Intelligence (AI), Mixed Reality (MR), and Internet of Things (IoT) [7,8]. In addition, standardization organizations have been developing different standards concerning digital twins, such as ISO 23247 Automation systems and integration—Digital twin framework for manufacturing [9], ISO/TR 24464:2000 Automation systems and integration—Industrial data—Visualization elements of digital twins [10], the ISO/IEC AWI 30172 Digital Twin-Use cases [11], and ISO/IEC AWI 30173 Digital twin—Concepts and terminology [12]. This interest in developing digital twins and employing them in industry has motivated this research to find a solution to guide the development of digital twins in actual industrial applications.

1.2 Problem statement and research questions

Developing digital twins has been frequently addressed by ad hoc approaches, where the digital twin can be developed for a specific use case with a specific purpose. This problem lies in the accessibility of the models of the physical systems. As an example, a single inverted pendulum system can be represented mathematically, as the model has been developed and published since 1960 [13]. In a simple form, an inverted pendulum problem has one variable to be monitored (the angle of the pendulum with respect to the horizon) and one variable to be actuated (the linear acceleration of the revolute joint). Thus, building a digital twin of an inverted pendulum is a straightforward task. However, in industrial applications, and specifically in production processes, a production system may include thousands of sensors scattered throughout the various process stages to collect data on every single parameter. This large amount of data points results in a big data problem [14]. Similarly, such a large production system may include tens or hundreds of actuation points to control the process. This also presents a multi-output problem. To solve such problems, data-driven modelling is employed, considering AI as a backbone for its modelling [15,16]. Thus, a trained AI model will mostly fit only a similar problem. This contributes to less flexible and scalable solutions [17]. Therefore, a research opportunity can be presented to build a systematic guideline for generally architecting digital twins of industrial applications.

In the context of industrial applications, building digital twins requires an understanding of how the physical system behaves and how the digital twin will be used. Furthermore, the requirement is specifically extended to understanding the manufacturing process and the involved resources. In fact, although the literature is rich with publications targeting digital twins and their use, there is a noticeable lack of research targeting the architecture and reference models of these digital twins. In addition, digital twins' implementation may require dynamic changes to address any changes that occur in the physical system. Thus, the architecture of the digital twin requires a flexible implementation to reduce the development and deployment times of the digital twins.

In this regard, this research aims to answer the following research questions:

RQ1. Why digital twins have become a cornerstone for Industry 4.0 and Lean manufacturing?

RQ2. What are the features needed for building a digital twin for industrial processes?

RQ3. What are the technologies that permit building a suitable knowledge-based system for developing Digital Twins' architecture?

RQ4. How to build a modular, scalable, reusable, interoperable, and composable Digital Twin?

1.3 Methodology

The research methodology in this doctoral research adopted an inductive research method. In other words, it breaks down the main problem into smaller and more manageable components. This approach aligns with the primary goal of finding solutions for architecting digital twins for industrial applications, regardless of the nature of the use case. To achieve that, the following points list the various stages that have been implemented in reaching the target objective:

1. **Collecting and Analysing Information**: This stage involves conducting research and a review of technologies and topics related to Industry 4.0. This review aims at allowing the doctoral candidate to position the research and focus on selecting and investigating the usage of available tools, techniques, and methods. Standards represent a rich source of information for this stage.
2. **Building Simple Prototypes**: At this first stage, the aim is to familiarize the doctoral candidate with the development work needed for solutions in the manufacturing domain. This includes developing a simulator of physical systems and developing applications to control this simulator and the physical system simultaneously.
3. **Selecting the Proper Technologies**: In this stage, the aim is to select the necessary tools, methods, and techniques for developing digital twins. At this stage, the candidate can examine the different directions for developing a digital twin. Moreover, existing standards can be utilized to keep the research aligned with industrial environments.

4. **Develop and Use the Prototype of the Final Envisioned Solution**: This stage represents the development of the proposed solution by the doctoral candidate. The outcome includes the development of the ontology model that describes the digital twin architecture, building the reasoning rules, and building the tools to manage the construction of the digital twin.
5. **Develop the Generic Paradigm of Architecting the Digital Twin**: This is the last stage where the doctoral candidate can produce the envisioned reference model to describe the architecture of digital twins. At this stage, the doctoral candidate has gained the knowledge and understanding of how to develop and use a digital twin (described in the previous stage), which allows for the creation of the methodology description.

1.4 Objectives

This doctoral research aims to create a methodology for architecting digital twins for industrial applications. Therefore, the main objectives are:

Obj 1. To contextualize the definition of process digital twins within industrial applications.

Obj 2. To develop an ontology model that defines the methodology for architecting digital twins.

Obj 3. To present a digital twin definition that allows modularity, scalability, reusability, interoperability, and composability.

1.5 Contributions

The main contribution of this doctoral research includes:

Con 1. A methodology for architecting process digital twins for industrial applications, regardless of the nature of the production process.

Con 2. A reference model for multi-view digital twins that guarantees modularity, scalability, reusability, interoperability, and composability.

Con 3. An industrial process digital twin architecture that adopts and extends standardized reference models and definitions.

1.6 Thesis outline

This thesis is a compilation of research peer-reviewed scientific publications. After the introduction, Chapter 2 provides a brief review of the literature related to the thesis topics. Additionally, Chapter 2 includes a summary of these technologies and their relations with the published articles. Then, Chapter 3 includes the core topic of the doctoral research. Moreover, a section highlights the contribution of each published research paper to the overall dissertation topic and to the aforementioned research questions. Afterward, the conclusion in Chapter 4 presents the knowledge learned and how this research can be utilized in future research. Finally, the published papers are appended to the thesis.

2 LITERATURE REVIEW

Factory automation has evolved over the past decades due to advancements in various fields, including information and communications technologies, computation, embedded systems, artificial intelligence, and electronics [18]. Factory automation has evolved over the past decades due to advancements in various fields, including information and communications technologies, computation, embedded systems, artificial intelligence, and electronics Figure 1. Through multidisciplinary research, the selected topics, highlighted with a thick black border and coloured background, are relevant to the published articles. Consequently, this chapter provides a brief introduction and highlights the relationship to the research work for each topic.

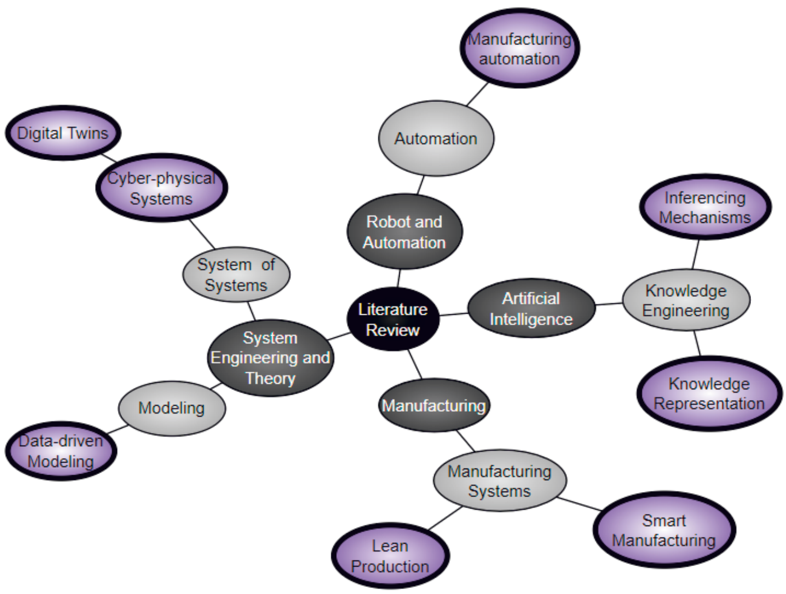


Figure 1. Taxonomy of related topics to the doctoral research

2.1 Robots and automation

2.1.1 Manufacturing automation

Manufacturing automation, also known as factory automation, refers to the use of technology to automate production processes in factories [19,20]. Initially, the objective was to improve the performance of the production system. For example, this included reducing production time, cutting production costs, and increasing production volume, among other goals [21]. However, with such objectives, new challenges have emerged, including employment reduction, concerns about human safety, and the introduction of more complex products and production processes [22]. With a substantial focus on developing solutions for factory automation that enhance the impact of industries on societal and environmental levels and align with "do no harm" policies [23], research organizations, lawmakers, and government bodies have introduced paradigms and standards to guide development in this demanding sector. For instance, the Industry 5.0 concept aims to build upon the advancements of the Industry 4.0 concept by focusing on human-centric design, sustainable industry, and resilient production networks [24].

Factory automation considers robotics a cornerstone for achieving flexible and agile production systems. Thanks to their dexterity and versatility, robots are increasingly employed in factories [25,26]. However, this utilization has had negative effects on employment in the industry [27]. To address these challenges, the trend is shifting toward collaboration and supporting humans, a concept known as "keeping the human in the loop." This collaboration, also referred to as Human-Robot Collaboration (HRC), involves physical interactions between humans and robots [28,29]. The goal of these interactions is to complete tasks where both humans and robots perform operations on the same product simultaneously. Additionally, human-robot interaction includes other forms such as social, visual, vocal, and emotional interaction, with the aim of maximizing interaction by providing trustworthiness, safety, and efficiency [30,31].

In addition, manufacturing automation involves controlling production resources through methods, techniques, and algorithms that optimize production. These control applications can be implemented at both the factory device level and the enterprise level. This integration between the factory shop floor and the enterprise level results in highly connected factories, often referred to as Smart Factories. This connectivity enables rapid planning and scheduling [32], enhances resilience by adapting to sudden changes in a fast and autonomous manner [33], and leads to overall value chain improvement by continuously monitoring the product and production lifecycle [34].

2.2 Manufacturing

2.2.1 Smart manufacturing

Smart manufacturing is a term used for addressing the use of highly connected systems and advanced technologies for maximising the factories performance. These technologies might include internet of things (IoT), artificial intelligence (AI), and big data analytics, cloud computing and edge computing [35]. With such capabilities, factories can use the data to create knowledge on how the factory perform and how the production can be improved. A possible example is to optimize the manufacturing process and improve the efficiency and competitiveness of a manufacturing facility [36]. In addition to that, and with the large amount of collected data, systems that are deployed may include forecast features to predict and estimate the production capacity [37].

Smart manufacturing may include the concept of distributed manufacturing as well [38,39]. With high levels of connectivity, factories can be connected to form collaborative manufacturing where each factory contribute to the production with its capabilities. This has been a research topic for many projects such as the C2NET[3] and AIDEAS[4] EU projects. This concept improves the efficiency of the overall production systems. However, it introduces new challenges like quality assurance issues [40]. In more details, factories are required to share detailed information about the production lifecycle to support the tracking and genealogy activities. Such an approach is possible if the manufacturing resources can share information in a real-time manner [41]. Then, mitigating measures can be implemented to avoid disruption in production.

[3] https://cordis.europa.eu/project/id/636909

[4] https://cordis.europa.eu/project/id/101057294

2.2.2 Lean production

Lean production is a manufacturing philosophy which is based on the "Lean Management" concept that aims to eliminate waste and increase efficiency in the production process. Toyota has developed the lean management concept in the 1950s, and since that, it has been adopted by a wide range of industries [42,43]. Waste in this context refers to the non-added value activities, efforts and resources that adds to the value chain an undesirable overhead, like the set-up time for production resources. To reduce, or ideally, eliminate this waste, production systems must encapsulate flexible and adaptable approaches for managing the resource, planning the manufacturing processes, planning, and scheduling the logistics among others [44,45]. The implementation methodology of the lean production typically follows a set of principles, including defining the value of the product, identifying the value stream, creating continuous flow, employing pull production strategy, and pursuing perfection. As an example, utilizing of DT technologies to simulate and help predict production performance as in [46]. Overall, lean production can help companies to increase efficiency, reduce costs, and improve customer satisfaction by focusing on continuous improvement and eliminating waste. It requires a culture of collaboration and a commitment to continuous learning and improvement [43].

2.3 Artificial intelligence

2.3.1 Knowledge representation

Knowledge representation is the way in which reasoned information is encoded and structured in a computer system. It is a fundamental aspect of artificial intelligence and refers to the methods and formalisms used to represent knowledge and reason about it [47]. There are several approaches to knowledge representation, including semantics networks, ontologies, rule-based systems, and logical representations among others [48]. Semantic networks are graphical representations that show the relationships between concepts and the properties of those concepts. In industrial application, knowledge-based system that uses semantic networks have been used intensively to build models for describing the production systems [49]. Semantic networks can be used to represent hierarchical, categorical, and causal relationships.

Ontologies, on the other hand, are formal representations of a specific domain of knowledge, including the concepts, relationships, and rules that define that domain [49,50]. Ontologies are often used in artificial intelligence and the semantic web to provide a common vocabulary for exchanging information. As an example, Figure 2. depicts an ontology model that describes a production system. This model is part of the implementation presented in publication V. One of the most commonly used ontology representation approaches is the Resource Definition Framework (RDF) where, each axiom is constructed by a subject, predicate and object. As an example, "Lion Are Mammals." In this axiom, "Lions" entity is the subject, "are" entity is the object, and "Mammals" entity is the predicate.

Rule-based systems are systems that represent knowledge as a set of rules or if-then statements. These rules can be used to make inferences and draw conclusions based on the available data. As an example, Toichoa et al. presented a rule-based tool for adapting robot parameters using the human emotional state in [51]. Logical representations are representations that use logical principles and formalisms, such as first-order logic or modal logic, to represent and reason about knowledge. Overall, the choice of knowledge representation depends on the specific problem that needs to be solved and the characteristics of the data that is available.

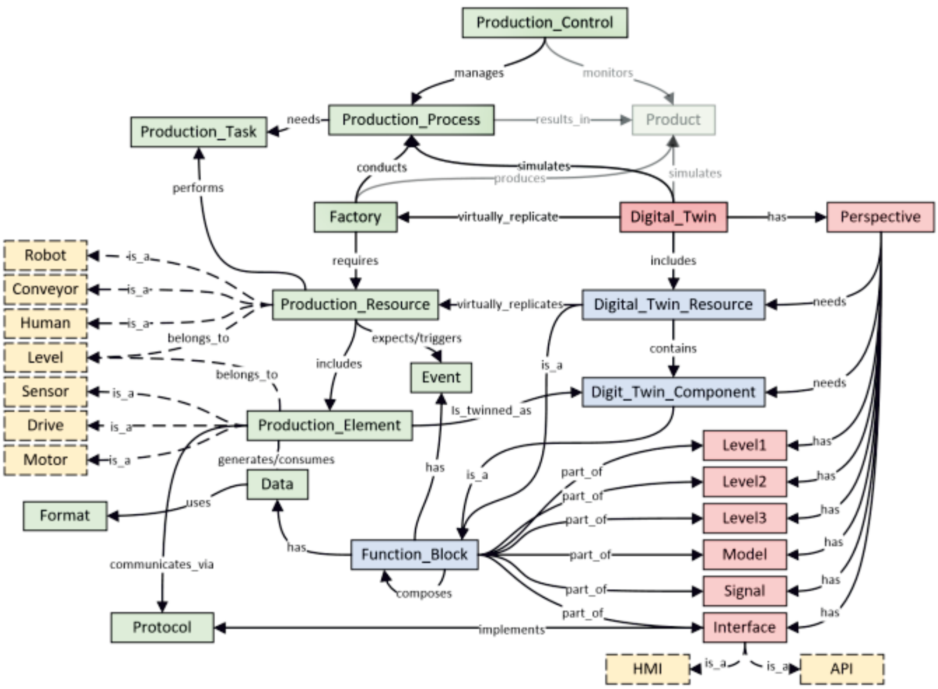


Figure 2. An ontology model that defines digital twin description [4]

2.3.2 Inferencing mechanisms

According to the Oxford dictionary[5], "to infer" means "to reach an opinion or decide that something is true on the basis of information that is available." In the context of knowledge-based systems, inferencing, which is also known as reasoning, refers to the creations of new knowledge from the existing knowledge. In the domain of Knowledge-based systems, Ramis Ferrer and Martinez Lastra in [52] describe the inference as concluding implicit knowledge form explicit knowledge. Reasoning may include different methods. Most commonly used methods are the forward reasoning and backward reasoning. A forward reasoning, also known as forward chaining, is the process of reaching a conclusion by starting in the known facts [53]. Then the reasoning can be narrowed to reach the targeted goal. Meanwhile, the backward reasoning, also known as backward chaining, starts by stating the target gaol as a fact. Then start introducing new goal until reach the known facts. Both of these approaches can be implemented using rule based. One of the common reasoning languages is the Semantic Web Rule Language (SWRL) [14].

2.4 System engineering and system theory

2.4.1 Digital twins

A digital twin is a virtual representation of a physical object or system that is used to simulate and analyse its behaviour and performance. Digital twins can be used in a variety of applications, such as manufacturing, transportation, and construction, to improve efficiency, optimize resource allocation, and reduce risks [54]. In fact, digital twins are built by collecting and understanding the information form the physical world [55,56]. This information is used to create virtual models of the physical system [57]. As a consequence, these models can be used to test and optimize the design of the physical system, predict its performance under different conditions, and monitor its operation in real-time [58,59]. In the literature, Digital Twins demonstrate mirroring, shadowing and threading aspects [60]. The mirroring aspect represents the ability of the digital twin to model the physical system. As presented by Saracco et al. [61], the shadowing aspect represents the ability of the digital twin to be synchronised with the physical system. The threading aspect represents the

[5] https://www.oxfordlearnersdictionaries.com/definition/english/infer?q=infer

ability of the digital twin to evolve and generate knowledge from the twining activity. In fact, these aspects provide a symbiotic feature of the digital twin where both physical and digital systems benefit from the twining activity [62]. In this regard, the term Symbiotic Autonomous Systems (SAS) refers to intelligent and cognitive systems that are able to autonomously exchange knowledge and information to improve their capabilities [63]. In relation to digital twins, the Cognitive Digital Twins (GDT)s addresses the ability of the digital twin to improve it is interaction with the physical system which in return, feeds the physical system with possible improvements.

In manufacturing domain, there are two main digital twins implementations, product digital twins and process digital twins [4]. Product digital twins are virtual representations of physical products, such as aircraft, automobiles, or industrial equipment during its production phase [64]. They can be used to optimize the design, test performance, monitor the operation of these products, tracking and genealogy of the physical products [65]. Process digital twins, on the other hand, are virtual representations of manufacturing processes including its resources, supply chains, and other business processes. They can be used to optimize the flow of materials, reduce waste and downtime, and improve efficiency [66–68]. Overall, Digital Twins have become valuable tools for system developers, integrators and operators as these tools provide monitoring, simulation capabilities. Adding to that, a set of distributed DTs can function as Cyber Physical System of Systems (CPSoS) where each DT can contribute to a holistic model of the physical system [69]. Such a system may increase the robustness of critical systems [70,71]. Moreover, DTs can be projected to standards like ISA 95 to represents their functional modes [4]. As an example, a digital twin can be operating in an online mode which means it is connected to the physical systems and able to collect information to enrich the digital models. Then, the digital twin may operate in an offline mode to simulate the different strategies or situations. This also can be extended if several digital twins are operating simultaneously as aforementioned as a CPSoS.

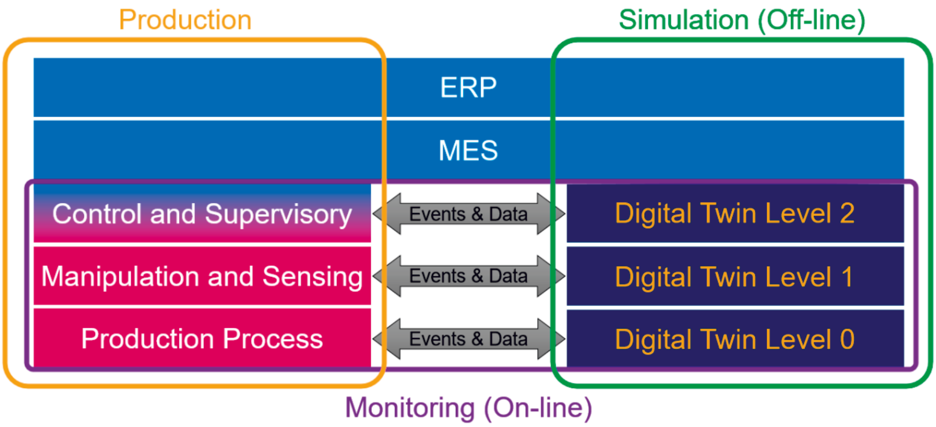


Figure 3. Digital twins' operational modes [4]

Commercially, there are solutions that are available and marketed as digital twins. Some of these solutions are general purpose and some are specific and targeting specific industry of use case. Table 1. presents a summary of the available commercial digital twins based on the research in [4].

Table 1. Summary of Commercially available DTs adopted from [4]

Group	Description	Example
I	This group include cloud-based often referred to as IoT platforms. These platforms mainly operate on a data-centric approach, where users upload data or connect IoT devices that transmit data to the platform. The platform then offers a predefined or customizable set of functions and operations for modelling, analysing, and processing the data. This category is highly versatile and adaptable, catering to a variety of domains for end users.	Amazon, Google, Microsoft, and Oracle cloud systems
II	This group includes digital twins specific to particular vendors. In other words, it comprises companies that offer industrial systems, devices, controllers, and hardware necessitating specialized tools for simulating their operations. Within this group, the simulation of systems is highly precise and detailed. Moreover, it facilitates comprehensive	ABB's RobotStudio, Siemens's TIA Portal, Omron's ACE Software, and various vendors

	visualization for end users to optimize interaction. Nonetheless, DTs in this group are rigid and not suitable for multi-vendor systems.	of robotics and PLCs
III	This group encompasses application-specific Digital Twins, focusing on particular applications or domains. Digital Twins in this group are tailored to address specific problems within a designated domain. Given the targeted nature of this category's digital twins towards specific applications, industrial digital twins are predominantly situated within this classification.	SolidWorks, SAP, ICONICS, Visual Components, and Siemens' Tecnomatix.
IV	This group includes systems able to simulate materials in high realistic presentations. The digital twins in this group utilize their built-in physics engines and modelling capabilities. This group is not commonly employed in industrial applications. This groups are foreseen to be suitable for Human Digital Twins (HDT).	Blender, Autodesk's Maya, NVIDIA Omniverse

2.4.2 Cyber-physical systems

Cyber-physical systems (CPS) are systems that involve the integration of computing and communication capabilities with physical processes and devices [72]. CPS are found in a wide range of applications, including manufacturing, transportation, healthcare, and energy, and can be used to improve efficiency, safety, and reliability [73–75]. The concept of CPS can be extended to the System of Systems (SoS) where several systems that individually provide certain services, can collaborate to create new bigger system that provides new service(s) [76]. Primarily, the characteristics of the CPS concept include integration of computing and communication capabilities with physical processes and devices, real-time operation, distributed operation, and complexity and diversity in technology [77].

Due to its driving economic impact, high competitiveness, and technology incubation, the industrial domain is at the core of the CPS concept where manufacturing systems can be modelled and controlled by cyber systems [78]. As an example, Industrial Cyber-physical Systems (ICPS) can be used for robotic manipulation modelling and controlling as presented in [79,80]. ICPS typically involve the use of sensors and actuators to collect and act on data from the physical industrial world. Moreover, ICPSs involves computing and communication systems to process and transmit this data. They can be used to monitor and control physical processes in real-time, as well as to optimize resource allocation and decision-making.

2.4.3 Data-driven modelling

Data-driven modelling is a type of modelling that uses data to create, test, and refine models of a system or process [81]. It uses AI algorithms for building these models based on the available collected data. Thus, data-driven modelling is commonly used in a variety of fields, including engineering, finance, and biology, to understand and predict the behaviour of complex systems [82].

Building a model using data driven approach requires completing mainly six steps. Firstly, data collection which involves gathering data from a variety of sources, such as sensors, simulations, or experiments [1,83]. Secondly, data cleaning and preparation. This step includes s preparing the data for analysis, which may include tasks such as filtering out noise or missing data, and scaling or normalizing the data [84,85]. Thirdly, model selection which involves selecting a model or algorithm that is appropriate for the problem at hand, based on the characteristics of the data and the desired output. For instance, Cruz et al. presents a novel approach for dynamically selecting the algorithm and the model that suits the available data [86]. Fourthly, model training that involves using the data to train the model, which typically involves adjusting the model's parameters to fit the data as closely as possible. Then, model validation where assessing the model on a separate dataset takes place to ensure that it is able to make accurate predictions or decisions. Finally, model refinement which includes adjusting the model parameters to seek better results or performance. the model or try a different model.

In industrial applications, data-driven approach is commonly used for building models of the manufacturing processes [87]. Depends on the quality of the training data and the training process, the data-driven models can be use in critical systems [88]. In addition, and due to its learning capabilities, data-driven solution can be used for highly complex systems where traditional analytical models are not accurate enough [89].

2.5 Summary

As previously presented in this chapter, developing digital twins for industrial applications requires a multidisciplinary approach where several technologies, sectors, and domains may contribute to the development. A digital twin can bridge fields like computer science, communication technologies, AI, automation, among others. To simplify the presentation, the technologies from these fields can be labelled as central technologies, as they are necessary for developing any digital twin. Then, the application domain, where the digital twin will be used, contributes to the utilization of the digital twin. For example, developing a digital twin for a transportation system requires knowledge of both the central technologies and transportation technologies, methods, and activities. Similarly, and in addition to knowledge of the central technologies, developing a digital twin for industrial automation requires understanding of factory automation, production management, and industrial data spaces.

In the context of this research, developing a digital twin for manufacturing processes combines application technologies such as factory automation, production scheduling, process modelling, and industrial data collection and management with central technologies like CPS, data-driven modelling, and AI, which includes knowledge representation and inferencing mechanisms. Concerning application technologies, and in addition to the topics in sections 2.1 and 2.2, it is important to emphasize the significance of existing manufacturing-related standards and research based on these standards. For instance, ISA-88, ISA-95, IEC 61499, IEC 61131, and ISO 23247. These standards enable the development of the digital twin to follow a systematic approach. On the other hand, regarding central technologies, a digital twin can be seen as a specific implementation of a CPS. In other words, creating a generic approach for developing a digital twin will require knowledge of developing CPSs. Additionally, knowledge representation and inferencing are necessary since one of the objectives involves providing guidelines for digital twin developers. Finally, data-driven technologies are needed because manufacturing processes are complex systems that generate substantial data points and can be predominantly modelled using data-driven approaches.

3 DIGITAL TWINS IN FACTORY AUTOMATION: METHODS, TOOLS, AND TECHNIQUES

3.1 Digital twin definition within the doctoral research

The definition of digital twins has been debated noticeably in the scientific community. While the majority agrees on the original definition of Micheal Grieves which was conceptualized in 2002 [90], the differences are mostly in the implementation part of developing a digital twin. As an example, features like fast forward simulation, 3D visualization, monitoring dashboards, among others are different between the DTs. Within this research, the main focus has been on developing digital twins for industrial processes. This includes production physical resources like machines and equipment, and production control systems and algorithms. In relation to that, this doctoral research presented the outcome of the reference model of a digital twin as multi-layer, multi-level and multi-perspective as depicted in Figure 4. in publication IV [4].

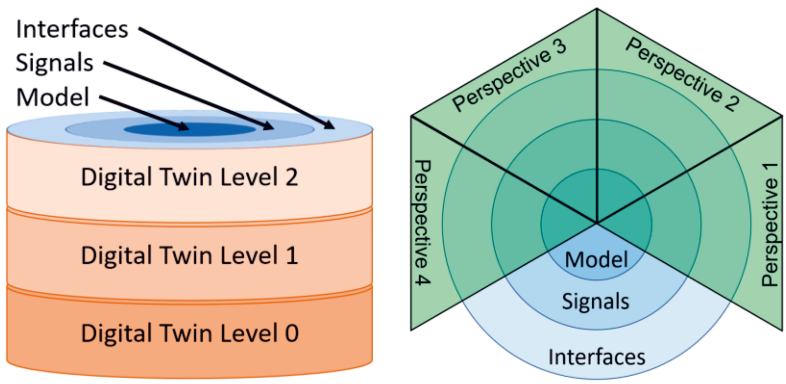


Figure 4. Digital twin reference model as presented in publication IV [4].

The novelty in this reference model lies in the ability of the developer to create multi-perspective of the digital twins that allows different views or insight on the physical system. In this regard, this feature allows the developers to dedicate the views only for the needed functionality. Then, the implementation of the reference model employs the methodology of architecting the digital twin where eighteen reasoning rules are introduced to populate the different levels, layers and perspectives. Such an implementation used the standard IES 61499 to present the digital twin a network of function blocks. To achieve these outcomes, five peer-reviewed journal papers have been published. These papers are listed in the upcoming section.

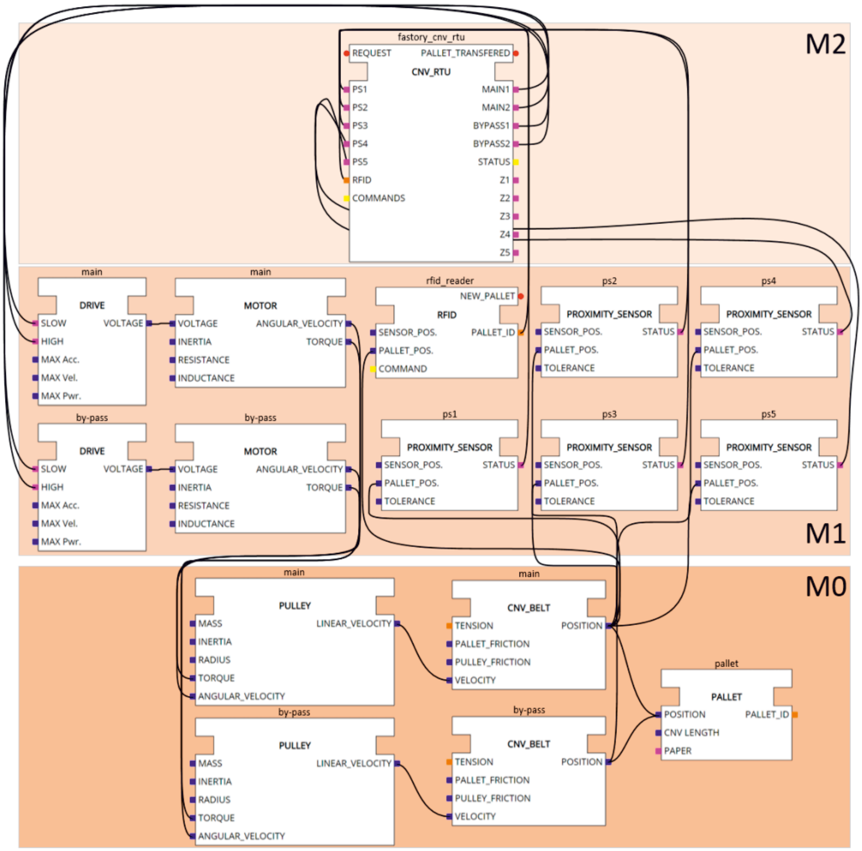


Figure 5. Snippet of the implementation of a digital twin from publication IV [4]

3.2 The refereed publications

This section presents the description and the contribution of the refereed articles in this thesis. As presented before, this thesis is a compendium of 5 peer-reviewed articles. The first article introduced the candidate with large number of resources in the literature to position the digital twin technologies in a scientific landscape. The second article demonstrate the integration of systems with MES and ERP systems as the target is building digital twins for manufacturing processes. Then, the third article illustrates the aspect of modelling information using knowledge-based system and data bases system. The fourth article presents the architecture of the digital twin that has been developed in the doctoral research. Finally, the fifth article presents the synthetically generated data set using digital twin and the possibility for exploiting this data set.

3.2.1 Publication I: Towards 'Lean Industry 4.0′ – Current trends and future perspectives

As a review article, this publication presents a systematic review on the topics of lean production and Industry 4.0. For the topic on lean production, DT technology can be valuable in providing a modelling and simulation platform to the factories to achieve optimum production with the available resources. For the topic on Industry 4.0, DT can be seen as an implementation of the Industry4.0 concepts. For this reason, research on DT technologies requires understanding industry 4.0 concepts. As an example, the interconnectivity between manufacturing resources is crucial for developing the digital replicas in the DT.

Publication I aims at positioning the research in the context of Industry 4.0 and lean manufacturing concepts. In more details, the paper contributes to the research by showing and proofing that process Digital Twins are the main building blocks of the concept of Digital Interconnected Factories (DIF) which is one of the pillars of the industry 4.0 and Lean Manufacturing concepts. Additionally, the publication provides clustering description on relation and contributions of technologies using keywords like cluster 1: Learning factory, lean management, lean, production management, optimization, discrete event simulation, cluster 2: Internet of things, cyber-physical systems, smart manufacturing, cluster 3: Lean manufacturing, smart factory, sustainability, cluster 4: Industry 4.0, digital factory, and cluster 5: Lean production, lean automation.

3.2.2 Publication II: Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems

This publication includes a review of the state of the art on manufacturing execution systems, data interoperability and knowledge-driven manufacturing execution systems. Regarding the manufacturing execution systems, the publication presents different definitions of the MES functions. These definitions are later used in the research to form the structure of the MES solution. Then, this MES solution adopted dynamic methods for using the MES functions in a general manner. The second main discipline is the data interoperability. In this topic, the paper reviews methods, and protocols for machine-machine communication. In addition, it highlights the use of Knowledge-based systems in industry. Finally, the paper reviews the open knowledge-drive MES approach from the eScop EU project. As a contribution to the literature, this paper presents a loosely coupled approach that can be implemented in manufacturing systems. DT technologies can use the finding in this paper to connect and couple with existing MES and ERP systems, which prove to be necessary for DTs to function.

Regarding the doctoral research, this publication contributes to the general understanding on the interrelations between the physical world and the digital world in the industrial applications. In more details, this publication introduces a generic and customizable knowledge-based solution for building applications that encapsulates the MES functionalities. These applications then utilized for controlling the shopfloor operations and processes. Furthermore, the use of web services and knowledge reasoning was essential for the proposed approach. Overall, this publication offers insights into MES, data interoperability, and knowledge-driven manufacturing execution systems, with potential applications in digital twins within the manufacturing industry.

3.2.3 Publication III: Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing

This publication includes a rich review on definitions, methods, and tools for managing database and knowledgebase systems. Firstly, the publication review research work related to data modelling and management including the information life cycle. Then, the paper presents a technology-oriented review on databases and knowledgebase systems. This review also includes techniques required for data transformation and harmonization which is necessary for developing digital twins. As a contribution to the literature, this publication presents a "All what the reader needs to know" about developing systems that will use database or knowledge base systems.

In relation to the doctoral research, data and knowledge management are critically important topic for working on digital twins. This paper presents a direct review of data and knowledge lifecycle in industrial applications. In addition, it covers key aspects such as data modelling, management, technology trends, and data transformation. The contribution of this paper includes understanding the needed technologies and techniques for managing the knowledge in a digital twin. But more importantly, the publication presents a key usage aspect of digital twins where synthetic data can be created form digital twins.

3.2.4 Publication IV: Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications

As this publication represents the core of this thesis, it includes a review on digital twins from three different points of view; scientific, commercial, and standards. Firstly, the publication presents the DT designs and approaches used by researchers and scholars to develop their DTs. This review also include review on how these DT are used. Then, an extensive review on commercial solution that helps for developing and building DT is presented. This commercial solutions' review paves the way on how DTs can be used. Finally, a review on existing standards related to DT is presented. Such review will help in positioning the presented approach for developing and architecting digital twins within the existing standards.

As aforementioned, building Digital Twins for manufacturing processes includes proper understanding on both the central and the application technologies. This understanding may include the importance of developing digital twins, the data and knowledge management in relation to digital twins, the usage of DT with respect to the manufacturing systems. Once these statements are realized, the important question is how to design and develop digital Twins? In other words, what are the proper steps for architecting Digital Twins? This publication contributes to the research by presenting a systemic approach for architecting DTs based on the process's description and the user needs.

3.2.5 Publication V: FASTory digital twin data

The paper presents the creation of data set generated by a Digital Twin. The paper contribution include understanding by example, how Digital Twins can be used for creating synthetic data sets for further analysing and improving manufacturing processes. The publication include access to data sets can be used for training artificial intelligent system or testing optimization methods for discrete manufacturing. These data sets have been created from a digital twin that developed in the work of publication III.

3.3 Summary

This section presents the contribution of the published papers to the doctoral research questions and objectives. This summary is listed in Table 2.

Table 2. Summary of the published paper

#	Publication title	Main results	Contribution to the doctoral research
I	Towards 'Lean Industry 4.0′ – Current trends and future perspectives	– Provides a systematic review on related topics to Industry4.0	RQ1, RQ2, Obj1, Con2.
II	Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems	– Developing a generic method for implementing MES functions in industrial environment	RQ1, RQ3, Obj2, Con1.
III	Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing	– Presenting benchmarking method for managing industrial data – Presenting a benchmarking test for comparing databases and knowledge bases	RQ3, Obj2, Con3.
IV	Ontology-Driven Guidelines for Architecting Digital Twins	– Intensive review on commercially available digital twins	RQ3, RQ4, Obj1, Obj 2,

	in Factory Automation Applications	– A review on standards that are related to developing digital twins. – Reference model for developing digital twins.	Obj3, Con2, Con3
V	FASTory digital twin data	– Synthetically generated dataset from a digital twin	RQ2, Con3.

4 CONCLUSIONS AND RECOMMENDATION FOR FUTURE WORKS

4.1 Conclusion

With the substantial innovation that digital twins' technologies are gaining, more research is expected to be conducted in the coming years. A digital twin is a multi-disciplinary topic which will continuously evolve as long as the dependents technologies evolve. In this work, the focus has been targeted towards developing digital twins for industrial processes. This dissertation presented answers to the original research questions which validates the need by the industry. Regarding RQ1, the presented research show how digital twins are completely aligned with the concept of Lean Production and Industry 4.0. This alignment can be seen in the need for building digital solutions that can interact with the physical systems and provide valuable insights. Furthermore, a digital twin is an essential tool for these concept as they allow simulation and optimization which is a core expected outcome form Lean Production and Industry 4.0.

Regarding RQ2, this doctoral research unravels several features to be considered in developing a digital twin. In this regard, the most important feature for developing a digital twin is the connectivity with the physical system. This includes collecting the data, processing the data and using the data to obtain the system's core model. This core model then can be used for simulation and interaction with the physical system. Secondly, the digital twin must continuously learn the behaviour of the physical system and adapt to changes in the operating conditions. This means that the digital twin will be able to represent all similar physical systems that are operating in different conditions. Other features that are essential for developing digital twins include proper interface with the human to maximize the interaction, expandability and flexibility to allow wider coverage of physical systems and use cases, and clarity and easiness of the usage to help the end users in maximizing the outcome of using the digital twins.

RQ3 has been addressed via the publications III, IV and V. In fact, and as a computer application, the information model of the digital twin is a core part where it affects the usage, the functionality and the performance of the digital twin. More precisely, the research show that using knowledge-based system provides human-like approach for architecting digital twins. Nonetheless, it has been demonstrated that using knowledge reasoning will require additional effort for building these reasoning rule. It is also important to mention that during the research, the lack of tools for building knowledge systems affected the outcome of this research.

Finally, RQ4 targets achieving specific attributes or abilities for digital twins. These abilities are answered clearly in Publication IV where the multi-level, multi-layer and multi perspective approach was presented. These features allow digital twins to be more flexible and versatile to be used in different configuration and environments. Moreover, this allows digital twins to be expanded to different domains than the one in this doctoral research like products digital twins, buildings digital twins and automotive digital twins. A key aspect in answering this research question lies under the usage and reliance on the existing standards. In this regard, using the existing standards elevated the outcome as standards have been already accepted by the scientific and industrial communities.

4.2 Future work

As highlighted before, the topic in this research holds reasonable need in the near future. Thus, this section presents some suggestions that can extend the research work. As highlighted in the conclusion of publication I, the lack of tools and models that support the use of knowledge-based approach affect the development of approach that uses the knowledge-based technologies. This opens research directions to look and develop tools that can manage knowledge-based systems for manufacturing processes but specifically, for developing digital twins.

As this research focused on the industrial processes, an extension to this research may include methodologies for developing digital twin for industrial products and further seeking approaches for interconnecting digital twins for processes and digital twins for products. This research direction has been highlighted in publication IV where the use of Spatial Computation (SC) concept can be further explored.

Looking at the concept of Industry 5.0, a research path can be seen in the adoption of the concept of Human-Centric Design (HCD). This also may lead to interesting research on the development of Human Digital Twins or as also know Human Realistic Avatar (HRA). Overall, the present research can be extended by employing new tools, techniques and methods that help in developing applications and digital twins for industrial applications.

REFERENCES

1. Ferrer, B.R.; Mohammed, W.M.; Chen, E.; Lastra, J.L.M. Connecting Web-Based IoT Devices to a Cloud-Based Manufacturing Platform. In Proceedings of the IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society; October 2017; pp. 8628–8633.

2. Grieves, M. Digital Twin: Manufacturing Excellence through Virtual Factory Replication. **2015**.

3. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inf.* **2019**, *15*, 2405–2415, doi:10.1109/TII.2018.2873186.

4. Mohammed, W.M.; Haber, R.E.; Martinez Lastra, J.L. Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications. *Machines* **2022**, *10*, 861, doi:10.3390/machines10100861.

5. Mohammed, W.M.; Lobov, A.; Ferrer, B.R.; Iarovyi, S.; Martinez Lastra, J.L. A Web-Based Simulator for a Discrete Manufacturing System. In Proceedings of the IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society; IEEE: Florence, October 2016; pp. 6583–6589.

6. Attaran, M.; Celik, B.G. Digital Twin: Benefits, Use Cases, Challenges, and Opportunities. *Decision Analytics Journal* **2023**, *6*, 100165, doi:10.1016/j.dajour.2023.100165.

7. Lago Alvarez, A.; Mohammed, W.M.; Vu, T.; Ahmadi, S.; Martinez Lastra, J.L. Enhancing Digital Twins of Semi-Automatic Production Lines by Digitizing Operator Skills. *Applied Sciences* **2023**, *13*, 1637, doi:10.3390/app13031637.

8. Castano, F.; Haber, R.E.; Mohammed, W.M.; Nejman, M.; Villalonga, A.; Lastra, J.L.M. Quality Monitoring of Complex Manufacturing Systems on the Basis of Model Driven Approach. *Smart Structures and Systems* **2020**, *26*, 495–506, doi:10.12989/SSS.2020.26.4.495.

9. International Organization for Standardization (ISO) ISO 23247 Automation Systems and Integration — Digital Twin Framework for Manufacturing 2021.

10. International Organization for Standardization (ISO) ISO/TR 24464:2020 Automation Systems and Integration — Industrial Data — Visualization Elements of Digital Twins 2020.

11. International Organization for Standardization (ISO) ISO/IEC AWI 30172 Digital Twin — Use Cases 2020.

12. International Organization for Standardization (ISO) ISO/IEC AWI 30173 Digital Twin — Concepts and Terminology 2020.

13. Lundberg, K.H.; Barton, T.W. History of Inverted-Pendulum Systems. *IFAC Proceedings Volumes* **2010**, *42*, 131–135, doi:10.3182/20091021-3-JP-2009.00025.

14. Ramis Ferrer, B.; Mohammed, W.M.; Ahmad, M.; Iarovyi, S.; Zhang, J.; Harrison, R.; Martinez Lastra, J.L. Comparing Ontologies and Databases: A Critical Review of Lifecycle Engineering Models in Manufacturing. *Knowl Inf Syst* **2021**, *63*, 1271–1304, doi:10.1007/s10115-021-01558-4.

15. Elahi, M.; Afolaranmi, S.O.; Mohammed, W.M.; Martinez Lastra, J.L. FASTory Assembly Line Power Consumption Data. *Data in Brief* **2023**, *48*, 109160, doi:10.1016/j.dib.2023.109160.

16. Mohammed, W.M.; Martinez Lastra, J.L. FASTory Digital Twin Data. *Data in Brief* **2021**, *35*, 106912, doi:10.1016/j.dib.2021.106912.

17. Elahi, M.; Afolaranmi, S.O.; Mohammed, W.M.; Martinez Lastra, J.L. Energy-Based Prognostics for Gradual Loss of Conveyor Belt Tension in Discrete Manufacturing Systems. *Energies* **2022**, *15*, 4705, doi:10.3390/en15134705.

18. Mohammed, W.M.; Ramis Ferrer, B.; Iarovyi, S.; Negri, E.; Fumagalli, L.; Lobov, A.; Martinez Lastra, J.L. Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems. *International Journal of Computer Integrated Manufacturing* **2018**, *31*, 262–274, doi:10.1080/0951192X.2017.1407874.

19. Lastra, J.L.M.; Delamer, M. Semantic Web Services in Factory Automation: Fundamental Insights and Research Roadmap. *IEEE Transactions on Industrial Informatics* **2006**, *2*, 1–11.

20. Luo, H.; Yin, S.; Gao, T.; Kaynak, O. Adaptive Configuration Technique for Decentralized Plug-and-Play Process Monitoring System. In Proceedings of the

IEEE EUROCON 2017 -17th International Conference on Smart Technologies; July 2017; pp. 479–484.

21. Iarovyi, S.; Mohammed, W.M.; Lobov, A.; Ferrer, B.R.; Lastra, J.L.M. Cyber–Physical Systems for Open-Knowledge-Driven Manufacturing Execution Systems. *Proc. IEEE* **2016**, *104*, 1142–1154, doi:10.1109/JPROC.2015.2509498.

22. Sauter, T. The Continuing Evolution of Integration in Manufacturing Automation. *IEEE Industrial Electronics Magazine* **2007**, *1*, 10–19, doi:10.1109/MIE.2007.357183.

23. United Nation No Harm Rule | UNEP Law and Environment Assistance Platform Available online: https://leap.unep.org/knowledge/glossary/no-harm-rule (accessed on 9 September 2023).

24. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, Conception and Perception. *Journal of Manufacturing Systems* **2021**, *61*, 530–535, doi:10.1016/j.jmsy.2021.10.006.

25. Mohammed, W.M.; Nejman, M.; Castano, F.; Martinez Lastra, J.L.; Strzelczak, S.; Villalonga, A. Training an Under-Actuated Gripper for Grasping Shallow Objects Using Reinforcement Learning. In Proceedings of the 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS); IEEE: Tampere, Finland, June 10 2020; pp. 493–498.

26. Malvido Fresnillo, P.; Vasudevan, S.; Mohammed, W.M.; Martinez Lastra, J.L.; Perez Garcia, J.A. Extending the Motion Planning Framework—MoveIt with Advanced Manipulation Functions for Industrial Applications. *Robotics and Computer-Integrated Manufacturing* **2023**, *83*, 102559, doi:10.1016/j.rcim.2023.102559.

27. Zhang, P.; Liang, H.; Yue, C. Technological Anxiety: How Robots Impact College Graduates' Informal Employment? *Journal of Asian Public Policy* **2023**, *0*, 1–17, doi:10.1080/17516234.2023.2170308.

28. Bejarano, R.; Ferrer, B.R.; Mohammed, W.M.; Martinez Lastra, J.L. Implementing a Human-Robot Collaborative Assembly Workstation. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN); IEEE: Helsinki, Finland, July 2019; pp. 557–564.

29. Rautiainen, S.; Pantano, M.; Traganos, K.; Ahmadi, S.; Saenz, J.; Mohammed, W.M.; Martinez Lastra, J.L. Multimodal Interface for Human–Robot Collaboration. *Machines* **2022**, *10*, 957, doi:10.3390/machines10100957.

30. De Miguel Lazaro, O.; Mohammed, W.M.; Ferrer, B.R.; Bejarano, R.; Martinez Lastra, J.L. An Approach for Adapting a Cobot Workstation to Human Operator within a Deep Learning Camera. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN); IEEE: Helsinki, Finland, July 2019; pp. 789–794.

31. Hormaza, L.A.; Mohammed, W.M.; Ferrer, B.R.; Bejarano, R.; Martinez Lastra, J.L. On-Line Training and Monitoring of Robot Tasks through Virtual Reality. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN); IEEE: Helsinki, Finland, July 2019; pp. 841–846.

32. Katasonov, A.; Lastusilta, T.; Korvola, T.; Saari, L.; Bendas, D.; Mohammed, W.M.; Lee, A.N. An Approach to Production Scheduling Optimization a Case of an Oil Lubrication and Hydraulic Systems Manufacturer. In Proceedings of the 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC); IEEE: Funchal, June 2017; pp. 1123–1130.

33. Mohammed, W.M.; Ferrer, B.R.; Iftikhar, U.; Lastra, J.L.M.; Simarro, J.H. Supporting a Cloud Platform with Streams of Factory Shop Floor Data in the Context of the Intustry 4.0. In Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); IEEE: Porto, Portugal, July 2018; pp. 786–791.

34. Marques, M.; Cunha, A.; Mohammed, W.M.; Jardim-Gonçalves, R.; Agostinho, C. IoT-Based Automatic Non-Conformity Detection: A Metalworking SME Use Case. In *Enterprise Interoperability VIII*; Popplewell, K., Thoben, K.-D., Knothe, T., Poler, R., Eds.; Proceedings of the I-ESA Conferences; Springer International Publishing: Cham, 2019; Vol. 9, pp. 155–165 ISBN 978-3-030-13692-5.

35. Ferrer, B.R.; Mohammed, W.M.; Martinez Lastra, J.L.; Villalonga, A.; Beruvides, G.; Castano, F.; Haber, R.E. Towards the Adoption of Cyber-Physical Systems of Systems Paradigm in Smart Manufacturing Environments. In Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); IEEE: Porto, July 2018; pp. 792–799.

36. Enyoghasi, C.; Badurdeen, F. Industry 4.0 for Sustainable Manufacturing: Opportunities at the Product, Process, and System Levels. *Resources, Conservation and Recycling* **2021**, *166*, 105362, doi:10.1016/j.resconrec.2020.105362.

37. Mikalef, P.; Krogstie, J.; Pappas, I.O.; Pavlou, P. Exploring the Relationship between Big Data Analytics Capability and Competitive Performance: The Mediating Roles of Dynamic and Operational Capabilities. *Information & Management* **2020**, *57*, 103169, doi:10.1016/j.im.2019.05.004.

38. Delamer, I.M.; Lastra, J.L.M. Service-Oriented Architecture for Distributed Publish/Subscribe Middleware in Electronics Production. *IEEE Transactions on Industrial Informatics* **2006**, *2*, 281–294, doi:10.1109/TII.2006.885188.

39. Okwudire, C.E.; Madhyastha, H.V. Distributed Manufacturing for and by the Masses. *Science* **2021**, *372*, 341–342, doi:10.1126/science.abg4924.

40. Phuyal, S.; Bista, D.; Bista, R. Challenges, Opportunities and Future Directions of Smart Manufacturing: A State of Art Review. *Sustainable Futures* **2020**, *2*, 100023, doi:10.1016/j.sftr.2020.100023.

41. Yu, Q.; Yang, H.; Lin, K.-Y.; Li, L. A Self-Organized Approach for Scheduling Semiconductor Manufacturing Systems. *J Intell Manuf* **2021**, *32*, 689–706, doi:10.1007/s10845-020-01678-8.

42. Kovalevskaya, D.; Pedersen, A.-C.; Holmen, E.; Kaloudis, A.; Ringen, G. Triads in Lean Management: Analyzing Buyer–Supplier-Supplier and Buyer–Supplier-Supplier's Supplier Relationships for Zero-Defect Manufacturing. *J Knowl Econ* **2022**, doi:10.1007/s13132-022-01094-8.

43. Ejsmont, K.; Gladysz, B.; Corti, D.; Castaño, F.; Mohammed, W.M.; Martinez Lastra, J.L. Towards 'Lean Industry 4.0′ – Current Trends and Future Perspectives. *Cogent Business & Management* **2020**, *7*, 1781995, doi:10.1080/23311975.2020.1781995.

44. Tiwari, P.; Sadeghi, J.K.; Eseonu, C. A Sustainable Lean Production Framework with a Case Implementation: Practice-Based View Theory. *Journal of Cleaner Production* **2020**, *277*, 123078, doi:10.1016/j.jclepro.2020.123078.

45. Yang, T.; Wen, Y.-F.; Hsieh, Z.-R.; Zhang, J. A Lean Production System Design for Semiconductor Crystal-Ingot Pulling Manufacturing Using Hybrid Taguchi Method and Simulation Optimization. *Assembly Automation* **2020**, *40*, 433–445, doi:10.1108/AA-11-2018-0193.

46. Barni, A.; Pietraroia, D.; Züst, S.; West, S.; Stoll, O. Digital Twin Based Optimization of a Manufacturing Execution System to Handle High Degrees of

Customer Specifications. *Journal of Manufacturing and Materials Processing* **2020**, *4*, 109, doi:10.3390/jmmp4040109.

47. Ferrer, B.R.; Mohammed, W.M.; Lobov, A.; Galera, A.M.; Lastra, J.L.M. Including Human Tasks as Semantic Resources in Manufacturing Ontology Models. In Proceedings of the IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society; IEEE: Beijing, October 2017; pp. 3466–3473.

48. Li, X.; Zhang, S.; Huang, R.; Huang, B.; Xu, C.; Zhang, Y. A Survey of Knowledge Representation Methods and Applications in Machining Process Planning. *Int J Adv Manuf Technol* **2018**, *98*, 3041–3059, doi:10.1007/s00170-018-2433-8.

49. Ferrer, B.R.; Mohammed, W.M.; L., J.; Lastra, M.; Strzelczak, S. A Semantic Workbench for Editing, Querying, Navigating and Distributing Ontologies for Cognitive Manufacturing. In Proceedings of the IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society; IEEE: Lisbon, Portugal, October 2019; pp. 2767–2772.

50. Seyedamir, A.; Ferrer, B.R.; Martinez Lastra, J.L. An ISA-95 Based Ontology for Manufacturing Systems Knowledge Description Extended with Semantic Rules. In Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); July 2018; pp. 374–380.

51. Toichoa Eyam, A.; Mohammed, W.M.; Martinez Lastra, J.L. Emotion-Driven Analysis and Control of Human-Robot Interactions in Collaborative Applications. *Sensors* **2021**, *21*, 4626, doi:10.3390/s21144626.

52. Ferrer, B.R.; Lastra, J.L.M. Towards the Encapsulation and Decentralisation of OKD-MES Services within Embedded Devices. *International Journal of Production Research* **2018**, *56*, 1286–1298, doi:10.1080/00207543.2017.1328141.

53. Jain, S.; Mehla, S.; Wagner, J. Chapter 9 - Ontology-Supported Rule-Based Reasoning for Emergency Management. In *Web Semantics*; Jain, S., Jain, V., Balas, V.E., Eds.; Academic Press, 2021; pp. 117–128 ISBN 978-0-12-822468-7.

54. Cheng, K.; Wang, Q.; Yang, D.; Dai, Q.; Wang, M. Digital-Twins-Driven Semi-Physical Simulation for Testing and Evaluation of Industrial Software in a Smart Manufacturing System. *Machines* **2022**, *10*, 388, doi:10.3390/machines10050388.

55. Ferrer, B.R.; Mohammed, W.M.; Martinez Lastra, J.L. A Solution for Processing Supply Chain Events within Ontology-Based Descriptions. In Proceedings of the IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society; IEEE: Florence, October 2016; pp. 4877–4883.

56. Qureshi, K.A.; Mohammed, W.M.; Ferrer, B.R.; Martinez Lastra, J.L.; Agostinho, C. Legacy Systems Interactions with the Supply Chain through the C2NET Cloud-Based Platform. In Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN); IEEE: Emden, July 2017; pp. 725–731.

57. Afolaranmi, S.O.; Ferrer, B.R.; Mohammed, W.M.; Martinez Lastra, J.L.; Ahmad, M.; Harrison, R. Providing an Access Control Layer to Web-Based Applications for the Industrial Domain. In Proceedings of the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN); IEEE: Emden, July 2017; pp. 1096–1102.

58. Muhammad, U.; Ferrer, B.R.; Mohammed, W.M.; Lastra, J.L.M. An Approach for Implementing Key Performance Indicators of a Discrete Manufacturing Simulator Based on the ISO 22400 Standard. In Proceedings of the 2018 IEEE Industrial Cyber-Physical Systems (ICPS); IEEE: St. Petersburg, May 2018; pp. 629–636.

59. Ramis Ferrer, B.; Muhammad, U.; Mohammed, W.; Martínez Lastra, J. Implementing and Visualizing ISO 22400 Key Performance Indicators for Monitoring Discrete Manufacturing Systems. *Machines* **2018**, *6*, 39, doi:10.3390/machines6030039.

60. Singh, M.; Fuenmayor, E.; Hinchy, E.; Qiao, Y.; Murray, N.; Devine, D. Digital Twin: Origin to Future. *ASI* **2021**, *4*, 36, doi:10.3390/asi4020036.

61. Saracco, R.; Grise, K.; Martinez, T. The Winding Path towards Symbiotic Autonomous Systems. *Phil. Trans. R. Soc. A.* **2021**, *379*, 20200361, doi:10.1098/rsta.2020.0361.

62. Kaynak, O.; He, W.; Flammini, F.; Liu, Z. Towards Symbiotic Autonomous Systems. *Phil. Trans. R. Soc. A.* **2021**, *379*, 20200359, doi:10.1098/rsta.2020.0359.

63. Wang, Y.; Karray, F.; Kwong, S.; Plataniotis, K.N.; Leung, H.; Hou, M.; Tunstel, E.; Rudas, I.J.; Trajkovic, L.; Kaynak, O.; et al. On the Philosophical, Cognitive and Mathematical Foundations of Symbiotic Autonomous Systems. *Phil. Trans. R. Soc. A.* **2021**, *379*, 20200362, doi:10.1098/rsta.2020.0362.

64. Lo, C.K.; Chen, C.H.; Zhong, R.Y. A Review of Digital Twin in Product Design and Development. *Advanced Engineering Informatics* **2021**, *48*, 101297, doi:10.1016/j.aei.2021.101297.

65. Mohammed, W.M. Encapsulation Of MES Functionalities As RESTful Web Services For Knowledge-Driven Manufacturing Systems. Master's Thesis, Tampere University of Technology: Tampere, Finland, 2017.

66. Ma, J.; Chen, H.; Zhang, Y.; Guo, H.; Ren, Y.; Mo, R.; Liu, L. A Digital Twin-Driven Production Management System for Production Workshop. *Int J Adv Manuf Technol* **2020**, *110*, 1385–1397, doi:10.1007/s00170-020-05977-5.

67. Errandonea, I.; Beltrán, S.; Arrizabalaga, S. Digital Twin for Maintenance: A Literature Review. *Computers in Industry* **2020**, *123*, 103316, doi:10.1016/j.compind.2020.103316.

68. Nasirahmadi, A.; Hensel, O. Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm. *Sensors* **2022**, *22*, 498, doi:10.3390/s22020498.

69. Biesinger, F.; Meike, D.; Kraß, B.; Weyrich, M. A Digital Twin for Production Planning Based on Cyber-Physical Systems: A Case Study for a Cyber-Physical System-Based Creation of a Digital Twin. *Procedia CIRP* **2019**, *79*, 355–360, doi:10.1016/j.procir.2019.02.087.

70. Stary, C. Digital Twin Generation: Re-Conceptualizing Agent Systems for Behavior-Centered Cyber-Physical System Development. *Sensors* **2021**, *21*, 1096, doi:10.3390/s21041096.

71. Lee, J.; Azamfar, M.; Singh, J.; Siahpour, S. Integration of Digital Twin and Deep Learning in Cyber-Physical Systems: Towards Smart Manufacturing. *IET Collaborative Intelligent Manufacturing* **2020**, *2*, 34–36, doi:10.1049/iet-cim.2020.0009.

72. Colombo, A.W. *Industrial Cloud-Based Cyber-Physical Systems: The IMC-AESOP Approach*; Springer International Publ: Cham, 2014; ISBN 978-3-319-05624-1.

73. Mohammed, W.M.; Malvido Fresnillo, P.; Vasudevan, S.; Gosar, Z.; Martinez Lastra, J.L. An Approach for Modeling Grasping Configuration Using Ontology-Based Taxonomy. In Proceedings of the 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS); IEEE: Tampere, Finland, June 10 2020; pp. 507–513.

74. Haber Guerra, R.E.; Cembrano Gennari, G.; Villagrá Serrano, J.; Jiménez Ruiz, A.R.; Margarit Taulé, J.M.; Fonseca Chácharo, L.; Gómez Álvarez-Arenas, T.E.; Camacho Sosa Dias, J.; Aparicio, S.; Carmona Galán, R.; et al. *Challenge 4: Cyberphysical Systems and Internet of Things*; Consejo Superior de Investigaciones Científicas (CSIC), 2021; pp. 76–89; ISBN 978-84-00-10756-7.

75. Lee, E.A. Cyber Physical Systems: Design Challenges. In Proceedings of the 2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC); IEEE: Orlando, FL, USA, May 2008; pp. 363–369.

76. Rüssmeier, N.; Lamm, A.; Hahn, A. A Generic Testbed for Simulation and Physical-Based Testing of Maritime Cyber-Physical System of Systems. *J. Phys.: Conf. Ser.* **2019**, *1357*, 012025, doi:10.1088/1742-6596/1357/1/012025.

77. Verma, R. Smart City Healthcare Cyber Physical System: Characteristics, Technologies and Challenges. *Wireless Pers Commun* **2022**, *122*, 1413–1433, doi:10.1007/s11277-021-08955-6.

78. Pivoto, D.G.S.; de Almeida, L.F.F.; da Rosa Righi, R.; Rodrigues, J.J.P.C.; Lugli, A.B.; Alberti, A.M. Cyber-Physical Systems Architectures for Industrial Internet of Things Applications in Industry 4.0: A Literature Review. *Journal of Manufacturing Systems* **2021**, *58*, 176–192, doi:10.1016/j.jmsy.2020.11.017.

79. Fresnillo, P.M.; Vasudevan, S.; Mohammed, W.M. An Approach for the Bimanual Manipulation of a Deformable Linear Object Using a Dual-Arm Industrial Robot: Cable Routing Use Case. In Proceedings of the 2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS); IEEE: Coventry, United Kingdom, May 24 2022; pp. 1–8.

80. Fresnillo, P.M.; Vasudevan, S.; Mohammed, W.M.; Martinez Lastra, J.L.; Laudante, G.; Pirozzi, S.; Galassi, K.; Palli, G. Deformable Objects Grasping and Shape Detection with Tactile Fingers and Industrial Grippers. In Proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS); IEEE: Victoria, BC, Canada, May 10 2021; pp. 525–530.

81. Berberich, J.; Köhler, J.; Müller, M.A.; Allgöwer, F. Data-Driven Model Predictive Control With Stability and Robustness Guarantees. *IEEE Transactions on Automatic Control* **2021**, *66*, 1702–1717, doi:10.1109/TAC.2020.3000182.

82. Shah, A.; Belyaev, P.; Ferrer, B.R.; Mohammed, W.M.; Martinez Lastra, J.L. Processing Mobility Traces for Activity Recognition in Smart Cities. In

Proceedings of the IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society; IEEE: Beijing, October 2017; pp. 8654–8661.

83. Iftikhar, U.; Mohammed, W.M.; Ferrer, B.R.; Lastra, J.L.M. A Framework for Data Collection, Transformation and Processing in Industrial Systems. In Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); IEEE: Porto, July 2018; pp. 707–712.

84. Mohammed, W.M.; Ferrer, B.R.; Martinez, J.L.; Sanchis, R.; Andres, B.; Agostinho, C. A Multi-Agent Approach for Processing Industrial Enterprise Data. In Proceedings of the 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC); IEEE: Funchal, June 2017; pp. 1209–1215.

85. Mohammed, W.M.; Ferrer, B.R.; Martinez Lastra, J.L.; Aleixo, D.; Agostinho, C. Configuring and Visualizing the Data Resources in a Cloud-Based Data Collection Framework. In Proceedings of the 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC); IEEE: Funchal, June 2017; pp. 1201–1208.

86. Cruz, Y.J.; Rivas, M.; Quiza, R.; Haber, R.E.; Castaño, F.; Villalonga, A. A Two-Step Machine Learning Approach for Dynamic Model Selection: A Case Study on a Micro Milling Process. *Computers in Industry* **2022**, *143*, 103764, doi:10.1016/j.compind.2022.103764.

87. Castano, F.; Cruz, Y.J.; Villalonga, A.; Haber, R.E. Data-Driven Insights on Time-to-Failure of Electromechanical Manufacturing Devices: A Procedure and Case Study. *IEEE Transactions on Industrial Informatics* **2023**, *19*, 7190–7200, doi:10.1109/TII.2022.3216629.

88. Jiang, Y.; Yin, S.; Kaynak, O. Data-Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond. *IEEE Access* **2018**, *6*, 47374–47384, doi:10.1109/ACCESS.2018.2866403.

89. Rossit, D.A.; Tohmé, F.; Frutos, M. A Data-Driven Scheduling Approach to Smart Manufacturing. *Journal of Industrial Information Integration* **2019**, *15*, 69–79, doi:10.1016/j.jii.2019.04.003.

90. Grieves, M.; Vickers, J. Origins of the Digital Twin Concept 2016.

PUBLICATIONS

This part of the thesis presents a reproduction of the original publications related to the doctoral research. As it can be seen, the reproduction of each publication has been requested to the corresponding organisation.

PUBLICATION
I

Towards 'Lean Industry 4.0′ – Current trends and future perspectives

Krzysztof Ejsmont, Bartlomiej Gladysz, Donatella Corti, Fernando Castaño, Wael M Mohammed, Jose L Martinez Lastra

Cogent Business & Management, 2020, 7 (1)
Doi: 10.1080/23311975.2020.1781995

Taylor & Francis
Taylor & Francis Group

cogent business & management

Cogent Business & Management

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/oabm20

Towards 'Lean Industry 4.0′ – Current trends and future perspectives

Krzysztof Ejsmont , Bartlomiej Gladysz , Donatella Corti , Fernando Castaño , Wael M. Mohammed & Jose L. Martinez Lastra |

Ejsmont et al., *Cogent Business & Management* (2020), 7: 1781995
https://doi.org/10.1080/23311975.2020.1781995

cogent business & management

OPERATIONS, INFORMATION & TECHNOLOGY | REVIEW ARTICLE

Towards 'Lean Industry 4.0' – Current trends and future perspectives

Krzysztof Ejsmont[1]*, Bartlomiej Gladysz[1], Donatella Corti[2], Fernando Castaño[3], Wael M. Mohammed[4] and Jose L. Martinez Lastra[4,5]

Received: 06 March 2020
Accepted: 05 June 2020

*Corresponding author: Krzysztof Ejsmont, Institute of Production Systems Organization, Faculty of Production Engineering, Warsaw University of Technology, Warsaw 02-524, Poland
E-mail: krzysztof.ejsmont@pw.edu.pl

Reviewing editor:
Pantea Foroudi, Middlesex University, UK

Additional information is available at the end of the article

Abstract: The enterprises to be competitive are constantly looking for continuous increase of productivity, quality and level of services. With the development of Industry 4.0 concept, manufacturers are more confident about new advantages of automation and systems integration. Lean management is well developed and empirically proven effective managerial approach. Combining Lean and Industry 4.0 practices seems to be necessary evolutionary step for further raise the level of operational excellence (exploitation of finance, workload, materials, machines/devices). There is an increasing number of Industry 4.0 solutions used to reduce waste (as known from Lean Management). Therefore, the main objective of this article aims at presenting the results of a literature review on the concept of 'Lean Industry 4.0'. Dynamic methodology called "Systematic Literature Network Analysis (SLNA)" was used. It combines the Systematic Literature Review approach with the quantitative analysis of bibliographic networks to detect emerging topics and the dynamic evolution of the topics. The paper is a comprehensive systematization and rationalization of knowledge about the integration of LM and I4.0 concepts, identifies the most important research trends and defines directions for future research. The article contains a framework that presents the current state of knowledge in the area of Lean Industry 4.0.

Krzysztof Ejsmont

ABOUT THE AUTHOR

Krzysztof Ejsmont PhD in Management, is an assistant professor in Faculty of Production Engineering at Warsaw University of Technology (WUT), Poland. His research is focused on smart manufacturing and Industry 4.0 from production/operations management perspective. He participated in few national and international R&D projects, including funded by EU (NanoForCE, e-Scop). Currently he is the administrator of global research project on "Industry 4.0 in Production and Aeronautical Engineering – IPAE" (10 partners from Europe, USA, Japan, Australia and South Korea).
The other co-authors are researchers who participate in the IPAE project. They are all scientifically related to Industry 4.0 issues and have extensive experience in scientific activity and in the implementation of R&D projects.

PUBLIC INTEREST STATEMENT

Industry 4.0 refers to a series of changes in the way products and services are delivered. Thanks to modern technologies, it enables quick collection and analysis of data between machines, faster and more flexible response to occurring problems, but also more efficient processes to produce high quality products at lower costs. The basis of lean management is the elimination of waste while taking into account the role of the employee in creating the value of manufactured products and services provided. Many companies are interested in implementing both concepts. The paper is a comprehensive review of the combination of these two concepts (called 'Lean Industry 4.0') and presents what has been done so far in this area. The article may be interesting to both theoreticians and practitioners, as it indicates the possibilities offered by the use of integrated Lean Industry 4.0 solutions to eliminate problems related to manufacturing and logistics.

Subjects: Operations Management; Lean Manufacturing; Production Systems & Automation; Engineering Management

Keywords: bibliometrics; Industry 4.0; lean management; lean manufacturing; systematic literature network analysis

1. Introduction

The ability to produce customized products has become the basis for being competitive in a dynamic, globalized and digitally connected world. Customers are used to receiving products and services specially tailored to their needs. The growing expectations of customers along with the progressing quality requirements have led to a growing number of products in portfolios and indirectly influenced the increase in the complexity of the production environment (Westkämper et al., 2013).

Therefore, companies are looking for concepts that can reduce complexity in the industrial area, as well as contribute to increasing value and reducing all types of wastes. Two of the most popular concepts used for this purpose are: Lean Management (LM) and Industry 4.0 (I4.0 or I4).

The cornerstone of the Lean Management philosophy is to reduce waste in the value chain to reduce total lead time including all operations (also those non-value adding). It is also important to focus on the client's value in the process of continuous improvement, as well as considering the role of the employee in creating the value of products and services provided (Schuh, 2013; Womack et al., 1990).

The basis of Industry 4.0 is the ability to quickly collect, process, analyze and exchange large data sets between machines. Thanks to modern technologies such as: Cyber-Physical Systems (CPS) or Internet of Things (IoT), it is possible to react faster and more flexibly to existing problems, but also to more efficient value creation processes, while reducing costs (Oztemel & Gursev, 2018).

Both concepts seem to be helpful in solving the problems facing modern production. Therefore, the research questions seems reasonable: (1) If and how can both concepts complement each other? and (2) Which Industry 4.0 technologies can support specific lean principles/tools. Therefore, the article aims to thoroughly examine what has already been described in the literature on this issue.

The motivation to take up this topic is relatively little research on the importance of the relationship between LM and Industry 4.0 (Buer, Strandhagen et al., 2018; Kolberg et al., 2017; Sanders et al., 2016, 2017; Tortorella & Fettermann, 2018). There is no comprehensive framework that connects LM and Industry 4.0 (Kolberg & Zuehlke, 2015; Leyh et al., 2017). The need to develop a framework for the integration of LM and Industry 4.0 is indicated in the literature (Sanders et al., 2016). In the paper (Buer, Strandhagen et al., 2018) it was found that the area of LM and Industry 4.0 is still immature, and therefore the framework of integration of LM and Industry 4.0 has not yet been published. The need to understand how these concepts interact is also suggested. Despite the existence of papers that combines these two approaches, there is no comprehensive systematization of existing knowledge in this area. Organizing available knowledge, presenting current research streams, as well as indicating which areas require further research will help in developing a holistic framework of integration of LM and Industry 4.0.

The scientific thesis of the article is as follows: systematizing knowledge on the combining the LM and I4.0 concepts will facilitate the development of a framework for their integration and will be a starting point for further research.

Hence, the aim of the paper is to present the landscape of scientific literature on the concept of Lean Industry 4.0 (LI4) using the method of dynamic literature review called "Systematic

Literature Network Analysis (SLNA)" (Ciano et al., 2019; Colicchia et al., 2012; Strozzi et al., 2017). This methodology combines a systematic literature review and bibliographic network analysis using modern bibliometric tools. Adopting such an approach seems to be adequate to the subject being studied, given its novelty and interdisciplinary nature. The proposed approach has been positively verified for other contexts, e.g., creating an academic landscape of sustainability science (Kajikawa et al., 2007), review on how this field of ethical sourcing research has grown and evolved over the decades, providing implications for future research (Kim et al., 2018). In these works, SLNA confirmed its potential value in identifying trends, evolutionary trajectories and key issues that affect the development of knowledge in a given field in a more scientific and objective way compared to traditional descriptive reviews. Traditional reviews are mainly based on content analyzes that do not cover the evolutionary aspect of the direction of publication and are based on subjective criteria for the selection of articles and the classification of research input into pre-defined coding schemes. Instead, SLNA relies on objective measurements and algorithms to quantify emerging topics based on available literature.

2. Theoretical background

2.1. Lean management

Toyota began optimizing operations by continuous eliminating all kinds of waste in 1949. Taiichi Ohno, the founder of the Toyota Production System (TPS) has developed a set of synchronized methods and principles for controlling production plants, which became the basis of the Lean philosophy. According to him, the essence of Lean is to reduce the time from the customer's order to the final receipt of the product, by eliminating all activities that are considered waste and do not add value to the customer (Ohno, 1988). The first books on the theory of Lean Management (LM) were published in English in 1978 and gained special recognition in the automotive sector. Over the past few decades, many articles and books have been published that focus on the description and characterization of LM content (Tortorella & Fettermann, 2018). A large number of authors consider LM to be the most well-known management paradigm of recent times (Holweg, 2007; Womack et al., 1990).

Currently, the Lean concept is seen as "a set of management principles and techniques aimed at eliminating waste in the production process and increasing the flow of activities that, from the point of view of customers, increase the value of the product" (Taj, 2008; Womack & Jones, 1996b).

In the literature, lean management is translated through various principles, guidelines or rules. Based on TPS values, five general principles can be distinguished (Womack & Jones, 1996a; Womack et al., 1990):

(1) specify the value desired by the customer,
(2) identify a value stream for each product/service providing value to the client; all waste in the value stream can be questioned,
(3) ensure that the product flow is continuous,
(4) introduce the pull principle—provide services on order,
(5) strive for perfection through continuous improvement (kaizen).

The main idea of Lean is to eliminate all kinds of wastes (muda). Eight main types of waste have been identified in the literature (Liker, 2004; Ohno, 1988): (1) transport (2) inventory (3) motion (4) waiting (5) over-processing (6) overproduction (7) defects (8) skills.

The researchers (Sony, 2018) argue that the focus should be not only on the elimination of these 8 wastes, but also on the other two waste-generating elements: mura and muri. Mura refers to process variability and processes should be standardized to reduce it. Muri means excessive work load that can be prevented by creating ergonomic and safe working conditions. The three main types of LM activities are (1) evaluation, (2) improvement, and (3) performance monitoring.

LM is supported by a set of well-known tools for the operationalization of its goals, both at the strategic and operational level, and the basic philosophy treats human as the most important issue in all activities (Varela et al., 2019). The most popular LM methods, tools and techniques include, among others (Chiarini, 2011): Value Stream Mapping (VSM), 5 S, Total Productive Maintenance (TPM), Single Minutes Exchange of Die (SMED), Kanban, Poka-Yoke, Just-in-time (JIT), Hoshin Kanri, Takt time, Jidoka, Heijunka.

2.2. Industry 4.0

The beginnings of the use of information and communication technologies (ICT) in the manufacturing industry took place in the 1970 s. However, the term 'Industry 4.0′ was used for the first time in 2011 at the Hannover in Germany. The main ideas of Industry 4.0 were published in the same year (Kagermann et al., 2011), and also became a strategic initiative of the German government and was included in the "High-Tech Strategy 2020 Action Plan" (Kagermann et al., 2013). Similar strategies have also been implemented in other industrialized countries, e.g., USA ("Advanced Manufacturing Partnership"), China ('Made in China 2025′), United Kingdom ("Smart Factory") and others (Kumar et al., 2020).

Over time, the concept of Industry 4.0 has become synonymous with the fourth industrial revolution (Buer, Strandhagen et al., 2018). Kolberg and Zuehlke (2015) present Industry 4.0 as a further development of Computer Integrated Manufacturing (CIM) and thus as a network approach that complements CIM through ICT. This approach is supported by the integration of automation technologies, e.g., cyber-physical systems (CPS), collaborative robots, cloud computing and big data sets, with the production environment via Internet of Things (IoT) (Xu et al., 2018). Chukwuekwe et al. (2016) suggest the existence of key Industry 4.0 technologies such as cloud computing, 3D printing, CPS, IoT, Internet of Services (IoS) and big data. Embedded systems, semantic machine-machine communication, CPS and IoT enable connection of the physical and virtual world which is the main goal of Industry 4.0 (Xu et al., 2018). This gives the opportunity to connect the entire factory into a network, creating an intelligent environment. Digitally developed smart machines, storage systems and production facilities enable comprehensive integration based on ICT systems throughout the entire supply chain, from inbound logistics to production, marketing, outbound logistics and services (Kagermann et al., 2013).

From the manufacturing point of view, Industry 4.0 is understood as the movement of intelligent objects that independently coordinate their paths in the factory. Machines are able to implement these paths and communicate in real time with the appropriate warehouse. Information is used primarily to evaluate and control current processes (Kaufmann, 2015). Industry 4.0 significantly affects the manufacturing environment, resulting in radical changes in the implementation of operations. Unlike conventional forecast-based production planning, Industry 4.0 enables real-time production planning, along with dynamic self-realization (Sanders et al., 2016).

Despite the fact that Industry 4.0 is one of the most discussed topics among practitioners and academic teachers in the past few years, no single, commonly accepted definition of this concept has been developed (Buer, Strandhagen et al., 2018; Mrugalska & Wyrwicka, 2017). Researchers and practitioners have divided opinions on which elements create Industry 4.0, how these elements are interrelated and where Industry 4.0 applies. Studies available in the literature show over 100 different definitions of Industry 4.0 (Moeuf et al., 2017). According to the authors, the definition that well reflects the idea of Industry 4.0 is that proposed by Leyh et al. (2017): "Industry 4.0 describes the transition from centralized production towards production that is very flexible and self-controlled. Within this production, the products, all affected systems, and all of the process steps of the engineering, are digitized and interconnected to share and pass information and to distribute this information along the vertical and horizontal value chains and beyond in extensive value networks."

It can therefore be assumed that Industry 4.0 is a strategy to compete in the future. In the paper (Mrugalska & Wyrwicka, 2017) can be found that Industry 4.0 focuses on optimizing value chains due to autonomously controlled and dynamic production, and enables the creation of flexible manufacturing and logistics systems.

2.3. Linking lean management and Industry 4.0

The relationship between LM and Industry 4.0 is increasingly emphasized in operations management research (Mourtzis et al., 2017; Sartal & Vázquez, 2017). Despite the significant differences between LM and I4.0, both concepts have the same goal—to increase added value (Prinz et al., 2018). On scientific and practical grounds, it is therefore reasonable to conduct research aimed at answering the question: can both approaches be combined, and if so, how?

Ohno (1988) described two pillars needed to support TPS: just-in-time (JIT) and autonomy (Jidoka). These pillars are also important for LM (Bicheno & Holweg, 2009). JIT can be supported by a digital supply chain (Zelbst et al., 2014), and autonomy can be increased by CPS (Thoben et al., 2014). Therefore, some researchers state that lean implementation is a prerequisite for successful transformation of Industry 4.0 (Kaspar & Schneider, 2015; Staufen, 2016).

Despite various indications in studies that examine the relationship between LM and I4.0, there is still a lack of empirical evidence to support their findings. Buer, Strandhagen et al. (2018) emphasized that the literature on LM and Industry 4.0 is not consistent as to their integration. In addition, they argue that it is necessary to examine the impact of the combination of LM and I4.0 on the results of organizations and the impact of external factors on the relationship between these two concepts. Over the past few years, scientists and practitioners have conducted research on how both approaches, when implemented together in companies, can raise operational and financial results to a higher level (Kolberg & Zuehlke, 2015; Mrugalska & Wyrwicka, 2017; Sanders et al., 2016; Tortorella & Fettermann, 2018).

The growth potential due to LI4.0 is quantified in some papers (Prinz et al., 2018), but it is very difficult to verify the validity of these estimates, as no long-term studies on the effects of I4.0 have been conducted. So despite the fact that the integration of LM and I4.0 was the subject of many theoretical works and practical experiments—it is reasonable to conduct further research to better understand this relationship (Leyh et al., 2017).

The most important publications on the integration of LM and I4.0 along with their characteristics are presented in Table 1.

Based on Table 1, it can be concluded that despite many papers on the subject of LM and I4.0 integration, many of them are theoretical studies regarding only selected aspects of combining both concepts. There is no publication that would be a comprehensive review of the literature, which thanks to the use of modern bibliometric tools will organize all previous research efforts in many areas related to LM and I4.0. There is also no full picture of the current state of literature that will show what has been done so far and what should be the subject of future research.

3. Materials and methods

Two databases were considered for the literature analysis: Web of Science Core Collection (WoSCC) and Scopus. The reason was that they are the most-used databases when it comes to citations for field delineation (Strozzi et al., 2017). WoSCC and Scopus are also the leading databases of scholarly

Table 1. Papers about the integration of LM and I4.0

Paper	Title	Type of paper	Description of the content
(Ejsmont & Gladysz, 2020)	Lean Industry 4.0—Wastes versus Technology Framework	Literature review	Explanation how I4.0 solutions can eliminate 8 LM wastes; indication of potential support for lean techniques through I4.0
(Kamble et al., 2019)	Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies	Survey	Determining the impact of I4.0 technologies on LM practices and sustainable organizational performance
(Rossini et al., 2019)	The interrelation between Industry 4.0 and lean production: an empirical study on European manufacturers	Survey	Examine the impact of the relationship between the adoption of I4.0 technologies and the implementation of lean production (LP) practices at the level of improving the operational performance of manufacturers
(Varela et al., 2019)	Evaluation of the Relation between Lean Manufacturing, Industry 4.0, and Sustainability	Literature review/survey	Establishing relations and measuring the effects of LM and I4.0 in terms of sustainability
(Buer, Strandhagen et al., 2018)	The link between Industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda	Literature review	Identification of main research streams regarding the relationship between LM and I4.0 and a future research program
(Mayr et al., 2018)	Lean 4.0—A conceptual conjunction of lean management and Industry 4.0	Conceptual/use case	Considerations on the complementarity of LM and I4.0; examples of how I4.0 technologies can support specific lean methods
(Sony, 2018)	Industry 4.0 and lean management: a proposed integration model and research propositions	Conceptual	Theoretical model of integration of I4.0 and LM
(Tortorella & Fettermann, 2018)	Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies	Survey	Examine the relationship between (LP) practices and the implementation of I4.0; the impact of LP and I4.0 on companies performance
(Leyh et al., 2017)	Industry 4.0 and Lean Production—A Matching Relationship? An analysis of selected Industry 4.0 models	Conceptual	Analysis of I4.0 models/framework in the context of LP
(Mrugalska & Wyrwicka, 2017)	Towards Lean Production in Industry 4.0	Conceptual/case study	Case studies of LP and I4.0 integration
(Sanders et al., 2016)	Industry 4.0 Implies Lean Manufacturing: Research Activities in Industry 4.0 Function as Enablers for Lean Manufacturing	Literature review	Indication of which I4.0 technologies are able to support particular dimensions of LM
(Kolberg & Zuehlke, 2015)	Lean Automation enabled by Industry 4.0 Technologies	Conceptual/use case	Theoretical and practical examples of combining I4.0 and lean production

impact and are characterized by high quality of reported journals (Powell & Peterson, 2017). Due to restrictive indexing procedures in databases, their content is considered to be of high quality.

Scopus is therefore very similar to WoSCC. It has some advantages and disadvantages. The main advantages are that in Scopus there are almost 60% more records than in WoSCC (Zhao & Strotmann, 2015) and the database also contains in-press articles. The main disadvantage is that the available data are not as clean as the WoSCC data. This means that some documents are not clearly identified and can be treated as different nodes in the resulting citation network. The study (Wang & Waltman, 2016) has stated that WoSCC classifies journals more accurately than Scopus.

WoSCC is better than Scopus when we want to find more accurate citation information (Crew et al., 2016) and identify "high-influence" publications (Tabacaru, 2019). For this reason, it was decided to choose the WoSCC database.

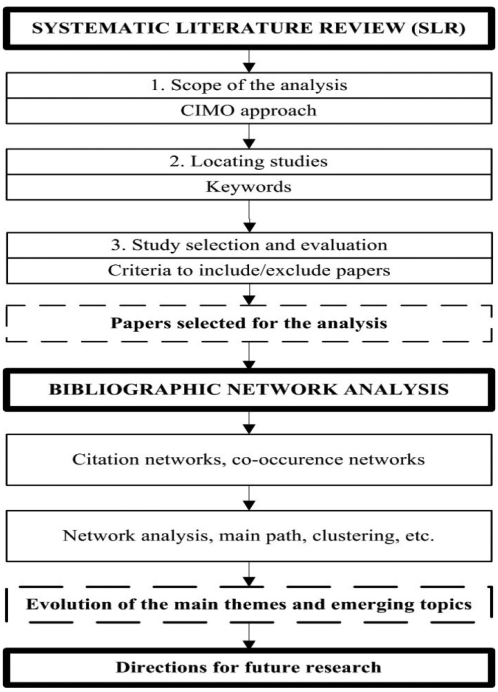
Systematic Literature Network Analysis (SLNA) is the procedure chosen for the selection and analysis of articles (Figure 1).

The first stage is a Systematic Literature Review (SLR). The scope of the study is identified by three steps:

- Scope of the analysis;

In order to formulate the research question and conduct a proper literature review, Denyer and Tranfield (2009) proposed to answer the questions related to Context, Intervention, Mechanism

Figure 1. Systematic Literature Network Analysis (SLNA) based on (Strozzi **et al.**, 2017**).**



and Outcome (CIMO); The other possible approach is the systematic literature review strategy proposed by Levy and Ellis (2006) i.e. choose, know, understand, apply, examine, combine and evaluate;

- Identifying studies “keywords, time, type of documents, language”;
- Study selection and evaluation;

The result of this stage will be a set of selected documents.

The second stage of the SLNA methodology is the analysis and visualization of the bibliographic network. In the paper, the main attention will be devoted to the citation network and the keywords network, in accordance with the research methodology presented in Figure 1.

To conduct the analysis, it was decided to use 2 software applications, i.e. VOSviewer (https://www.vosviewer.com), CiteSpace (http://cluster.cis.drexel.edu/~cchen/citespace/).

The VOSviewer program (Van Eck & Waltman, 2010) is especially useful when working with a multi-element data set. The software is based on the approach to grouping and bibliometric mapping of the network, introduced by Waltman et al. (2010). It allows visualization of similarities (VOS) in several forms (network, overlay, density). The software, after selecting the type of analysis and counting method, as well as providing the minimum number of thresholds, e.g. keywords or citations, allows to create a coexistence network. The VOSviewer was used to create a citation network, to designate global and local citation and for creating co-occurrence network of author’s keywords.

CiteSpace is an application for visualizing and analyzing trends and patterns in scientific literature. It was designed as a tool for progressive visualization of fields of knowledge (Chen, 2004). CiteSpace focuses on finding critical points in the development of a field or domain, especially intellectual turning points and key points. Detailed case studies are provided in (Chen, 2006) and other papers. The application supports structural and temporal analyzes of various networks from scientific studies (e.g., keywords). CiteSpace has a burst detection algorithm that makes it possible to detect series on keywords. The output of an algorithm is a list of popular keywords over time (topic bursts). The burst weight represents the size of the change in the keyword frequency. The CiteSpace was used to conduct a burst detection analysis. The primary source of input for application is Web of Science.

4. Systematic literature review

4.1. Scope of the analysis and identifying studies

This paper explores the concept of ‘Lean Industry 4.0′. This concept assumes the integration of Lean and Industry 4.0 mainly through modern technologies enabling the reduction of 8 major wastes.

The selection of keywords used in the construction of the searching query was carried out as follows: various terms and abbreviations related to the words “Lean” and ‘Industry 4.0′ have been identified in the literature.

Since the word “lean” was defined in literature in many ways, synonyms and abbreviations of that word had to be established. This was to correctly select all documents related to the Lean concept. Krafcik (1988) first used the word “lean”, which referred to “Toyota Production System”—TPS. It is worth noting that TPS is a major precursor of the more generic “Lean Manufacturing” (Womack et al., 1990). Taiichi Ohno (1988) and Eiji Toyoda, Japanese industrial engineers, developed the system between 1948 and 1975. Lean Production principles such as Just-In-Time (JIT) and other quality management philosophies such as Total Quality Management (TQM) or Six Sigma, have become substitute terms among scientists and practitioners (Delbridge & Oliver, 1991). Given the large number of synonyms and abbreviations, four literature reviews (Ciano et al., 2019; Hasle et al., 2012; Jasti & Kodali, 2015; Martínez-Jurado & Moyano-Fuentes, 2014) suggest the following keywords: “Lean”, “LM”, “LP”, “Just-in-time”, “JIT”, “Toyota Production System”, “TPS”, “Total Quality Management”, “TQM”, “Six Sigma”. Looking at the approaches and tools in lean, two main key concepts are value and waste (Chroneer & Wallstrom, 2016). Since the main essence of lean philosophy is to eliminate all kinds of wastes (muda)—“waste” was also recognized as the keyword. As for ‘Industry 4.0′, there is no term in the literature that could be considered a synonym for it. As other names/abbreviations we can find: “4.0 Industry”, ‘Industrie 4.0′, ‘I4.0′, ‘I4′, “Fourth Industrial Revolution”, “4th Industrial Revolution”. Although the term ‘Industry 4.0′ appeared for the first time in 2011, a full description of this concept has been published in 2013 (Kagermann et al., 2013).

The query was formulated as follows:

< (“Lean” OR “LM” OR “LP” OR “Just-in-time” OR “JIT” OR “Toyota Production System” OR “TPS” OR “Total Quality Management” OR “TQM” OR “Six Sigma” OR “waste”) AND (‘Industry 4.0′ OR ‘Industrie 4.0′ OR “4.0 Industry” OR ‘I4.0′ OR ‘I4′ OR “Fourth Industrial Revolution” OR “4th Industrial Revolution”) >.

This step in the analysis is very important, because the results may change if other keywords are used. Considering the novelty of the examined issue, the selection of keywords refers to the combination of the Lean and Industry 4.0 concepts and finding their proper synonyms/abbreviations. The main purpose was therefore to extract from the literature the most used terms. To achieve the desired result, it is important to use various tools to extract information from a set of studies. Many terms closely related to Lean (e.g., types of losses) and Industry 4.0 (e.g., Cyber-Physical Systems, Internet of Things) were not included in the set of selected keywords, because this would significantly reduce the number of papers found. This selection was made in accordance with the purpose of this article, i.e. presenting the landscape of scientific literature on the concept of LI4. The selected set of keywords allows the appearance of specific concepts and related issues and trends using the adopted methodology and its bibliographic analysis tool.

4.2. Study selection and evaluation

The identified searching query was entered in the “Topic” field in WoSCC at the beginning of January 2020 (02.01.2020). 243 documents were identified. The aim was select papers about LI4, which was the main purpose of their analysis. Only documents published in English were considered. Papers dated 2011–2019 were investigated. Authors also considered reference lists of found articles from important references missed in the database search.

The rationale for selecting this time period is because the term ‘Industry 4.0′ was used for the first time in 2011, and the basic concept of the fourth industrial revolution was then described.

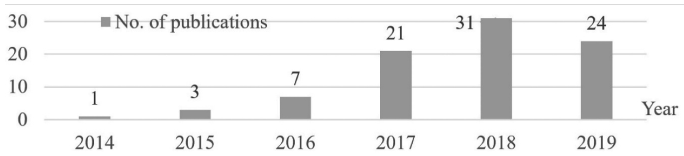
87 relevant studies were selected after screening abstracts and keywords of found papers. In case of doubt, after reading the abstract and familiarizing with the keywords, does the article concern the relationship between Lean and Industry 4.0—the full text has been read. Only publications with full versions in the WoSCC database were taken for further analysis.

4.3. Structure of the papers selected for the analysis

Interest in the topic LI4 over the years 2014–2019 (the oldest selected paper was published in 2014) is presented in Figure 2.

The largest part of selected papers was proceedings papers (51.72%) and articles (44.83%). Book chapters, reviews and early accesses constituted only 6.90% of the entire selected group (2.30%

Figure 2. Number of papers in the field of Lean Industry 4.0.



each). It is worth to mention that further research should approach also white papers as many technology providers tend to publish in that way. However, the quality of such papers is not peer-reviewed. Therefore, they were excluded as they are incomparable in their nature with scientific publications disseminated through channels of conferences and journals.

Figure 2 shows that in the period 2014–2016—the interest in the topic was not significant. In 2017, there was a clear increase in interest on the LI4 concept (21 papers). In 2018 another significant increase was recorded—close to 50% compared to 2017 (31).

Such a dynamic growth in interest in this subject emphasizes its importance and relevance. There were 24 papers identified in 2019, but the search was done in the beginning of January 2020 and probably not all publications from 2019 were indexed in the WoSCC database.

Research areas to which selected papers relate are presented in Figure 3.

The selected studies are dominated by one research area: engineering (64 papers). This proves that in most publications, the authors focus on aspects related to production, in which the combination of the concept of Lean and Industry 4.0 is based on engineering solutions for improving manufacturing. The next most numerous research areas related to the analyzed issue are computer science (13) and operations research, management science (12). This may indicate that in these studies the focus is on the development of IT tools/algorithms using the technologies of Industry 4.0, improving lean manufacturing. Other research areas are less numerous in terms of papers and cover many different fields, which may indicate the interdisciplinary nature of the LI4 issue.

As shown in Figure 4, most papers were created in Germany (29.89%), Italy (16.09%) and Poland (9.20%). Other studies constitute a small percentage of the total. Europe's dominance in this topic

Figure 3. Research areas of selected papers.

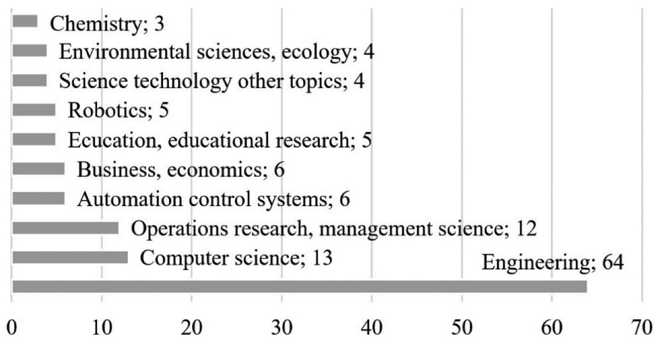
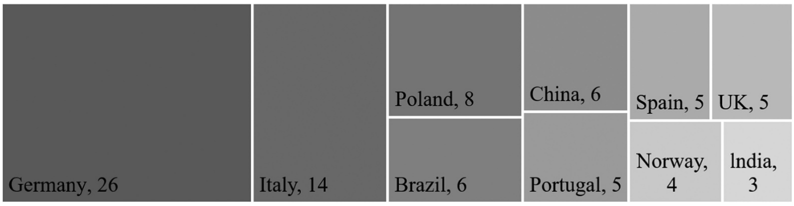


Figure 4. Countries/regions of selected papers.



can be clearly seen, which can be explained by the fact that Industry 4.0 was founded in Germany and this concept is very popular especially on the European continent.

Due to the type of paper, conceptual studies pre-dominate (38). Then there are literature reviews (18) and case studies (11). The least numerous are empirical researches (11) and surveys (9). Therefore, theoretical papers dominantly outweigh the practical ones.

5. Bibliographic network analysis

5.1. Citation network analysis (CNA)

A citation network is a network where the nodes are papers and links mean that there are citations between them. Thanks to this, we can observe the flow of knowledge, as well as which works are linked with citations. This, in turn, makes it possible to isolate clusters (smaller networks), which include papers in which each must have at least one connection with another within the cluster. This enables, among others easier definition of the thematic scope of the cluster.

Figure 5 shows a citation network based on selected papers, using an overlay visualization form. As a result, it became possible to observe which publications were characterized by the largest number of links with others (weights) in the entire network (87 articles). At the same time, the total number of citations in the WoSCC database was presented using a color scale.

Because CNA is a citation-based method, publications that do not have a single link (40 papers) are excluded from the analysis because they are unrelated. In fact, citation analysis can only be

Figure 5. Citation network.

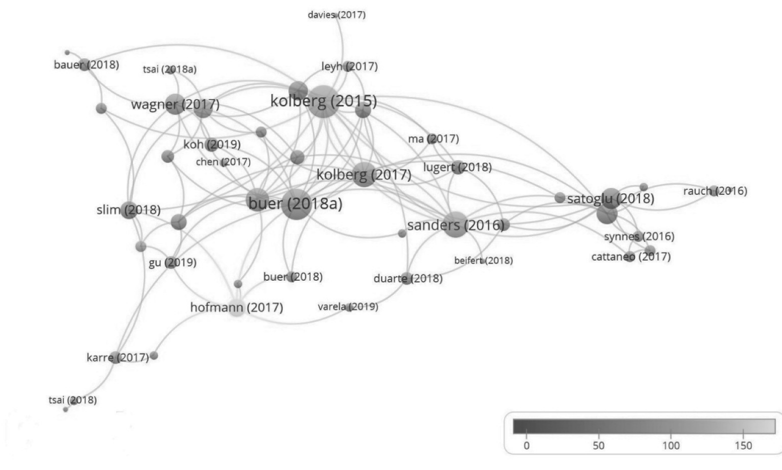


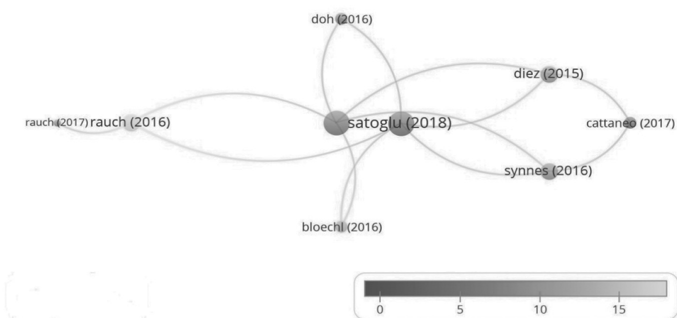
Table 2. Structure of the two largest clusters

Cluster 1		
Paper	Links within the cluster 1	Links within the citation network
(Satoglu et al., 2018a)	5	9
(Satoglu et al., 2018b)	5	9
(Rauch et al., 2016)	3	3
(Synnes & Welo, 2016)	3	3
(Diez et al., 2015)	3	3
(Cattaneo et al., 2017)	2	3
(Bloechl & Schneider, 2016)	2	3
(Doh et al., 2016)	2	2
(Rauch et al., 2017)	1	1
Paper	Links within the cluster 2	Links within the citation network
(Gu et al., 2019)	4	4
(Hofmann & Ruesch, 2017)	4	7
(Ciano et al., 2019)	3	6
(Tortorella & Fettermann, 2018)	3	10
(Yin et al., 2018)	3	3
(Slim et al., 2018)	2	7
(Fettermann et al., 2018)	2	2
(Buer, Fragapane et al., 2018)	1	3

applied to linked papers. A network constructed in this way consists of 47 nodes and 117 links (Figure 5). CNA gives the best results, when clusters consist of a large number of nodes because the amount of information that can be extracted is much larger than information from small clusters (Strozzi et al., 2014). Based on these assumptions, only two largest clusters were analyzed (Table 2).

It should be noted, that not all articles included in a given cluster must be closely related to a particular topic. This is due to the fact that sometimes the authors cite papers that practically do not relate to the topic discussed in the article or the main theme of the cluster (Rauch et al., 2016; Synnes & Welo, 2016). This may also be the case in literature review articles (Cattaneo et al., 2017; Ciano et al., 2019; Slim et al., 2018). Therefore, when using modern bibliometric tools, the traditional review of papers should not be overlooked.

Figure 6. Cluster 1—links and number of citations.



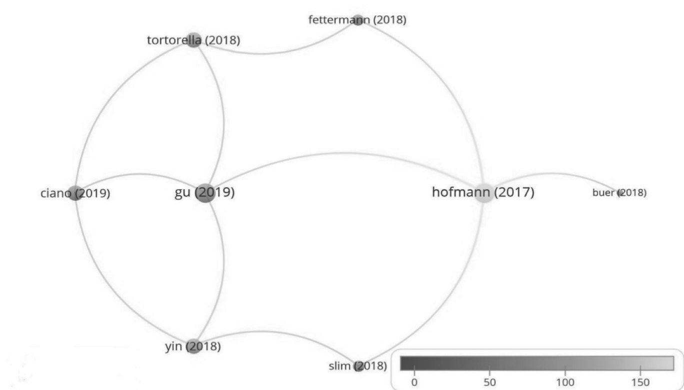
5.1.1. Cluster 1 CNA: Smart lean transformation using Industry 4.0 solutions

First cluster (Figure 6) is consisted of nine publications. Those publications may be labelled together as related to the smart lean transformation enabled through Industry 4.0 tools and technologies and vice versa. Two papers (Satoglu et al., 2018a, 2018b) address relation between particular (subjectively chosen by the authors) Industry 4.0 technologies and lean manufacturing tools or wastes. Authors discussed how specific technologies can support specific tools or lead to elimination of specific waste. However, those papers lack details and cover broad scope of technologies, but each only partially. Authors proved that lean and I4.0 are not mutually exclusive, by showing examples of their successful co-existence. Even though, this proof is very general as no real details and effects are discussed. The second sub-cluster within this cluster is related to product and process development issues (Rauch et al., 2016, 2017; Synnes & Welo, 2016). Authors discussed how the principles of lean product development could be achieved and supported by I4.0 technologies. Again, the choices seem to be subjective. Paper (Rauch et al., 2017) is more specific and it is focused on specifics of critical factors for lean product development in Italian small and medium enterprises. The third sub-cluster consists of papers which discuss some lean approaches and their support for coping with I4.0 challenges and implementation or I4.0 approaches to cope with lean challenges or the mix of both mentioned. It includes Hoshin Kanri (Diez et al., 2015), simulation game for lean and intelligent production logistics (learning factory) (Bloechl & Schneider, 2016), integration of information systems and technologies in the stages of the chain value manufacturing (Doh et al., 2016), lean thinking principles implementation in the context of smart factory (Cattaneo et al., 2017).

5.1.2. Cluster 2 CNA: The impact of Industry 4.0 on the improvement (main context: reduction of wastes) in production systems and logistics

The second cluster (Figure 7) consists of 8 papers. Those papers were related to Industry 4.0 technologies contribution to improvements of operations, production and logistics management. They did not directly indicate specific frameworks or methodologies of improvements (like lean management/manufacturing in case of cluster 1). It is worth to mention that this cluster is consisted of extensive journal papers (while cluster 1 included many conference papers). Four papers were from one journal (International Journal of Production Research) and may be therefore considered as a sub-cluster. One paper discussed how lean was addressed by the mentioned journal (bibliometric analysis) (Ciano et al., 2019). Second paper discussed a case of I4.0 and lean production implementation in the context of Brazilian companies (Tortorella & Fettermann, 2018), so it is somehow similar in approach to the paper (Rauch et al., 2017) (the context of Italian SMEs). Third paper proposed an integrated architecture for implementing extended

Figure 7. Cluster 2—links and number of citations.



producer responsibility in the context of Industry 4.0 (Gu et al., 2019). Fourth paper discussed the evolution from Industry 2.0 to the Industry 4.0 (Yin et al., 2018). Three of four papers (Buer, Fragapane et al., 2018; Fettermann et al., 2018; Hofmann & Ruesch, 2017) left in the cluster should be considered as related to general issues of the impact of the use of Industry 4.0 technologies on logistics, operations management and continuous improvement. The last paper discussed convergence and contradictions of lean and I4.0 for inventive design of smart production systems.

Analyzing topical areas and the scope of papers included in both discussed clusters, it is worth to mention that papers within cluster 2 are rather loosely topically connected, and therefore CNA does not provide in their case solid base for clustering. However, cluster 1 is mainly focused of relations between lean features like techniques, methods, set of wastes and their possible support by the applications of I4.0 technologies. Some of the papers also tackle the problems of vice versa relation, i.e. how lean approach may be used when implementing I4.0 technologies.

5.2. Global and local citation score analysis

Global citation score (GCS) analysis can be used to detect groundbreaking publications. GCS shows the total number of citations in the WoSCC database. Studies with high GCS are considered seminal or have a significant impact on the area of knowledge to which they relate (Knoke & Yang, 2008). In other words, GCS allows the identification of papers that form the basis of a given field, which are often used by other authors to develop their publications. Citations are counted from the entire WoSCC database, even if they are from articles that have not been identified or selected.

Table 3 presents the 10 most frequently cited papers ranked by their GCS. Their local citation score (LCS) is also given, which shows the number of citations the publication obtained in the citation network of selected studies. By comparing GCS and LCS, groundbreaking papers can be identified that has received a small number of citations within the citation network, but has a significant number of citations in the entire WoSCC database.

According to Table 3, among the 10 most frequently cited papers, six studies do not belong to any of the two largest clusters analyzed. Table 3 confirms that some papers belonging to the two largest clusters are groundbreaking, and not just a lot of citations. In addition, 6 publications not belonging to any of the largest clusters confirm the main topics that are currently the objects of research of scientists (e.g., Industry 4.0 supporting lean automation, the impact of Industry 4.0 on production systems).

Table 3. GCS and LCS of the 10 most cited papers

Rank	Paper	GCS	LCS	Part of one of the two largest clusters
1	(Hofmann & Ruesch, 2017)	163	7	Yes
2	(Kolberg & Zuehlke, 2015)	89	19	No
3	(Sanders et al., 2016)	61	13	No
4	(Mrugalska & Wyrwicka, 2017)	44	8	No
5	(Yin et al., 2018)	43	3	Yes
6	(Tortorella & Fettermann, 2018)	37	11	Yes
7	(Kolberg et al., 2017)	34	12	No
8	(Wagner et al., 2017)	27	9	No
9	(Buer, Strandhagen et al., 2018)	24	17	Yes
10	(Bonilla et al., 2018)	21	0	No

Table 4. Ranking of the 10 most cited papers in 2019

Rank	Paper	GCS	No. of citations in 2019	Citations per year since publication
1	(Hofmann & Ruesch, 2017)	163	100	33.33
2	(Yin et al., 2018)	43	34	17
3	(Tortorella & Fettermann, 2018)	37	30	15
4	(Buer, Strandhagen et al., 2018)	24	20	10
5	(Gu et al., 2019)	10	10	10
6	(Bonilla et al., 2018)	21	19	9.5
7	(Mrugalska & Wyrwicka, 2017)	44	26	8.66
8	(Sanders et al., 2016)	61	32	8
9	(Wagner et al., 2017)	27	20	6.66
10	(Kolberg et al., 2017)	34	17	5.66

In order to identify recent groundbreaking studies that could have a potentially large impact and promising scientific input on LI4, papers were ranked according to the number of citations received in the entire WoSCC database in 2019, divided by the number of years since the year of publication. This allowed the identification of those studies that have (potentially) low GCS, but have recently gained considerable interest from the scientific community. In fact, this process "weighed" citations received in 2019 based on the "lifespan" of papers. The ranking of papers elaborated in this way is shown in Table 4. Thanks to this, it became possible to identify one article (Gu et al., 2019) that was not included in the previous ranking (see Table 3). It is also important to notice that the high GCS value does not always mean that the study have a large impact and promising scientific input on LI4.

The papers presented in Table 4 suggest that the latest breakthrough literature is heading towards topics related to the integration of Industry 4.0 with lean automation and lean production (Kolberg et al., 2017; Mrugalska & Wyrwicka, 2017; Tortorella & Fettermann, 2018; Wagner et al., 2017), the use of modern technologies (CPS, IoT, sensors) in production processes and logistics (Gu et al., 2019; Hofmann & Ruesch, 2017), integration of Lean and Industry 4.0 in the context of improving manufacturing systems (Buer, Strandhagen et al., 2018; Sanders et al., 2016; Yin et al., 2018). There is also a study presenting the impact of Industry 4.0 on sustainable development in the context of reducing wastes (Bonilla et al., 2018).

It is worth noting that paper (Gu et al., 2019), which, despite being published in 2019, already has 10 citations and belongs to the second biggest cluster. Usually, papers get more citations in the years following publication. In this case, it may indicate that it could be a breakthrough study, setting further directions of research. The paper proposes an integrated architecture to achieve effective and efficient extended producer responsibility (EPR) using Industry 4.0 technologies. The authors promote the sharing of information, sustainability and reduction of wastes.

5.3. Co-occurrence network of author's keywords

The minimum number of author's keywords in the set of selected documents was 3. After setting a higher value, the number of searched keywords and clusters will decrease, which may lead to the omission of an important issue that has not yet been sufficiently investigated and described or is simply not properly exposed in the paper. For example, for the value of 4 for a minimum number of author's keywords in a set of selected documents, there will be 15 keywords and 4 clusters, and for the value of 5 there will be 10 keywords and 2 clusters. Setting a lower than 3 value means that too many words will be treated as keywords. Three as the number of keywords was also proposed as a reference in the publication (Ciano et al., 2019). Using the VOSviewer program, a network (Figure 8) consisting of 16 nodes corresponding to 5 clusters was obtained. Occurrences were used

Figure 8. Co-occurrence network of author's keywords.

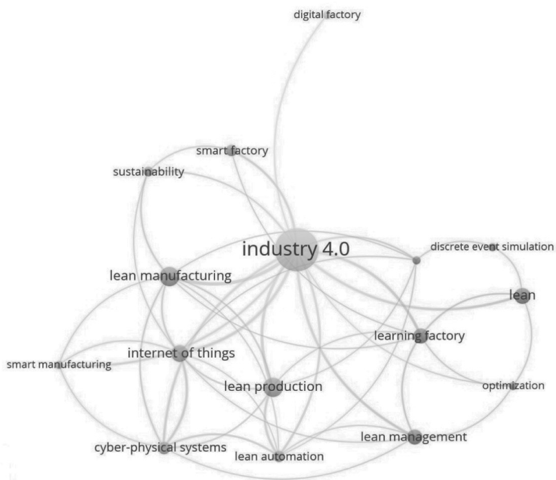


Table 5. Information about author's keywords			
Author's keywords	**Total link strength**	**Occurrences**	**Links**
Industry 4.0	59	48	14
Lean manufacturing	21	12	9
Lean production	19	12	7
Internet of things	19	10	9
Cyber-physical systems	17	6	7
Lean automation	15	5	7
Learning factory	14	8	7
Lean management	11	8	6
Lean	9	9	4
Production management	7	3	6
Smart factory	6	5	3
Sustainability	6	4	4
Smart manufacturing	6	3	4
Optimization	4	3	4
Discrete event simulation	2	3	2
Digital factory	1	3	1

as weights. The size and clarity of the node suggests the frequency of its occurrence in the analyzed set. In turn, the proximity of the location of the elements indicates more frequent than in the case of distant ones, co-occurrence in specific sets.

There were 47 links in the developed co-occurrence network and total link strength was 108. Total link strength attribute indicates the total strength of the co-occurrence of a given keyword with other keywords. The higher its value, the more frequently a given keyword coexists with others and is more relevant to the network. Detailed information on selected author's keywords is provided in Table 5. In order to understand the research trajectories, keywords are listed according to their total link strength, i.e. their importance in the cluster (Waltman et al., 2010).

The obtained clusters were described referring to the papers in which the searched keywords appeared. This allowed to present the results of research in given areas related to LI4.

5.3.1. Cluster 1 author's keywords: Learning factory, lean management, lean, production management, optimization, discrete event simulation

Based on the definition given in (Veza et al., 2015): "A Learning Factory is an environment to support a practice-based engineering curriculum with the possibility of learning the necessary tools and methods, using real life and didactical equipment". Learning factories can successfully provide an appropriate environment for the education of students as well as industry employees (Chryssolouris et al., 2016). There are several modern learning techniques, such as project-based learning or problem-based learning that are widely used and integrated into the practical environment provided by learning factories (Ahmad et al., 2018). The literature describes several examples of learning factories that allow to practice the application of Lean and Industry 4.0 tools. In the learning factory wbk Institute of Production Science, the impact of implementing Lean and Industry 4.0 tools on key performance indicators (KPI) for participating production planners was described (Hofmann et al., 2019). The Institute of Innovation and Industrial Management has been running a learning factory (LeanLab) since 2014. Its goal is to improve the level of academic education, industrial training and practical research in the field of industrial engineering and logistics (Karre et al., 2017). Study (Küsters et al., 2017) describes the functioning of The Textile Learning Factory 4.0 at the Institut für Textiltechnik der RWTH Aachen University in Aachen, Germany. In the literature, we can also find contributions related to the development of design guidelines for I4.0 learning factories and examples of their use in the Smart Mini Factory (Rauch et al., 2019), description of the transition process from Lean learning factory to a learning factory for intelligent production logistics (Bloechl & Schneider, 2016), the theoretical study presenting the framework for connecting Lean and Industry 4.0 in a learning factory (Prinz et al., 2018).

Changes and requirements regarding production management principles, in particular changes in customer requirements in the era of Industry 4.0, are not clear (Yin et al., 2018). Many authors have confirmed in their papers that Lean Manufacturing and Industry 4.0 are not mutually exclusive. They can be integrated with each other for effective production management (Duarte & Cruz-Machado, 2018; Sanders et al., 2016; Satoglu et al., 2018a). Industry 4.0 can provide support through continuous resource management due to the availability of detailed information in real time at every stage of the production process (Gabriel & Pessl, 2016). Therefore, data monitoring provides information on resource consumption and enables flexible production management (Bonilla et al., 2018). Emerging Industry 4.0 technologies enabling the collection of production management data offer the opportunity to receive accurate information and feedback for reliable production planning and control (Dallasega et al., 2017; Reuter et al., 2016). Paper (Araújo et al., 2018) presents the development of an intelligent and automated system for lean industrial production, ensuring maximum productivity and efficiency of the production process. The study (Hrušecká et al., 2018) presents the discrete-event simulation model for increasing the efficiency of warehouse management, which had a direct positive impact on production management.

A review of existing in the literature frameworks, methods and methodology of lean connection and discrete event simulations, along with their characteristics is presented in (Goienetxea Uriarte et al., 2019). Discrete event simulation (DES) is according to many authors (Jahangirian et al., 2010; Negahban & Smith, 2014) the most popular method of simulation. According to the authors (Stadnicka & Antonelli, 2019), discrete event simulations or experiments should be carried out on similar work cells already operating in the factory to determine the correct allocation of tasks. Siemens Tecnomatix Plan Simulation software is recognized as the basic discrete event simulation software that helps create digital models of logistics systems, such as production, to test performance and optimize the system. It also helps in developing what-if scenarios. The usefulness of this tool for conducting virtual experiments on production and logistics systems has been documented by various researchers (Kikolski, 2017; Siderska, 2016; Zupan & Herakovic, 2015). However, there are many other software applications for discrete event simulation and the choice of the most appropriate should be case sensitive. Discrete event simulation is a powerful tool for modeling complex dynamical systems. An example of using the 3D software environment to build a discrete simulation model can be found in the literature (Grube et al., 2019). The use of discrete event simulation to develop Value Stream Mapping (VSM) and the simulation of many production-related parameters such as lead time, added value, stock, utilization were also identified (Trebuna et al., 2019). An example of using discrete event simulation related to physical objects placed on the Digital Twin Module (DTM) is found in (Grube et al., 2019).

5.3.2. Cluster 2 author's keywords: Internet of things, cyber-physical systems, smart manufacturing

Thanks to the intelligent production system, production processes are more flexible, intelligent and agile and are well adapted to meet the challenges of a dynamic and global market (Zhong et al., 2017). Smart manufacturing (sometimes synonymous with Industry 4.0 [Kang et al., 2016]), directs existing production systems towards the development of an open, digital, automated and intelligent production platform for information applications in the industrial network (Kamble, Gunasekaran, Sharma et al., 2018; Vaidya et al., 2018). The concept of smart manufacturing is based on the integration of the IoT and CPS to create Cyber-physical Production Systems (CPPS), which results in the continuous generation of large amounts of data known as Big Data (Basios & Loucopoulos, 2017). CPS is responsible for technology integration, and IoT enables smart data collection, storage, analysis, and sharing technologies (Kamble et al., 2019). The use of machine learning based on embedded sensors in the CPS is presented in (Castaño et al., 2017). In response to the German concept of Industry 4.0, the United States proposed a Smart Manufacturing Plan (Smart Manufacturing Leadership Coalition, 2011) and suggested connecting everything using the IoT (Porter & Heppelmann, 2014, 2015). The Internet of Things and smart manufacturing are the core of Industry 4.0 (Tsai & Lai, 2018). Publication (Duarte & Cruz-Machado, 2018) presents the following definition of smart manufacturing: production equipped with sensors and autonomous systems that will allow to optimize operations with minimal employee intervention (Roblek et al., 2016; Shrouf et al., 2014). The use of sensoring systems and signal analysis to monitor tool wear during the production process is shown in (Beruvides et al., 2013). Automatic selection of optimal parameters based on simple soft-computing methods in the micromilling process is presented in (La Fe-Perdomo et al., 2019). There are also technical studies in the literature that show how available optimization methods can contribute to improving industrial efficiency (Beruvides et al., 2016). It will be possible to produce smaller lots of different types more efficiently (Wang et al., 2016). Smart manufacturing is also defined as the ability to solve current and upcoming problems using an open infrastructure, which allows faster implementation of solutions, building advantage and additional value in the process (Odważny et al., 2018). Some studies suggest that smart manufacturing in Industry 4.0 may be the key to implementing mass customization (MC) strategies. The authors in (Zawadzki & Żywicki, 2016) suggested smart product design and production control for efficient operations in the smart factory to implement the MC strategy. Paper (Cattaneo et al., 2017) contains results of the literature review on smart manufacturing and lean. Paper (Yin et al., 2018) presents an example of a smart manufacturing system for Industry 4.0, and study (Zhang et al., 2017) describes a cloud-based smart manufacturing paradigm based on CPS.

5.3.3. Cluster 3 author's keywords: Lean manufacturing, smart factory, sustainability

A smart factory is the foundation of Industry 4.0 (Germany Trade & Invest (GTAI), 2014). A smart factory will be more flexible, dynamic and intelligent (Roblek et al., 2016). People, systems and objects found in it are connected with each other (Germany Trade & Invest (GTAI), 2014). The IoT is the main technology enabling a smart factory (Germany Trade & Invest (GTAI), 2014; Shrouf et al., 2014). A smart factory is usually associated with lean practices (Tortorella & Fettermann, 2018) and sustainable development (Kusiak, 2018). The smart factory has been designed in accordance with sustainable and business practices, emphasizing flexibility, adaptability and self-adaptability, learning ability, fault tolerance and risk management (Duarte & Cruz-Machado, 2018; Germany Trade & Invest (GTAI), 2014). Sarkis and Zhu (2018) highlighted waste as a link between lean practices and sustainable development. Kusiak (2018) added a new element to this perspective, namely the role of sustainable development in smart production. In particular, the ubiquitous ICT infrastructure of Industry 4.0 can be a factor conducive to the implementation of sustainable production (Gu et al., 2019). Therefore, the relationship between lean, sustainable development, and Industry 4.0, converging on a circular economy, can be a promising future prospect (Ciano et al., 2019). The main research trends identified are increased interest in combining Lean and simulation in the context of Industry 4.0 and their combination with optimization, Six Sigma, as well as sustainable development (Goienetxea Uriarte et al., 2019). Many authors share the opinion that Industry 4.0 technologies will contribute to the organization's sustainable development goals (Carvalho et al., 2018; Lin et al., 2018; Luthra & Mangla, 2018; Kamble, Gunasekaran, Gawankar et al., 2018; Tortorella & Fettermann, 2018; Zhong et al., 2017). An assessment of the relationship between Lean Manufacturing, Industry 4.0 and sustainability is available in the paper (Varela et al., 2019). Study (Slim et al., 2018) presents an analysis of convergence and contradictions occurring between Lean and Industry 4.0 for the innovative design of smart manufacturing within the smart factory. It is also worth paying attention to two very extensive reviews of the literature on: the concept of smart factory (Strozzi et al., 2017) and smart factory in combination with sustainable development and green growth concepts (Odważny et al., 2018).

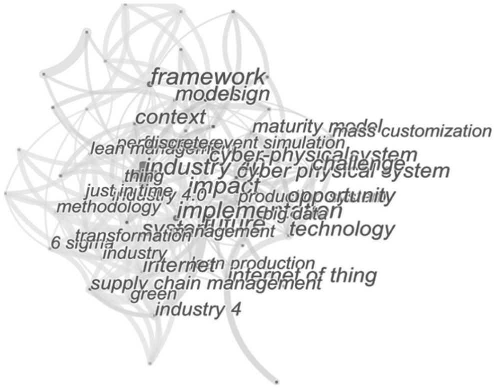
5.3.4. Cluster 4 author's keywords: Industry 4.0, digital factory

The term digital factory, which is one of the key concepts in Industry 4.0, is used interchangeably with a virtual factory or a digital twin (Mrugalska & Wyrwicka, 2017). An example of the digital twin application is presented in work (Guerra et al., 2019). The digital factory is defined as an integrated simulation model providing advanced decision support capabilities (Jain, Lechevalier et al., 2016; Jain, Shao et al., 2016). Simulation is therefore an important tool in the context of Industry 4.0 (Goienetxea Uriarte et al., 2019). The publication (Dombrowski et al., 2019) presents the concept of Digital Factory 4.0, which is a teaching-learning environment in which participants along with a partner cooperating with the industry can independently solve complex tasks having a practical context. Analysis and explanation of the problem is possible based on a visit to a real object. In publication (Tsai, 2018), the author state that Industry 4.0 of the textile industry requires a digital factory. CPS and IoT were recognized as the basic technologies. Publication (Dallasega et al., 2017) presents state-of-the-art planning and real-time production control in the digital factory of the future. An important requirement of the Digital Factory is to provide stakeholders with information and knowledge support during the decision-making process. Just-in-time information retrieval (JITIR) in a digital factory environment is designed to provide stakeholder support through proactive but non-invasive delivery of required information at the right time based on user context. Decision-making activities are taken throughout the entire life cycle of the factory and include location and network planning, material handling and equipment design, process planning or factory operation (Constantinescu et al., 2014; Westkämper, 2006).

5.3.5. Cluster 5 author's keywords: Lean production, lean automation

The integration of Lean Production (LP) practices with Industry 4.0 technologies has been defined as Lean Automation (LA). Its assumption is greater flexibility, changeability and shorter information flow to meet future customer requirements (Kolberg & Zuehlke, 2015). Due to the potential benefits of implementing Industry 4.0 technology, several authors (Gjeldum et al., 2016; Sanders et al., 2016) argued for the existence of new available application areas for LA. An example here is the implementation of flexible, efficient and affordable CPS (Kolberg & Zuehlke, 2015). Currently, LA (Jidoka) using CPS is considered a profitable and effective approach to improving the flexibility of production systems. The publication (Ma et al., 2017) proposed the concept of Smart Lean Automation Engine Enabled by Cyber-Physical Systems Technologies (SLAE–CPS). Another example of LA is the use of modern ICT technologies for the Kanban method (electronic Kanban) (Kolberg

Figure 9. Burst detection algorithm applied to author's keywords.



Keywords	Year	Strength	Begin	End	2014 - 2019
industry 40	2014	6.5907	**2015**	2017	

et al., 2017; Lugert et al., 2018). The next stage will be the integration of a larger number of available LP methods with the developed interface and the creation of other LA solutions using this interface (Kolberg et al., 2017). LA levels are presented in (Satoglu et al., 2018a, 2018b).

5.4. Burst detection analysis

Available literature from a research field can be treated as a series of topics that appear, gain popularity and intensity during a certain period, and then disappear. The appearance of a given topic in the document stream is signaled by a "burst of activity". Some features increase rapidly as a topic appears (e.g., citations). Kleinberg (2003) developed a formal approach to modeling "bursts" in such a way that they can be reliably and effectively identified.

The Kleinberg approach is based on stream modeling using an infinite state automaton in which bursts occur naturally as state changes. In some ways this is an analogy to queuing theory models for burst network traffic.

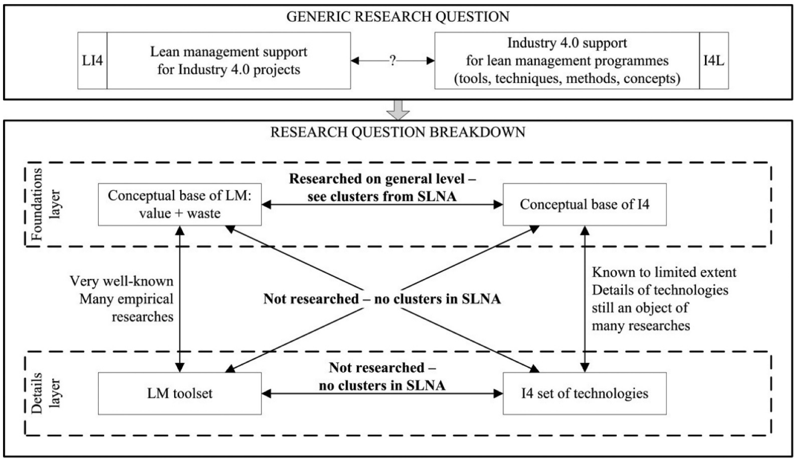
In the paper the Kleinberg's algorithm was applied for author's keywords. The results of applying the burst detection algorithm using the CiteSpace application are shown in Figure 9.

Authors' keywords generated (Figure 9) mostly coincide with those obtained by the VOSviewer program. Just one keyword (Industry 4.0) with the strongest citation was obtained. This is mainly because the analysis period is too short to indicate emerging keywords (bursts). Burst was found only for Industry 4.0 keyword. Therefore, it is obvious that all the papers relate to Industry 4.0 keyword as it effects directly from the designed query. Burst detection has not indicated any emerging and ending topics as the Industry 4.0 is relatively new itself. Considering citations, the only paper with burst was (Kagermann et al., 2013), which is the first one giving comprehensive description of I4.0. Therefore, this analysis also has not indicated any trends.

6. Discussion and further research

According to Dombrowski et al. (2017) two main perspectives are available in the literature: LM is considered a prerequisite for the introduction of I4.0 tools or I4.0 is the LM promoter. The third perspective may be the thesis that the combination of both concepts gives positive synergy effects. This is confirmed by the literature review made by Ejsmont and Gladysz (2020). However, it should be emphasized that the implementation of LM and I4.0 can influence iteratively. Therefore, progress does not have to be sequential (Nyhuis et al., 2017). Mrugalska and Wyrwicka (2017)

Figure 10. Comprehensive framework of Lean Industry 4.0.



support the statement that lean and I4.0 can coexist and support each other. Vogel-Heuser et al. (2017) rejected the contradiction between LM and I4.0. What's more, I4.0 can be helpful in implementing lean and overcome existing obstacles in this process (Sanders et al., 2016).

Combining the results of several analyzes, the paper gave a general picture of the state of scientific work and research trajectories regarding the concept of LI4. Thanks to this, it became possible to set out some ideas for future research directions.

Analyzing the available literature on LI4, interest in this topic has been increasing in recent years. Despite the growing interest of scientists, there is a research gap regarding the assessment of the combination of I4.0 and LM (Slim et al., 2018). The current state of knowledge in the area of Lean Industry 4.0 is presented in the Figure 10.

Based on the created framework (Figure 10), it can be concluded that:

- Lean concept has been verified empirically and there is available extensive literature on empirical and field studies,
- Industry 4.0 is a relatively new concept with few examples of industrial applications described and researched by scientists,
- the impact of I4 on lean and lean on I4 has not been empirically verified and no qualitative, nor quantitative data is widely available in scientific papers,
- LI4 is presented mainly as a concept in which theoretical models are built and verified only by analogy and logical construction.

It would also be interesting to include in the research the precursors of the I4 concept, i.e. CIM (Computer Integrated Manufacturing) or FMS (Flexible Manufacturing Systems) and verify literature on their synergy/contradiction with lean.

It should be noted that the most detailed information about I4, which appears in keywords is discrete-event simulation and big data. Most papers do not present practical examples and empirical evidence, but only describe some general concepts and frameworks. There are no specific lean methods and tools and no specific technologies/solutions I4 listed in keywords.

Therefore, it is highly desirable to conduct further mixed and multi-method research—quantitative and qualitative. It could be organized in the following form:

- Quantitatively—diagnosis, diagnostic surveys, surveys, interviews—initially necessary to narrow down to a selected region/industry and conduct pilot studies,
- Qualitatively—case studies of organizations implementing lean programs and using I4 technologies/tools.

The authors were also tempted to formulate a preliminary hypothesis, which should be verified in further studies: some LM tools and some I4 technologies are used together frequently, while others are less common, which is the signpost of stronger synergies between some pair of Lean/I4 and I4/Lean tools/technologies, and possible contradictions between some other pairs.

Applying citation network analysis, global and local citation scores, co-occurrence network of author's keywords, and using the Kleinberg series detection algorithm, it became possible to identify the main research trajectories for LI4 and indicate its main advantages. Scientists are working on topics related to the use of Industry 4.0 technologies to improve the performance, productivity, efficiency, efficacy and effectiveness of manufacturing systems (Iarovyi et al., 2015). This is indirectly related to the elimination of wastes, which in turn is the basic assumption of LM. It is worth emphasizing that Industry 4.0 is the concept that forms the basis of all clusters, and this concept focuses on the support for other available concepts, including those related to LM (e.g., just-in-time, kanban). In the future, research efforts are needed to investigate the reasons why some Industry 4.0 technologies (e.g., 3D printing, augmented/virtual reality) currently appear to be less popular in the scientific community in the context of LI4.

Based on the literature review carried out in the article, many advantages of combining the LM and I4.0 concepts were identified. The vast majority of advantages are associated with the use of Industry 4.0 technology to support Lean principles/tools (Mayr et al., 2018; Mrugalska & Wyrwicka, 2017; Sanders et al., 2016). For example, the paper (Pereira et al., 2019) presents the matrix "Lean tools supported by Industry 4.0 technologies", (Ejsmont & Gladysz, 2020) presents complex review of Industry 4.0 solutions that may support lean techniques, tools and methods, and (Mayr et al., 2018) presents combinations of I4.0 tools and lean methods. Shah and Ward (2007) conducted a comprehensive, multi-step study to identify a multidimensional lean structure. They quantified the conceptual definition and proposed measuring LM in ten factors, which in turn were divided into 4 categories:

- Supplier factors: supplier feedback, JIT delivery by suppliers, supplier development;
- Customer factor: customer involvement;
- Process factors: pull production, continuous flow, setup time reduction;
- Control and human factors: TPM, statistical process control, employee involvement.

The advantages of combining LM and I4.0 can be demonstrated grounding on the presented above categories of factors (Pereira et al., 2019; Sanders et al., 2016):

a. Supplier factors

- RFID, cloud computing and IoT allow better adaptation to production needs,
- CPS, RFID, IoT support the exchange of information between customers, manufacturers and suppliers (shortening lead times and response to customer complaints),
- digital performance tables to speed up response times.

b. Customer factor

- Customer involvement at the product development stage due to IoT, cloud computing or platforms cooperating with I4.0 technologies.

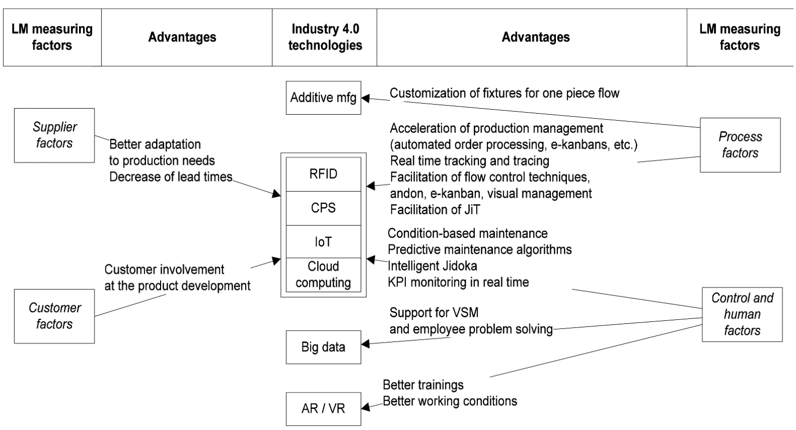
c. Process factors

- access to operational data in real time due to the use of CPS and RFID speeds up and facilitates production management (e.g., automatic order processing, level control using e-kanban or kanban 4.0),
- CPS and RFID makes it possible to monitor the status of details (finished products or work in progress) and to track the quantity and location in real time, as well as communication with machines,
- CPS provide simplified use of andon and e-kanban systems as well as other production flow control techniques,
- additive manufacturing technology for the production of customized spare parts and ensuring one-piece flow,
- sensors, embedded devices and software facilitate more efficient changes,
- IoT and IIoT facilitate JIT implementation.

d. Control and human factors

- CPS are able to collect real-time data on maintenance needs and automatically send signals to the service (e.g., automatic notifications from machines when errors or failures are detected),
- predictive algorithms streamline automatic maintenance,
- quality control based on real data supports self-control (e.g., intelligent Jidoka),
- the use of smart devices, CPS, Big Data, Data Analytics for registration and monitoring of KPI in real time,
- a combination of digital standard operating procedures with cloud technology and augmented reality (AR) to improve operator performance,
- sensors and training in virtual reality (VR) to improve working conditions,
- AR and VR enable testing of dangerous situations,

Figure 11. Lean Industry 4.0 advantages.



- hybrid jobs with collaborative robots,
- Big Data and Data Analytics streamline VSM procedures and facilitate employee problem solving.

Figure 11 depicts summary of advantages that could be achieved through applications of I4.0 technologies as a support for lean management principles. It is clear that the vast majority of synergies is seen in the categories of process factors and control & human factors. However, categories are interrelated and advantages classified as process or control and human factors affect supplier and customer factors as well. E.g. customization of fixtures for one-piece flow, could be also employed for suppliers and customers relations, flow control techniques could be used in-house shop-floor wise, but also in external supply chains.

The conducted research shows that relatively little attention has been devoted to organizational aspects related to the implementation of LI4. The results obtained show that the papers focus mainly on management aspects and responding to changing production requirements, but only in a conceptual way (Sony, 2018). A lot is said on conceptual level of frameworks, but detailed models and extensive case studies are still missing. LI4 models are also missing details considering waste elimination or the combination of I4.0 specific technologies with specific LM tools and techniques (Ejsmont & Gladysz, 2020). Hence, an interesting research agenda can be in-depth research on organizational impact, change management and integration of human resources in the development of LI4, as well as the development of applied models (including reference models and best practices) in these contexts.

The analyzed literature emphasized the fact that LI4 is a concept that goes beyond the enterprise and covers the extended supply chain (Gu et al., 2019). However, LI4 in the extended supply chain is still not discussed in the literature. Another important research direction is the integration of horizontal and vertical systems of Industry 4.0 and LM. This may lead to the need for a holistic approach to LI4 in the supply chain through further research that could also provide empirical evidence (of synergies and contradictions) to complement few theoretical studies.

Most of the publications are of a conceptual, philosophical or review nature and concern only preliminary considerations regarding the possibilities of synergy I4.0 and LM. The descriptive approach dominates, and the presented possibilities of integration of Industry 4.0 and LM for waste reduction or methodological foundations for their cooperation are at a very general level. Studies have shown that despite the existence of models, frames and architectures related to the functioning of LI4 (Arica & Powell, 2014; Kamble, Gunasekaran, Gawankar et al., 2018; Sony, 2018; Xu & Chen, 2017), these are mainly conceptual works. It should be noted that they are rarely supported by empirical researches or case studies confirming the statements presented in the analyzed articles. There are also no conclusions based on practice. Therefore, it would be advisable to examine the functioning of LI4 in an industrial environment in order to supplement the results of the literature study. For example, surveys or case studies of companies located on different continents (not just in Europe) would be beneficial to give a slightly broader view of how LI4 is and can be implemented in the industrial community.

It would be also interesting to examine the barriers and contradictions that influence the synergy of Industry 4.0 and LM. Empirical research and case studies would be desirable as they represent a minority of the available work.

Further research should cover more available papers. In this purpose, studies on LI4 can also be identified in other databases such as: Scopus, IEEE, etc. For more complete results, it should be also consider extending the query of Anglo-Saxon terminology, i.e. “smart manufacturing” and “smart factory”, which appear in some clusters. Analysis can also include German-language papers due to the source of I4. It is also worth considering the inclusion of “white papers”, but this would require establishing a rational method of their selection. It will be also very important to carry out additional analyzes, e.g., co-authorship, co-citation, bibliographic coupling. This will allow obtaining information such as: how many scientists deal with a given topic, what countries/organizations are represented by the authors, what are the relations between authors etc.

Future actions should also focus on issues that have not yet been thoroughly researched and described on scientific grounds. As an example, practical verification of the possibilities of Lean and Industry 4.0 integration to reduce 8 major types of LM wastes can be given. The literature lacks both theoretical models combining Lean and Industry 4.0, as well as case studies or empirical research results in a quantitative (measurable) form. This is important because only then it will be possible to interpret them objectively and to clearly state whether Lean and Industry 4.0 are complementary or not.

7. Conclusions

Even though the literature can identify the works of linking Lean and Industry 4.0, these studies have some limitations. Kolberg and Zuehlke (2015) conclude that a common framework is lacking because the concepts are discussed in an example without a structured approach. This is also confirmed by research carried out by Ejsmont and Gladysz (2020). They show that the majority of Lean and Industry 4.0 combinations are presented at a very general level and there is a lack of quantitative data verifying the benefits of the combination of Lean and Industry 4.0. Descriptive approach dominates, and the presented possibilities of integration of LM and Industry 4.0 in the field of e.g., waste reduction are at a very basic level. There is definitely a lack of empirical research and case studies confirming the synergy of both concepts. Only Wagner et al. (2017) presented the matrix to illustrate the impact of eight I4.0 tools on several lean principles that include specific methods. However, only the effect of CPS on JIT is described in detail. Ejsmont and Gladysz (2020) presented '8 Wastes—Industry 4.0′ framework. Despite this, Lean Industry 4.0 concept still should be thoroughly investigated, delivering frameworks to eliminate contradictions and strengthen synergies.

From the best of authors' knowledge, this work is the first attempt to comprehensively systematize existing scientific knowledge about the concept of LI4. By conducting quantitative bibliometric analyzes using algorithms and software tools, it is possible to obtain a full landscape of knowledge about the conducted study. It is also possible to identify some research trends, as well as to determine the directions of future research, which will cover the dynamic evolution of LI4.

The main theoretical implication of this study is to broaden knowledge of LI4 by analyzing the on-going developments, new trends and emerging topics which have not been sufficiently addressed and require further research. An additional contribution is the use of SLNA methodology to examine a relatively new contributions reflected in literature. This may facilitate the application of the adopted procedure in other areas. The obtained research results enabled the use of citation networks and keyword networks in their main clusters. This can be helpful for scientists to further develop the topic by identifying key issues, emerging trends and evolutionary trajectories.

In relation to the performed analysis, based on the author's keywords, there some benefits of the presented study. The selected author's keywords enable identification of detailed information on the latest topics and issues discussed (five described clusters). Analysis of co-occurrence network of author's keywords has also given the opportunity to isolate topics that have not yet been thoroughly studied (e.g., the role of simulation or Big Data in the concept of LI4, the use of LI4 in supply chains). The results obtained were confirmed and enriched by the use of a burst detection algorithm, which was the last step in the adopted methodology.

The paper also has practical implications: it is the first comprehensive study on the issue of LI4, and also presents information on the current state of knowledge and sets further development directions. Thanks to bibliometric analyzes, critical areas of development of the LI4 concept have been identified, and information has been provided on what the scientific and industrial environment should focus on to enable effective connection of LM and I4.0.

The main disadvantage of the used methodology is the inclusion of keywords that may not always fully correspond to the content of the reports or may contribute to an erroneous determination of the real meaning of the article for a given area of knowledge. Related keywords can sometimes overlook important details of the study being considered. Some keywords may also be omitted when building the network, as they may not meet the coexistence requirements. Another disadvantage is that the data for bibliometric analyzes were taken from one database—WoSCC, which, although quite comprehensive and prestigious, contains only a small part of all scientific publications. One drawback may also be that researchers very often cite works that already have a large number of citations. This is mainly due to the fact that such articles are considered as reliable sources of information, and also have a certain reputation and popularity.

Despite the limitations presented, the results obtained should be considered interesting. The usability of the SLNA method was proven, which thanks to, among others using tools to visualize the citation network and co-occurrence network of author's keywords, enables the support of dynamic analyzes for the presentation of knowledge, or allows the identification of activities to promote and develop further research of a given issue.

Disclosure statement
No potential conflict of interest was reported by the authors.

Supplementary material
The supplemental data for this article can be accessed here.

Acknowledgements
This research was funded by The Polish National Agency for Academic Exchange under grant no. PPI/APM/2018/1/00047 entitled 'Industry 4.0 in Production and Aeronautical Engineering' (International Academic Partnerships Programme).

Funding
This work was supported by the The Polish National Agency for Academic Exchange [PPI/APM/2018/1/00047] and Industry 4.0 in Production and Aeronautical Engineering (International Academic Partnerships Program).

Author details
Krzysztof Ejsmont[1]
E-mail: krzysztof.ejsmont@pw.edu.pl
ORCID ID: http://orcid.org/0000-0003-1516-0878
Bartlomiej Gladysz[1]
E-mail: bartlomiej.gladysz@pw.edu.pl
ORCID ID: http://orcid.org/0000-0003-0619-0194
Donatella Corti[2]
E-mail: donatella.corti@supsi.ch
Fernando Castaño[3]
E-mail: fernando.castano@car.upm-csic.es
ORCID ID: http://orcid.org/0000-0002-4801-9224
Wael M. Mohammed[4]
E-mail: wael.mohammed@tuni.fi
ORCID ID: http://orcid.org/0000-0002-8364-3348
Jose L. Martinez Lastra[4,5]
E-mail: jose.martinezlastra@tuni.fi
ORCID ID: http://orcid.org/0000-0001-6227-3408
[1] Institute of Production Systems Organization, Faculty of Production Engineering, Warsaw University of Technology, Warsaw, Poland.
[2] Institute of Systems and Technologies for Sustainable Production, Department of Innovative Technologies, The University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno, Switzerland.
[3] Centre for Automation and Robotics, Polytechnic University of Madrid (UPM) – Spanish National Research Council (CSIC), Madrid, Spain.
[4] FAST-Lab, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 600, Tampere FI-334100, Finland.
[5] CINTECX, Universidade De Vigo, Campus Universitario as Lagoas, Marcosende, Vigo 36310, Spain.

Citation information
Cite this article as: Towards 'Lean Industry 4.0' – Current trends and future perspectives, Krzysztof Ejsmont, Bartlomiej Gladysz, Donatella Corti, Fernando Castaño, Wael M. Mohammed & Jose L. Martinez Lastra, *Cogent Business & Management* (2020), 7: 1781995.

References
Ahmad, R., Masse, C., Jituri, S., Doucette, J., & Mertiny, P. (2018). Alberta learning factory for training reconfigurable assembly process value stream mapping. *Procedia Manufacturing*, *23*, 237–242. https://doi.org/10.1016/j.promfg.2018.04.023

Araújo, A. F., Varela, M. L. R., Gomes, M. S., Barreto, R. C. C., & Trojanowska, J. (2018). Development of an intelligent and automated system for lean industrial production, adding maximum productivity and efficiency in the production process. In A. Hamrol, O. Ciszak, S. Legutko, & M. Jurczyk (Eds.), *Advances in manufacturing. lecture notes in mechanical engineering* (pp. 131–140). Springer. https://doi.org/10.1007/978-3-319-68619-6_13

Arica, E., & Powell, D. J. (2014). A framework for ICT-enabled real-time production planning and control. *Advances in Manufacturing*, *2*(2), 158–164. https://doi.org/10.1007/s40436-014-0070-5

Basios, A., & Loucopoulos, P. (2017). Six Sigma DMAIC enhanced with capability modelling. *IEEE 19th conference on business informatics* (pp. 55–62). https://doi.org/10.1109/CBI.2017.70

Beruvides, G., Quiza, R., Del Toro, R., & Haber, R. E. (2013). Sensoring systems and signal analysis to monitor tool wear in microdrilling operations on a sintered tungsten-copper composite material. *Sensors and*

Actuators. A, Physical, *199*, 165–175. https://doi.org/10.1016/j.sna.2013.05.021

Beruvides, G., Quiza, R., & Haber, R. E. (2016). Multi-objective optimization based on an improved cross-entropy method. A case study of a micro-scale manufacturing process. *Information Sciences*, *334–335*, 161–173. https://doi.org/10.1016/j.ins.2015.11.040

Bicheno, J., & Holweg, M. (2009). *The lean toolbox*. PICSIE Books.

Bloechl, S. J., & Schneider, M. (2016). Simulation game for intelligent production logistics – The PuLL (R) learning factory. *Procedia CIRP*, *54*, 130–135. https://doi.org/10.1016/j.procir.2016.04.100

Bonilla, S. H., Silva, H. R. O., da Silva, M. T., Gonçalves, R. F., & Sacomano, J. B. (2018). Industry 4.0 and sustainability implications: A scenario-based analysis of the impacts and challenges. *Sustainability*, *10*(10), 3740. https://doi.org/10.3390/su10103740

Buer, S.-V., Fragapane, G. I., & Strandhagen, J. O. (2018). The data-driven process improvement cycle: Using digitalization for continuous improvement. *IFAC-PapersOnLine*, *51*(11), 1035–1040. https://doi.org/10.1016/j.ifacol.2018.08.471

Buer, S.-V., Strandhagen, J. O., & Chan, F. T. S. (2018). The link between Industry 4.0 and lean manufacturing: Mapping current research and establishing a research agenda. *International Journal of Production Research*, *56*(8), 2924–2940. https://doi.org/10.1080/00207543.2018.1442945

Carvalho, N., Chaim, O., Cazarini, E., & Gerolamo, M. (2018). Manufacturing in the fourth industrial revolution: A positive prospect in sustainable manufacturing. *Procedia Manufacturing*, *21*, 671–678. https://doi.org/10.1016/j.promfg.2018.02.170

Castaño, F., Beruvides, G., Haber, R. E., & Artuñedo, A. (2017). Obstacle recognition based on machine learning for on-chip LiDAR sensors in a cyber-physical system. *Sensors*, *17*(9), 2109. https://doi.org/10.3390/s17092109

Cattaneo, L., Rossi, M., Negri, E., Powell, D., & Terzi, S. (2017). Lean thinking in the digital era. In J. Rios, A. Bernard, A. Bouras, & S. Foufou (Eds.), *Product lifecycle management and the industry of the future – PLM 2017* (pp. 371–381). Springer. https://doi.org/10.1007/978-3-319-72905-3_33

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *101*(Suppl. 1), 5303–5310. https://doi.org/10.1073/pnas.0307513100

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377. https://doi.org/10.1002/asi.20317

Chiarini, A. (2011). Integrating lean thinking into ISO 9001: A first guideline. *International Journal of Lean Six Sigma*, *2*(2), 96–117. https://doi.org/10.1108/20401461111135000

Chroneer, D., & Wallstrom, P. (2016). Exploring waste and value in a lean context. *International Journal of Business & Management*, *11*(10), 282–297. https://doi.org/10.5539/ijbm.v11n10p282

Chryssolouris, G., Mavrikios, D., & Rentzos, L. (2016). The teaching factory: A manufacturing education paradigm. *Procedia CIRP*, *57*, 44–48. https://doi.org/10.1016/j.procir.2016.11.009

Chukwuekwe, D., Schjølberg, P., Rødseth, H., & Stuber, A. (2016). Reliable, robust and resilient systems: Towards development of a predictive maintenance concept within the Industry 4.0 environment. *Proceedings of the concept within the Industry 4.0 environment, euromaintenance*, 30 May–1 June.

Ciano, M. P., Pozzi, R., Rossi, T., & Strozzi, F. (2019). How IJPR has addressed 'lean': A literature review using bibliometric tools. *International Journal of Production Research*, *57*(15–16), 5284–5317. https://doi.org/10.1080/00207543.2019.1566667

Colicchia, C., Strozzi, F., & Wilding, R. (2012). Supply chain risk management: A new methodology for a systematic literature review. *Supply Chain Management: An International Journal*, *17*(4), 403–418. https://doi.org/10.1108/13598541211246558

Constantinescu, C. L., Francalanza, E., Matarazzo, D., & Balkan, O. (2014). Information support and interactive planning in the digital factory: Approach and industry-driven evaluation. *Procedia CIRP*, *25*, 269–275. https://doi.org/10.1016/j.procir.2014.10.038

Crew, K., Schoenborn, M., Stemper, J., & Lilyard, C. (2016). E-Journal metrics for collection management: Exploring disciplinary usage differences in scopus and web of science. *Evidence Based Library and Information Practice*, *11*(2), 97–120. https://doi.org/10.18438/B85P87

Dallasega, P., Rojas, R. A., Rauch, E., & Matt, D. T. (2017). Simulation based validation of supply chain effects through ICT enabled real-time-capability in ETO production planning. *Procedia Manufacturing*, *11*, 846–853. https://doi.org/10.1016/j.promfg.2017.07.187

Delbridge, R., & Oliver, N. (1991). Narrowing the gap? Stock turns in the Japanese and Western car industries. *International Journal of Production Research*, *29*(10), 2083–2095. https://doi.org/10.1080/00207549108948068

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. Buchanan & A. Bryman (Eds.), *The Sage handbook of organizational research methods* (pp. 671–689). Sage Publications.

Diez, J. V., Ordieres-Mere, J., & Nuber, G. (2015). The HOSHIN KANRI TREE. Cross-plant lean shopfloor management. *Procedia CIRP*, *32*, 150–155. https://doi.org/10.1016/j.procir.2015.02.120

Doh, S. W., Deschamps, F., & Pinheiro De Lima, E. (2016). Systems integration in the lean manufacturing systems value chain to meet Industry 4.0 requirements. In M. Borsato, N. Wognum, M. Peruzzini, J. Stjepandic, & W. J. C. Verhagen (Eds.), *Transdisciplinary engineering: crossing boundaries 4* (pp. 642–650). IOS Press. https://doi.org/10.3233/978-1-61499-703-0-642

Dombrowski, U., Richter, T., & Krenkel, P. (2017). Interdependencies of Industrie 4.0 & lean production systems: A use cases analysis. *Procedia Manufacturing*, *11*, 1061–1068. https://doi.org/10.1016/j.promfg.2017.07.217

Dombrowski, U., Wullbrandt, J., & Fochler, S. (2019). Center of excellence for Lean Enterprises 4.0. *Procedia Manufacturing*, *31*, 66–71. https://doi.org/10.1016/j.promfg.2019.03.011

Duarte, S., & Cruz-Machado, V. (2018). Exploring linkages between lean and green supply chain and the Industry 4.0. In M. Zhu, M. Gen, A. Hajiyev, & F. L. Cooke (Eds.), *Proceedings of the eleventh international conference on management science and engineering management* (pp. 1242–1252). Springer. https://doi.org/10.1007/978-3-319-59280-0_103

Ejsmont, K., & Gladysz, B. (2020). Lean Industry 4.0 – Wastes versus technology framework. In K. Panuwatwanich & C.-H. Ko. (Eds.), *10th international conference on Engineering, Project and*

Production Management (EPPM 2019) (pp. 537–546). Springer. https://doi.org/10.1007/978-981-15-1910-9_44

Fettermann, D. C., Sa Cavalcante, C. G., de Almeida, T. D., & Tortorella, G. L. (2018). How does Industry 4.0 contribute to operations management? *Journal of Industrial and Production Engineering*, *35*(4), 255–268. https://doi.org/10.1080/21681015.2018.1462863

Gabriel, M., & Pessl, E. (2016). Industry 4.0 and sustainability impacts: Critical discussion of sustainability aspects with a special focus on future of work and ecological consequences. *Annals of Faculty Engineering Hunedoara – International Journal of Engineeringis*, *14*(2), 131–136.

Germany Trade & Invest (GTAI). (2014). *Industry 4.0-Smart manufacturing for the future*. Germany Trade and Invest. Retrieved November 19, 2019, from http://www.inovasyon.org/pdf/gtai.industrie4.0_smart.manufact.for.future.july.2014.pdf

Gjeldum, N., Mladineo, M., & Veza, I. (2016). Transfer of model of innovative smart factory to croatian economy using lean learning factory. *Procedia CIRP*, *54*, 158–163. https://doi.org/10.1016/j.procir.2016.06.096

Goienetxea Uriarte, A., Ng, A. H. C., & Uredna Moris, M. (2019). Bringing together lean and simulation: A comprehensive review. *International Journal of Production Research*, *57*(20), 1–32. https://doi.org/10.1080/00207543.2019.1643512

Grube, D., Malik, A. A., & Bilberg, A. (2019). SMEs can touch Industry 4.0 in the smart learning factory. *Procedia Manufacturing*, *31*, 219–224. https://doi.org/10.1016/j.promfg.2019.03.035

Gu, F., Guo, J., Hall, P., & Gu, X. (2019). An integrated architecture for implementing extended producer responsibility in the context of Industry 4.0. *International Journal of Production Research*, *57*(5), 1458–1477. https://doi.org/10.1080/00207543.2018.1489161

Guerra, R. H., Quiza, R., Villalonga, A., Arenas, J., & Castano, F. (2019). Digital twin-based optimization for ultraprecision motion systems with backlash and friction. *IEEE Access*, *7*, 93462–93472. https://doi.org/10.1109/ACCESS.2019.2928141

Hasle, P., Bojesen, A., Langaa Jensen, P., & Bramming, P. (2012). Lean and the working environment: A review of the literature. *International Journal of Operations & Production Management*, *32*(7), 829–849. https://doi.org/10.1108/01443571211250103

Hofmann, C., Staehr, T., Cohen, S., Stricker, N., Haefner, B., & Lanza, G. (2019). Augmented go & see: An approach for improved bottleneck identification in production lines. *Procedia Manufacturing*, *31*, 148–154. https://doi.org/10.1016/j.promfg.2019.03.023

Hofmann, E., & Ruesch, M. (2017). Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry*, *89*, 23–34. https://doi.org/10.1016/j.compind.2017.04.002

Holweg, M. (2007). The genealogy of lean production. *Journal of Operations Management*, *25*(2), 420–437. https://doi.org/10.1016/j.jom.2006.04.001

Hrušecká, D., Adla, R., Krayem, S., & Pivnička, M. (2018). Event-B model for increasing the efficiency of warehouse management. *Polish Journal of Management Studies*, *17*(2), 63–74. https://doi.org/10.17512/pjms.2018.17.2.06

Iarovyi, S., Lastra, J. L. M., Haber, R., & Del Toro, R. 2015. "From artificial cognitive systems and open architectures to cognitive manufacturing systems." *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)* (pp. 1225–1232). https://doi.org/10.1109/INDIN.2015.7281910

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research*, *203*(1), 1–13. https://doi.org/10.1016/j.ejor.2009.06.004

Jain, S., Lechevalier, D., Woo, J., & Shin, S. J. 2016. "Towards a virtual factory prototype." *Winter simulation conference – WSC 2015*. https://doi.org/10.1109/WSC.2015.7408333

Jain, S., Shao, G., & Shin, S. J. 2016. "Virtual factory based data analytics." *12th IEEE international conference on control and automation – ICCA 2016*. https://doi.org/10.1109/ICCA.2016.7505298

Jasti, N. V. K., & Kodali, R. (2015). Lean production: Literature review and trends. *International Journal of Production Research*, *53*(3), 867–885. https://doi.org/10.1080/00207543.2014.937508

Kagermann, H., Lukas, W., & Wahlster, W. (2011). *Industrie 4.0 – Mitdem Internet er Dinge auf dem Wegzur 4. Industriellen Revolution* [Industry 4.0: With the Internet of Things towards 4th industrial revolution]. *VDI nachrichten*, nr 13-2011. Retrieved May 25, 2020, from https://www.dfki.de/fileadmin/user_upload/DFKI/Medien/News_Media/Presse/Presse-Highlights/vdinach2011a13-ind4.0-Internet-Dinge.pdf

Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group*. National Academy of Science and Engineering. Retrieved November 9, 2019, from https://www.din.de/blob/76902/e8cac883f42bf28536e7e8165993f1fd/recommendations-for-implementing-industry-4-0-data.pdf

Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: An analysis of the citation network. *Sustainability Science*, *2*(2), 221–231. https://doi.org/10.1007/s11625-007-0027-8

Kamble, S. S., Gunasekaran, A., & Dhone, N. C. (2019). Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies. *International Journal of Production Research*, *57*(20), 1–19. https://doi.org/10.1080/00207543.2019.1630772

Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2018). Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. *Process Safety and Environmental Protection*, *117*, 408–425. https://doi.org/10.1016/j.psep.2018.05.009

Kamble, S. S., Gunasekaran, A., & Sharma, R. (2018). Analysis of the driving and dependence power of barriers to adopt Industry 4.0 in Indian manufacturing industry. *Computers in Industry*, *101*, 107–119. https://doi.org/10.1016/j.compind.2018.06.004

Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., Kim, B. H., & Noh, S. D. (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *3*(1), 111–128. https://doi.org/10.1007/s40684-016-0015-5

Karre, H., Hammer, M., Kleindienst, M., & Ramsauer, C. (2017). Transition towards an Industry 4.0 state of the LeanLab at Graz University of technology. *Procedia Manufacturing*, *9*, 206–213. https://doi.org/10.1016/j.promfg.2017.04.006

Kaspar, S., & Schneider, M. (2015). Lean und Industrie 4.0 in Der Intralogistik: Effizienzsteigerung Durch Kombination Der Beiden Ansätze [Lean and Industry 4.0 in the field of intra logistics. Efficiency improvement by combination of the two approaches]. *ProductivITy - Kompetenz in Produktion und Logistik*, *20*(5), 17–20. https://www.industrie-management.de/sites/industrie-management.de/files/sonder-drucke/kaspar-Lean-und-Industrie-40in-der-Intralogistik_productivity_15-5.pdf

Kaufmann, T. (2015). *Geschäftsmodelle in Industrie 4.0 und dem Internet der Dinge: Der Weg vom Anspruch in die Wirklichkeit* [Business models in Industry 4.0 and the Internet of Things: The path from aspiration to reality]. Springer. https://doi.org/10.1007/978-3-658-10272-2

Kikolski, M. (2017). Study of production scenarios with the use of simulation models. *Procedia Engineering*, *182*, 321–328. https://doi.org/10.1016/j.proeng.2017.03.102

Kim, S., Colicchia, C., & Menachof, D. (2018). Ethical sourcing: An analysis of the literature and implications for future research. *Journal of Business Ethics*, *152*(4), 1033–1052. https://doi.org/10.1007/s10551-016-3266-8

Kleinberg, J. (2003). Burst and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, *7*(4), 373–397. https://doi.org/10.1023/A:1024940629314

Knoke, D., & Yang, S. (2008). *Social network analysis*. Sage.

Kolberg, D., Knobloch, J., & Zuehlke, D. (2017). Towards a lean automation interface for workstations. *International Journal of Production Research*, *55*(10), 2845–2856. https://doi.org/10.1080/00207543.2016.1223384

Kolberg, D., & Zuehlke, D. (2015). Lean automation enabled by Industry 4.0 technologies. *IFAC-PapersOnLine*, *48*(3), 1870–1875. https://doi.org/10.1016/j.ifacol.2015.06.359

Krafcik, J. F. (1988). Triumph of the lean production system. *Sloan Management Review*, *30*(1), 41–52.

Kumar, S., Suhaib, M., & Asjad, M. (2020). Industry 4.0: Complex, disruptive, but inevitable. *Management and Production Engineering Review*, *11*(1), 43–51. https://doi.org/10.24425/mper.2020.132942

Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, *56*(1–2), 508–517. https://doi.org/10.1080/00207543.2017.1351644

Küsters, D., Praß, N., & Gloy, Y.-S. (2017). Textile Learning Factory 4.0 – Preparing Germany's Textile industry for the digital future. *Procedia Manufacturing*, *9*, 214–221. https://doi.org/10.1016/j.promfg.2017.04.035

La Fe-Perdomo, I., Beruvides, G., Quiza, R., Haber, R., & Rivas, M. (2019). Automatic selection of optimal parameters based on simple soft-computing methods: A case study of micromilling processes. *IEEE Transactions on Industrial Informatics*, *15*(2), 800–811. https://doi.org/10.1109/TII.2018.2816971

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *9*, 181–212. https://doi.org/10.28945/479

Leyh, C., Martin, S., & Schäffer, T. (2017). Industry 4.0 and lean production – A matching relationship? An analysis of selected Industry 4.0 models. *Proceedings of the federated conference on computer science and information systems. ACSIS* (Vol. 11, pp. 989–993). https://doi.org/10.15439/2017F365

Liker, J. K. (2004). *The Toyota Way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill.

Lin, D., Lee, C. K. M., Lau, H., & Yang, Y. (2018). Strategic response to Industry 4.0: An empirical investigation on the Chinese automotive industry. *Industrial Management & Data Systems*, *118*(3), 589–605. https://doi.org/10.1108/IMDS-09-2017-0403

Lugert, A., Batz, A., & Winkler, H. (2018). Empirical assessment of the future adequacy of value stream mapping in manufacturing industries. *Journal of Manufacturing Technology Management*, *29*(5), 886–906. https://doi.org/10.1108/JMTM-11–2017-0236

Luthra, S., & Mangla, S. K. (2018). Evaluating challenges to Industry 4.0 initiatives for supply chain sustainability in emerging economies. *Process Safety and Environmental Protection*, *117*, 168–179. https://doi.org/10.1016/j.psep.2018.04.018

Ma, J., Wang, Q., & Zhao, Z. (2017). SLAE–CPS: Smart lean automation engine enabled by cyber-physical systems technologies. *Sensors*, *17*(7), 1500. https://doi.org/10.3390/s17071500

Martínez-Jurado, P. J., & Moyano-Fuentes, J. (2014). Lean management, supply chain management and sustainability: A literature review. *Journal of Cleaner Production*, *85*, 134–150. https://doi.org/10.1016/j.jclepro.2013.09.042

Mayr, A., Weigelt, M., Kühl, A., Grimm, S., Erll, A., Potzel, M., & Franke, J. (2018). Lean 4.0 - A conceptual conjunction of lean management and Industry 4.0. *Procedia CIRP*, *72*, 622–628. https://doi.org/10.1016/j.procir.2018.03.292

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., & Barbaray, R. (2017). The industrial management of SMEs in the Era of Industry 4.0. *International Journal of Production Research*, *56*(3), 1–19. https://doi.org/10.1080/00207543.2017.1372647

Mourtzis, D., Fotia, S., & Vlachou, E. (2017). Lean rules extraction methodology for lean PSS design via key performance indicators monitoring. *Journal of Manufacturing Systems*, *42*, 223–233. https://doi.org/10.1016/j.jmsy.2016.12.014

Mrugalska, B., & Wyrwicka, M. K. (2017). Towards lean production in Industry 4.0. *Procedia Engineering*, *182*, 466–473. https://doi.org/10.1016/j.proeng.2017.03.135

Negahban, A., & Smith, J. S. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems*, *33*(2), 241–261. https://doi.org/10.1016/j.jmsy.2013.12.007

Nyhuis, P., Hübner, M., Quirico, M., Schäfers, P., & Schmidt, M. (2017). Veränderung in der Produktionsplanung und –steuerung [Change in production planning and control]. In G. Reinhart (Ed.), *Handbuch Industrie 4.0: Geschäftsmodelle, Prozesse, Technik* [Handbook Industry 4.0: Business models, processes, technology] (pp. 31–50). Hanser. https://doi.org/10.3139/9783446449893.002

Odważny, F., Wojtkowiak, D., Cyplik, P., & Adamczak, M. (2018). Smart factory within sustainable development and green growth concepts. *LogForum*, *14*(4), 467–477. https://doi.org/10.17270/J.LOG.2018.301

Ohno, T. (1988). *Toyota production system: Beyond large-scale production*. CRC Press.

Oztemel, E., & Gursev, S. (2018). Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, *31*, 127-182. https://doi.org/10.1007/s10845-018-1433-8

Pereira, A. C., Dinis-Carvalho, J., Alves, A. C., & Arezes, P. (2019). How Industry 4.0 can enhance lean practices. *FME Transactions*, *47*(4), 810–822. https://doi.org/10.5937/fmet1904810P

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, *92*(11), 64–88. https://hbr.org/2014/11/how-smart-connected-products-are-transforming-competition

Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard Business Review*, *93*(10), 97–114. https://hbr.org/2015/10/how-smart-connected-products-are-transforming-companies

Powell, K., & Peterson, S. R. (2017). Coverage and quality: A comparison of web of science and scopus databases for reporting faculty nursing publication metrics. *Nursing Outlook*, *65*(5), 572–578. https://doi.org/10.1016/j.outlook.2017.03.004

Prinz, C., Kreggenfeld, N., & Kuhlenkotter, B. (2018). Lean meets Industrie 4.0 – A practical approach to interlink the method world and cyber-physical world. *Procedia Manufacturing*, *23*, 21–26. https://doi.org/10.1016/j.promfg.2018.03.155

Rauch, E., Dallasega, P., & Matt, D. T. (2016). The way from Lean Product Development (LPD) to Smart Product Development (SPD). *Procedia CIRP*, *50*, 26–31. https://doi.org/10.1016/j.procir.2016.05.081

Rauch, E., Dallasega, P., & Matt, D. T. (2017). Critical factors for introducing lean product development to small and medium sized enterprises in Italy. *Procedia CIRP*, *60*, 362–367. https://doi.org/10.1016/j.procir.2017.01.031

Rauch, E., Morandell, F., & Matt, D. T. (2019). AD design guidelines for implementing I4.0 learning factories. *Procedia Manufacturing*, *31*, 239–244. https://doi.org/10.1016/j.promfg.2019.03.038

Reuter, C., Brambring, F., & Hempel, T. (2016). Approach of information provision for high resolution production control. *Arabian Journal of Business and Management Review*, *S2*(1), 1–9. https://www.hilarispublisher.com/abstract/approach-of-information-provision-for-high-resolution-production-control-20882.html

Roblek, V., Mesko, M., & Krapez, A. (2016). A complex view of industry 4.0. *SAGE Open*, *6*(2), 1–11. https://doi.org/10.1177/2158244016653987

Rossini, M., Costa, F., Tortorella, G. L., & Portioli-Staudacher, A. (2019). The interrelation between Industry 4.0 and lean production: An empirical study on European manufacturers. *The International Journal of Advanced Manufacturing Technology*, *102* (9–12), 3963–3976. https://doi.org/10.1007/s00170-019-03441-7

Sanders, A., Elangeswaran, C., & Wulfsberg, J. (2016). Industry 4.0 implies lean manufacturing: research activities in Industry 4.0 function as enablers for lean manufacturing. *JIEM*, *9*(3), 811–833. https://doi.org/http://dx.doi.10.3926/jiem.1940

Sanders, A., Subramanian, K. R. K., Redlich, T., & Wulfsberg, J. P. (2017). Industry 4.0 and lean management–synergy or contradiction? In H. Lödding, R. Riedel, K.-D. Thoben, G. von Cieminski, & D. Kiritsis (Eds.), *The path to intelligent, collaborative and sustainable manufacturing* (pp. 341–349). Springer. https://doi.org/10.1007/978-3-319-66926-7_39

Sarkis, J., & Zhu, Q. (2018). Environmental sustainability and production: Taking the road less travelled. *International Journal of Production Research*, *56*(1–2), 743–759. https://doi.org/10.1080/00207543.2017.1365182

Sartal, A., & Vázquez, X. H. (2017). Implementing information technologies and operational excellence: Planning, emergence and randomness in the survival of adaptive manufacturing systems. *Journal of Manufacturing Systems*, *45*, 1–16. https://doi.org/10.1016/j.jmsy.2017.07.007

Satoglu, S., Ustundag, A., Cevikcan, E., & Durmusoglu, M. B. (2018a). Lean production systems for Industry 4.0. In A. Ustundag & E. Cevikcan (Eds.), *Industry 4.0: Managing the digital transformation* (pp. 43–59). Springer. https://doi.org/10.1007/978-3-319-57870-5_3

Satoglu, S., Ustundag, A., Cevikcan, E., & Durmusoglu, M. B. (2018b). Lean transformation integrated with Industry 4.0 implementation methodology. In F. Calisir & H. C. Agdag (Eds.), *Industrial engineering in the Industry 4.0 Era* (pp. 97–107). Springer. https://doi.org/10.1007/978-3-319-71225-3_9

Schuh, G. (2013). *Lean innovation*. Springer. https://doi.org/10.1007/978-3-540-76915-6

Shah, R., & Ward, P. T. (2007). Defining and developing measures of lean production. *Journal of Operations Management*, *25*(4), 785–805. https://doi.org/10.1016/j.jom.2007.01.019

Shrouf, F., Ordieres, J., & Miragliotta, G. (2014). Smart factories in industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm. *IEEE international conference on industrial engineering and engineering management* (pp. 697–701). https://doi.org/10.1109/IEEM.2014.7058728

Siderska, J. (2016). Application of tecnomatix plant simulation for modeling production and logistics processes. *Business, Management and Education*, *14* (1), 64–73. https://doi.org/10.3846/bme.2016.316

Slim, R., Remy, H., & Amadou, C. (2018). Convergence and contradiction between Lean and Industry 4.0 for inventive Design of smart production systems. In D. Cavallucci, R. De Guio, & S. Koziołek (Eds.), *Automated invention for smart industries* (pp. 141–153). Springer. https://doi.org/10.1007/978-3-030-02456-7_12

Smart Manufacturing Leadership Coalition. (2011). *Implementing 21st century smart manufacturing*. Workshop Summary Report. Retrieved November 9, 2019, from https://www.controlglobal.com/assets/11WPpdf/110621_SMLC-smart-manufacturing.pdf

Sony, M. (2018). Industry 4.0 and lean management: A proposed integration model and research propositions. *International Journal of Production & Manufacturing Research*, *6*(1), 416–432. https://doi.org/10.1080/21693277.2018.1540949

Stadnicka, D., & Antonelli, D. (2019). Human-robot collaborative work cell implementation through lean thinking. *International Journal of Computer Integrated Manufacturing*, *32*(6), 580–595. https://doi.org/10.1080/0951192X.2019.1599437

Staufen, A. G. (2016). *Deutscher Industrie 4.0 Index 2015* [German Industry 4.0 Index 2015]. Staufen AG.

Strozzi, F., Bono, F., & Gutierrez, E. (2014). Bibliometric and citation network analysis: Percolation and graph spectral analysis in the framework of wireless sensor network technology. *EUR scientific and technical research series* (pp. JRC92443).

Strozzi, F., Colicchia, C., Creazza, A., & Noè, C. (2017). Literature review on the 'Smart Factory' concept using bibliometric tools. *International Journal of Production Research*, *55*(22), 6572–6591. https://doi.org/10.1080/00207543.2017.1326643

Synnes, E. L., & Welo, T. (2016). Enhancing integrative capabilities through lean product and process development. *Procedia CIRP*, *54*, 221–226. https://doi.org/10.1016/j.procir.2016.05.090

Tabacaru, S. (2019). *Web of science versus scopus: Journal coverage overlap analysis*. Texas A&M University Libraries.

Taj, S. (2008). Lean manufacturing performance in China: Assessment of 65 manufacturing plants. *Journal of Manufacturing Technology Management*, *19*(2), 217–234. https://doi.org/10.1108/17410380810847927

Thoben, K. D., M. Veigt, D. Lappe, M. Franke, M. Kück, D. Kolberg, I. Fahl, R. Zimmerling, J. Schlick, P. Stephan, P. Guth. (2014, June 4–5) Towards networking logistics resources to enable a demand-driven material supply for lean production systems – basic concept and potential of a cyber-Physical logistics system. *7th international scientific symposium on logistics* (pp. 42–69).

Tortorella, G. L., & Fettermann, D. (2018). Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies. *International Journal of Production Research*, *56*(8), 2975–2987. https://doi.org/10.1080/00207543.2017.1391420

Trebuna, P., Pekarcikova, M., & Edl, M. (2019). Digital value stream mapping using the tecnomatix plant simulation software. *International Journal of Simulation Modelling*, *18*(1), 19–32. https://doi.org/10.2507/IJSIMM18(1)455

Tsai, W.-H. (2018). Green production planning and control for the textile industry by using mathematical programming and Industry 4.0 techniques. *Energies*, *11*(8), 2072. https://doi.org/10.3390/en11082072

Tsai, W.-H., & Lai, S.-Y. (2018). Green production planning and control model with ABC under Industry 4.0 for the Paper Industry. *Sustainability*, *10*(8), 2932. https://doi.org/10.3390/su10082932

Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0 – A Glimpse. *Procedia Manufacturing*, *20*, 233–238. https://doi.org/10.1016/j.promfg.2018.02.034

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Varela, L., Araújo, A., Ávila, P., Castro, H., & Putnik, G. (2019). Evaluation of the relation between lean manufacturing, Industry 4.0, and sustainability. *Sustainability*, *11*(5), 1439. https://doi.org/10.3390/su11051439

Veza, I., Gjeldum, N., & Mladineo, M. (2015). Lean learning factory at FESB – University of Split. *Procedia CIRP*, *32*, 132–137. https://doi.org/10.1016/j.procir.2015.02.223

Vogel-Heuser, B., Bauernhansl, T., & Ten Hompel, M. (eds). (2017). *Handbuch Industrie 4.0 Bd.4: Allgemeine Grundlagen* [Handbook Industry 4.0 Bd.4: General principles]. Springer Vieweg. https://doi.org/10.1007/978-3-662-53254-6

Wagner, T., Herrmann, C., & Thiede, S. (2017). Industry 4.0 impacts on lean production systems. *Procedia CIRP*, *63*, 125–131. https://doi.org/10.1016/j.procir.2017.02.041

Waltman, L., Van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. https://doi.org/10.1016/j.joi.2010.07.002

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of web of science and scopus. *Journal of Informetrics*, *10*(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factory of industrie 4.0: An outlook. *International Journal of Distributed Sensor Networks*, *12*(1), 1–10. https://doi.org/10.1155/2016/3159805

Westkämper, E. (2006). New paradigm in manufacturing engineering: Factory life cycle. *Production Engineering*, *13*(1), 143–146. http://publica.fraunhofer.de/documents/N-43846.html

Westkämper, E., Spath, D., Constantinescu, C., & Lentes, J. (eds). (2013). *Digitale Produktion* [Digital production]. Springer. https://doi.org/10.1007/978-3-642-20259-9

Womack, J. P., & Jones, D. T. (1996a). Beyond Toyota: How to root out waste and pursue perfection. *Harvard Business Review*, *74*(5), 140–158. https://hbr.org/1996/09/how-to-root-out-waste-and-pursue-perfection

Womack, J. P., & Jones, D. T. (1996b). *Lean thinking: Banish waste and create wealth in your corporation*. Simon & Schuster.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *Machine that changed the world*. Rawson.

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. *International Journal of Production Research*, *56*(8), 2941–2962. https://doi.org/10.1080/00207543.2018.1444806

Xu, Y., & Chen, M. (2017). An Internet of Things based framework to enhance just-in-time manufacturing. *Journal of Engineering Manufacture*, *232*(13), 2353–2363. https://doi.org/10.1177/0954405417731467

Yin, Y., Stecke, K. E., & Li, D. (2018). The evolution of production systems from Industry 2.0 through Industry 4.0. *International Journal of Production Research*, *56*(1–2), 848–861. https://doi.org/10.1080/00207543.2017.1403664

Zawadzki, P., & Żywicki, K. (2016). Smart product design and production control for effective mass customization in the Industry 4.0 concept. *Management and Production Engineering Review*, *7*(3), 105–112. https://doi.org/10.1515/mper-2016-0030

Zelbst, P. J., Green, K. W. J., Sower, V. E., & Abshire, R. D. (2014). Impact of RFID and information sharing on JIT, TQM and operational performance. *Management Research Review*, *37*(11), 970–989. https://doi.org/10.1108/MRR-10-2014-273

Zhang, Z., Li, X., Wang, X., & Cheng, H. (2017). Decentralized cyber-physical systems: A paradigm for cloud-based smart factory of Industry 4.0. In L. Thames & D. Schaefer (Eds.), *Cybersecurity for Industry 4.0* (pp. 127–171). Springer. https://doi.org/10.1007/978-3-319-50660-9_6

Zhao, D., & Strotmann, A. (2015). Analysis and visualization of citation networks. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *7*(1), 1–207. https://doi.org/10.2200/S00624ED1V01Y201501ICR039

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of Industry 4.0: A review. *Engineering*, *3*(5), 616–630. https://doi.org/10.1016/J.ENG.2017.05.015

Zupan, H., & Herakovic, N. (2015). Production line balancing with discrete event simulation: A case study. *IFAC-PapersOnLine*, *48*(3), 2305–2311. https://doi.org/10.1016/j.ifacol.2015.06.431

PUBLICATION II

Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems

Wael M. Mohammed, Borja Ramis Ferrer, Sergii Iarovyi, Elisa Negri, Luca Fumagalli, Andrei Lobov, and Jose L. Martinez Lastra

This is an Author's Accepted Manuscript of an article published by Taylor & Francis Group in International Journal of Computer Integrated Manufacturing in 2018, available online: https://www.tandfonline.com/10.1080/0951192X.2017.1407874

Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems

Wael M. Mohammed[1], Borja Ramis Ferrer[1], Sergii Iarovyi[1], Elisa Negri[2], Luca Fumagalli[2], Andrei Lobov[1], and Jose L. Martinez Lastra[1]
[1] FAST-Lab, Tampere University of Technology, Tampere, Finland
P.O. 600, FI-33101 Tampere, Finland
[2] Department of Management Engineering, Politecnico di Milano, Milan, Italy
Piazza Leonardo da Vinci, 32, 20133 Milano - Italy
wael.mohammed@tut.fi, borja.ramisferrer@tut.fi, sergii.iarovyi@tut.fi, elisa.negri@polimi.it, luca1.fumagalli@polimi.it, andrei.lobov@tut.fi, jose.lastra@tut.fi

***Abstract*—Information technologies grow rapidly nowadays with the advance and extension of computing capabilities. This growth affects all dependent fields which use those technologies. Industrial automation field is not an exception. This publication describes a general and flexible architecture for implementing Manufacturing Execution System (MES) functions which can be deployed in any industrial cases regardless of the type of industry. These features are achieved by combining the flexibility of knowledge-driven systems with the vendor-independent property of RESTful web services. With deployment of such model, MES functions may gain more versatility and independency. This research work is a continuation of the development the OKD-MES (Open Knowledge-Driven Manufacturing Execution System) approach. The OKD-MES approach is provided by eScop project as a semantic-based solution for flexible and reconfigurable manufacturing execution system. Therefore, the aim of this research consists in presenting a MES functions architecture that could be implemented in OKD-MES in order to increase the flexibility of the manufacturing system. This research work has been conducted during the development phase of the eScop project.**

***Index Terms*—Knowledge-Driven Manufacturing Systems, Manufacturing Execution System Functions, Semantics.**

I. Introduction

Manufactruing systems' providers tend to exploit the latest technologies for increasing the efficiency of their production. This fact creates an intensive competition in the development of novel manufacturing systems. Currently, many research works are targeting the enhancement of the Manufacturing Execution System (MES). As known, in manufacturing systems scope, MES binds the upper level which is the ERP (Enterprise Resource Planning) with lower level (shop floor) in the factories. In this manner, different organizations define (categorize) the functionalities of MES into MES functions such as MESA, ISA95 or VDI [1]. These functions are the source of the bond that MES provides for the manufacturing system. Actually, one of such research work has been recently provided by the eScop[1] project. As a result of the project, a CPS (Cyber Physical System) has been provided as a flexible and reconfigurable solution. The eScop solution has been achieved within the Open Knowledge-Driven Manufacturing Execution System (OKD-MES) concept [2], [3]. The OKD-MES allows to enhance the efficiency of factories because e.g. it reduces the time and effort consumption, which is needed for configuration, maintenance and scheduling orders and resources, among other activities.

This article presents a compatible solution with the OKD-MES concept for implementing MES functions. Therefore, the contribution of this research work is to present a solution that can be employed in order to enhance the flexibility, re-configurability and interoperability of event-driven manufacturing systems. It should be noted that the presented solution is not restricted to systems that implement the OKD-MES concept. The rest of the paper is structured as follows: Section II describes the research background. Then, section III presents the architecture and implementation of the MES functions platform including its main components and their interactions. It presents an example for scenario case as well. Afterwards, Section IV provides the pros and cons of the approach. Finally, the section V concludes the article and presents the work to be done in the future.

II. State Of the Art

A. *Manufacturing Execution System*

Manufacturing Execution Systems (MES) are key supporting tools for production management. They play an important role as bridge between the ERP (Enterprise Resource Planning) systems and lower automation field layers [4].

In many companies such bridge is delegated to a human activity and this clearly creates a lack of opportunity for optimization. As an example, optimization might be applied on resources scheduling or labours exploitation. Moreover, such manual interaction often causes large efforts spent in the optimization of activities at ERP level, as well as important automation solutions implemented at shop floor level, to be useless.

Therefore, MES fills this gap in the automation chain, serving as a central component integrating higher and lower levels of the ISA-95[2] automation pyramid.

Nevertheless, MES cannot be seen as a simple and unique tool that transports information. Instead of that, MES represents a complex system where many functionalities are encompassed. The MESA[3] (Manufacturing Enterprise Solutions Association) defined 11 MES functions. These functions are listed in Table

[1] http://escop-project.eu
[2] https://www.isa.org/isa95/
[3] http://www.mesa.org/en/index.asp

I and described in more detail in [1]. Back in history, MESA organization created the list in the late 90s as a union of all the functions which MES may have. Nevertheless, the factory does not have to implement all the 11 functions to be optimally functional. In fact, the implementation of MES functions depends on several factors as e.g. the industry type or the automation level in the factory.

TABLE I. MESA INTERNATIONAL MES FONCTIONS BRIEFE DESCRIPTION[4]

Function	Description
Resource allocation and status	Manages resources information, providing detailed history and status on real time
Operations/Detail Scheduling	Provides sequencing, recognizing alternative and overlapping/parallel operations
Dispatching Production Units	Manages flow of production units
Document Control	Controls records/forms to be maintained with the production unit
Data Collection/Acquisition	Obtains the intra-operational production and parametric data from the factory floor
Labour Management	Provides optimisation of the labour exploitation
Quality Management	Provides real time information to assure product quality
Process Management	Monitors production and provides or corrects decision support to improve process activities
Maintenance Management	Tracks maintenance activities and provides instance solutions
Product Tracking and Genealogy	Provides the status information of work activities. Also it may generate historical information for the products that have been produced
Performance Analysis	Presents the performance (i.e. KPIs) of the facility for more study and analysis.

As highlighted in section I, another definition of MES functions may be found (ISA95 or VDI[4]). However, MESA definition separates the MES functions in terms of scope of functionality. Thus, eScop solution involved MESA definition as MES functions baseline.

Besides that, and from a research perspective, it is important to keep a comprehensive view embracing all possible functions. In fact, it is recommended for any of the possible new research architecture to support the MES functions in order to keep the highest possible level of applicability in industrial environments. Different researchers have discussed MES application and MES functions [5]–[15] and many of them postulate that the main research gaps to be filled in are:

i) The capability to have flexible functions that can guarantee complete automation of the proposed solution avoiding human activity by-pass;
ii) Complementary to the previous issue, the capability to involve the operators, thus fostering the concept of man-in-the-loop, enabling the functions to be implemented by operators that can communicate with the MES systems thank to PDA (Personal Digital Assistant), RFID (Radio-Frequency Identification) reader, Barcode scanner, tablets, etc.
iii) The interoperability of the MES functions that can be guaranteed only if all the functions are implemented within the same software package or if a flexible alternative is available.

As well, it should be noted that the last gap is the one to be fulfilled within the eScop project approach [1].

B. Interoperability

1) Communication

Communication is defined as the action of exchanging information. Additionally, in the field of information technology, the communication is seen as a protocol for the exchanging the information. The increased demand of such protocols led to founding the W3C (World Wide Web Consortium) by Tim Berners-Lee at MIT in 1994 [16]. At that time, the W3C presented the HTTP (Hypertext Transfer Protocol) protocol. Since then, the W3C has provided various amount of communication protocols for different use. On the other hand, in the same era, OASIS (Organization for the Advancement of Structured Information Standards) provided an architecture for managing the web services such as the so called SOA (Service Oriented Architecture) [17], [18]. Accordingly, it was considered as an advantage for industrial devices to support the SOA architecture. Moreover, the SOA exploits the DPWS (Devices Profile for Web Services) for web services discovery and management [19]. The DPWS is based on WSDL (Web Service Definition Language). In parallel to that, W3C provided REST (REpresentation State Transfer) definitions for the web services. REST is based on HTTP request/response method. The request for REST services can be seen as the CRUD (Create, Read, Update and Delete) term. REST represents web resources using a uniform set of "*stateless*" operations. Furthermore, REST architecture supports several different data formats as e.g. XML, HTML, plain text and JSON. Although REST is not standardized yet, it features light weight processes and fault-tolerant [20].

2) Semantics and Knowledge-Based Systems

Knowledge Representation (KR) and reasoning is a branch of Artificial Intelligence that describes and analyses human reasoning behaviour to support formal calculation and deduction. It defines symbols and languages that allow formalising knowledge in a precise semantics [21]. In other words, KR allows the definition of the logical consequences that are understandable and automatically derivable by computer systems with reasoning algorithms [22]. A key point in the knowledge formalization is the choice of the formal language, which must be sufficiently expressive to allow the description of the domain of interest and efficient enough for (1) not requiring too long reasoning time and (2) ensuring decidability [23]. The container of KR formal descriptions is usually called a Knowledge Base (KB), which stores complex structured and unstructured information through a finite set of propositions on the domain of interest written in the chosen language. KBSs (Knowledge-Based System) include both the syntax of the domain of interest (i.e. definition of rules to define acceptable interpretations of propositions) and a set of operators that provide a meaning or a value to the propositions [24]. On the other hand, KBSs offer a distributed data structure,

[4] http://www.vdi.de

contrarily to databases that provide a fixed data one [25]. They answer to different data needs and it is the main reason for the origination of knowledge bases as an alternative to the traditional hierarchical and relational databases: the KR should not follow a tabular structure with rows and columns, but it is more convenient to use object modeling with a hierarchy of classes, subclasses, relationships and instances. As described in [26], these features are perfectly provided by ontologies.

In literature, ontologies are defined as explicit specifications of a conceptualisation [27] for a shared understanding of information [28]. In particular, a domain ontology is an abstract representation of reality within a certain scope. Ontologies are the natural candidates for implementing KBSs, because they formalize knowledge about a domain improving expertise reusability in knowledge based systems [29]. By their nature, ontologies do not have a specific application domain, but they may be the means to represent knowledge in any domain, in order to make it shared, explicit and formal. In particular, in the manufacturing domain, ontologies have a high potential application for unambiguous communication, to create a shared terminology and semantic alignment, meta-data in computational form for the information infrastructure [30]. The uses of ontology in the manufacturing domain could be several: from the knowledge sharing and reuse, to supporting interoperability in different systems.

The advent of modern smart technologies and distributed control in manufacturing environments has brought to light a promising application of ontologies. In fact, a production system consists of different independent and smart modules that are aware of the capabilities that they can offer to the system but do not have any knowledge on the role they have to play together in the production from a systemic point of view [31], [32]. Ontologies are the perfect means for providing such kind of knowledge to distributed control architectures. In fact, in centralized control architectures, the different system components do not need control information on the role. Instead of that, they have inside the system: the logics of the system has been integrated in the design of the system. On the contrary, distributed control is made of smart components that act as independent elementary modules that perform their local control and require a centralized representation of their role in the system. This role must be formalized through a shared representation of the domain, characteristics that are supplied by ad hoc ontologies [33]. The modular approach of this kind of production systems allows to reduce building, ramp up and reconfiguration time of manufacturing automation systems significantly, because when a module is removed, modified or included, the knowledge of the system - on which the control is based - is easily updated.

The available semantic languages that can be used for implementing ontologies are several, each of them characterized by different syntax, reasoning capabilities and complexity among other features [34]. A comparison of semantic languages against a set of collected requirements from the manufacturing domain is found in [23]. The languages that are advised for this domain are the OWL[5] (Web Ontology Language) and the OWL sublanguages (Lite and DL). They are based on RDF[6] (Resource Description Framework) and are able to represent semantic information in a simple and meaningful manner (through the so-called Triples: Subject-Predicate-Object). They can be queried with the SPARQL language (recursive acronym for SPARQL Protocol and RDF Query Language) to retrieve the information stored in the KBS [23].

C. OKD-MES

Referring to the previous subsections, a particular approach for the implementation of the MES was proposed. This approach aims to provide high level of modularity and re-configurability for the MES solution and underlying automation field layers. The approach combines the capabilities of embedded devices and web services with the advanced approach for the knowledge management and will be referred further as Open Knowledge-driven MES (OKD-MES).

The OKD-MES provides a reference architecture including 4 core layers: Physical Layer (PHL), Representation Layer (RPL), Orchestration Layer (ORL), Visualization Layer (VIS), and a layer of loosely-coupled deployable MES functions as can be observed on Figure 1.

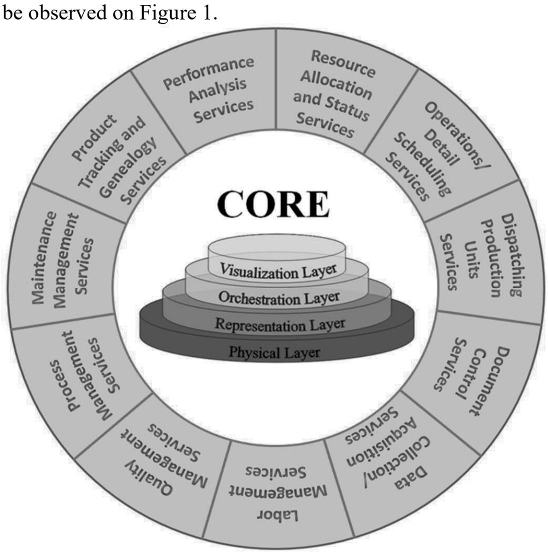


Figure 1 - OKD-MES concept [1]

The PHL addresses the problem of integration of factory shop-floor equipment to the OKD-MES ecosystem. PHL is deployed in the form of RESTful web-service enabled controllers. The PHL devices expose the contract and related services required to interact with the controlled equipment. As well the part of discovery mechanism is implemented in PHL allows to keep the system representation in sync with the real world.

The RPL hosts a manufacturing domain ontology that serves as semantic representation of the production systems, that is instantiated for the specific industrial application case. As such, it offers the capabilities of ontologies presented in section B2: allowing interoperability and knowledge sharing along cyber-physical systems in the OKD-MES [35]. Having a knowledge base in its core, the RPL is capable for some complex data

[5] https://www.w3.org/TR/owl-ref/

[6] https://www.w3.org/RDF/

querying and possibly reasoning. Considering the implementation independent service and capabilities description the RPL allows an interface for a dynamic dependency injection. In addition to the ontology, the RPL implements part of discovery mechanism which interacts with PHL devices, through the ontology service that exposes the ontology as a service on the web service architecture.

The coordination task in OKD-MES is resolved by orchestrators in the ORL. The orchestrator is capable to execute service compositions based on abstract process definitions. Such definitions can be resolved in interactions with RPL to executable service invocations which are handled by ORL. The process definitions are also defining the error and emergency handling in OKD-MES.

The Visualization Layer provides a generic interface between the users and the application. Each OKD-MES component can use declarative UI (User Interface) definition which will be materialized to a web application interface by VIS. The declarative definition allows a dynamic UI generation based on user specified rules.

The four layers mentioned above are providing the ground for implementation of MES functions. Technically, any MES function can be defined as process definition with corresponding UI. Unfortunately, such approach will lead to overcomplicated process definition and will reduce systems performance due to natural compromises between the flexibility and performance. In order to improve performance, the layer of loosely-coupled MES functions was introduced. Such layer enables implementation of the independent, discoverable modules which provide certain functionalities to be used in the OKD-MES. Such modules are exposing services to be injected to the other OKD-MES components using same approach as the one for PHL.

III. MES Platform Architecture

Lately, with the available computational resources, the manufacturing system's installation, configuration and running costs and time have been significantly reduced. This research work presents a dynamic, flexible and reconfigurable platform for providing the MES functionalities for a manufacturing system. The platform employs the KBS and RESTful web services to allow dynamic and autonomous interaction of the components in the manufacturing system. With such solution, it is expected for the manufacturing system to be used as a distributed solution or cloud application.

The suggested platform allows the user to define particular MES functions. According to the user needs, MES functions may provide web services. Accordingly, these services might require some logical or arithmetical calculations. While the calculations might be defined in the form of the functional scripts. This section presents the structure of the platform and defines the workflow of the platform. The section consists of four subsections; in the first one, the components of the platform are described. The second subsection presents the ontology model that has been used for building a knowledge-based system. Meanwhile, the third subsection shows the interactions with the provided solution. Finally, the forth subsection provides a case scenario as a proof of the concept.

A. *Components*

As the majority of web applications, this MES functions platform provides its own services through the web. As well, it is designed to be configurable in running time manners. Accordingly, the platform needs to have an interface with the external environment and systems. Moreover, it has to possess some capabilities to persist internal information. Moreover, the platform should provide the consumers (the manufacturing system) with a specific functionality. The mentioned requirements are satisfied by the following modules: Service Manager (SM) facilitates the interactions with the environment, Ontology Manager (OM) provides the tools to persist the internal information and Function Manager (FM) enables the processing of the information in the system and provides the binding to the corresponding interfaces. Some information about the component design is presented in following subsections, and implementation details can be found in subsection D.

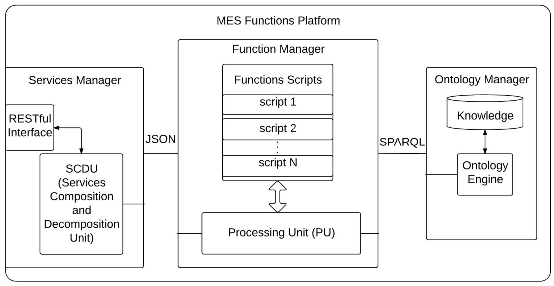


Figure 2: MES platform architecture

1) *Service Manager*

SM (Service Manager) provides the platform with a proper interface for the communications. The SM contains two main components; RESTful interface and SCDU (Services Composition and Decomposition Unit). The RESTful interface supplies the SM with the API (Application Program Interface) capabilities to provide and consume the RESTful services. The SCDU decomposes received RESTful requests or received responses for certain requests. Similarly, it composes the responses for the requested services or for the requests which the platform requires. As a result, SCDU provides the mapping between the knowledge structures used within the system with the generic RESTful concepts. As shown in Figure 2, SM transforms the incoming requests or returned responses into object notation and then passes it to the function manager for further actions. As well, SM allows the reversed flow where the objects defined in the Function Manager are transmitted through the SM to the RESTful environment.

2) *Ontology Manager*

Re-configurability and flexibility are considered as the main features of the proposed approach. Thus, the MES platform employs the KBS. As many KBS based on ontologies, the current approach requires the model to be defined. After that, the user can populate the model with information relevant for the application case. Within this approach, the ontology is managed by a specific component – the Ontology Manager. The OM provides the platform with the proper information via querying and updating the information model. Thus, the OM enables SPARQL-based services for providing the required functionality.

3) Function Manager

Finally, the third main component in the platform is the FM. The FM provides the platform with processing unit for managing the functional scripts. As well, the FM contains the functional scripts which are used to provide MES functionality. The functional scripts are defined by the user on system configuration with regard to the particular use case requirements. Besides, the FM binds the other managers (SM and OM). The FM contains the core of the MES functions platform which is called PU (Processing Unit). The PU provides the runtime environment for the functional scripts.

B. Ontology Model

The ontology model supplies the platform with the required information, i.e. configurations, services and functional scripts accessibility. As presented in the previous subsection, the ontology model is managed through the OM. Figure 3 shows the ontology model for the platform. It contains the following main classes: *OkdMesLayer*, *Configurations*, *MesFunction*, *Service*, *functionalScript* and *Parameter*.

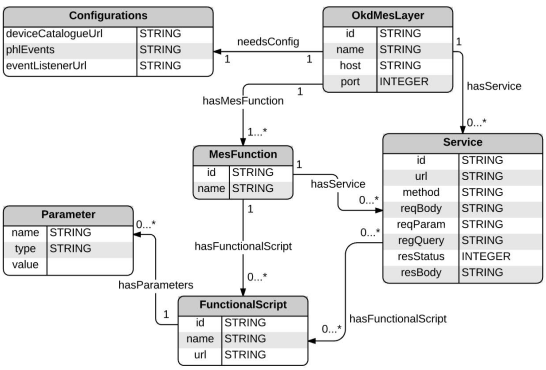


Figure 3: The ontology model

The *OkdMesLayer* class holds the information about the accessibility of the platform. This class includes four datatype properties; *id*, *name*, *host* and *port*. OkdMesLayer is linked to *Configurations*, *Service* and *MesFunction classes* using *needsConfig*, *hasService* and *hasMesFunction* object properties respectively. The *Configurations* class allows the *OkdMesLayer* class to hold configuration parameters regarding the other components in the manufacturing system. It contains deviceCatalogueUrl datatype property for exploring the manufacturing system services. phlEvents datatype property for providing a list of events which the platform should subscribe to. Finally, eventListenerUrl datatype property for determining the URL (Unique Resource Identifier) where the platform will receive notifications of the PHL events. Then, the second link connects the layer with the web services. This means that the platform might have some service instances which are not related to the MES functions.

Thirdly, the *MesFunction* class includes two datatype properties: *id* and *name*. The *MesFunction* class is linked to *Service* and/or *FunctionalScript* via *hasService* and *hasFunctionalScript* object properties**,** respectively. This means that the MES function could have background functions which run without a request from a service. Similarly, it can serve certain services without having a background functions. Then, the *Service* class contains eight datatype properties; *id*, *url*, *method*, *reqBody*, *reqQuery*, *reqPram*, *responseStatus* and *responseBody*. As a RESTful service, the *url* and the *method* are used for routing and validating the correctness of the request. *reqBody*, *reqParam* and r*eqQuery* hold the request information. Meanwhile, the *resStatus* and *resBody* represent the response for the requested service. In this context, the *resStatus* defines the response http status code. It should be noted that the value of the *resStatus* plays a role in the response of the received requests. More illustration is presented in the next subsection. The *Service* class is connected to the *FunctionalScript* class by *hasFucntionalScript* object property. The *FunctionalScript* class represents a logical and/or arithmetic set of operation which might be called by a service invocation or as a background function. The *FunctionalScript* class includes three datatype properties; *id*, *name* and *url*. The *name* is used for calling the function while the *url* presents the accessibility for the function script. In this way, the function script can be a cloud resource which is requested once it is needed. Then the *FunctionalScript* class is linked to *Parameter* class using *hasParameter* property. The Parameter class consists of *name*, *type* and *value* datatype properties. This class is used for two reasons:

1. for passing parameters to the functional scripts.
2. for storing variables in the platform where functional scripts can share data.

C. Interactions

The interactions of such an approach can be seen from two different points of views. Firstly, how the user will setup and run the platform. Secondly, how the platform will provide functionality to the manufacturing system. In this subsection, an illustration is provided for demonstrating the two integration scenarios.

1) Setting up the platform

Alike of any application, the MES platform requires a setting up before the user can run it in the manufacturing system. Therefore, an elucidation of the activities which the user should conduct for running the platform is presented in Figure 4.The user starts by applying a study of the feasibility of using MES functions platform in the manufacturing system. Once it is feasible, the user is subjected to populate the ontology model with proper instances. In this stage, the user is entitled for defining the web services that are required for the platform to serve. Moreover, the user determines the functional scripts that the platform requires. After wards the user uploads the instances to the platform.

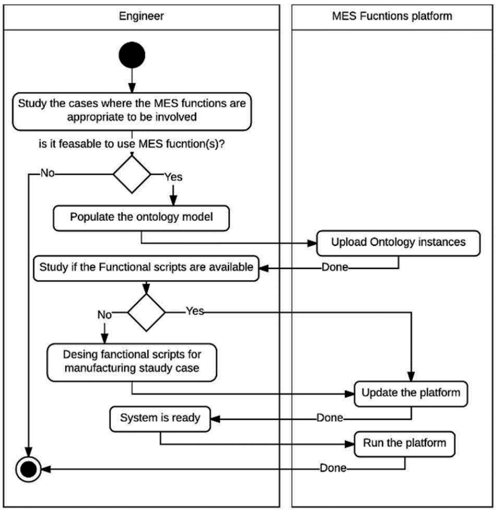


Figure 4:Setup activity diagram

Thereupon, the user examines the existence of the functional scripts. For instance, these functional scripts could be algorithms for optimization or KPI (Key Performance Indicators) formulas. Consequently, the user updates the platform. The update might be URLs for these functional scripts in the ontology instances or code scripts that need to be uploaded to the platform. Finally, the user runs the platform since it is populated with ontology instances and functional scripts.

2) Run time work flow

The second integration scenario addresses the runtime flow work of the platform. Once the platform is set on running mode, it subscribes to all events in the Configurations class in the ontology model.

The subscription of PHL events are an option of the user design to handle PHL information. An example can be seen in resource status and allocation MES function. The status of PHL resources is propagated through events notification. The sequence of the subscription of the PHL events is shown in Figure 5.

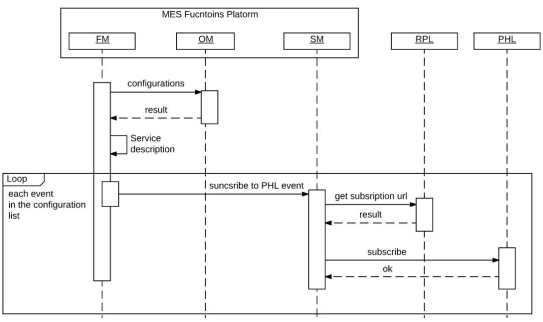


Figure 5: event subscription sequence

The subscription starts with reading the configurations of the platform. Afterwards, the FM requests the SM to perform subscription services for each eventID in the configuration of the platform.

Figure 7 presents a request-response life cycle in the platform. An application, orchestrator as an example, sends a request message to the platform. The request message reaches the SM first through the interface. Consequently, the SM sends a notification to the FM informing that a new request message has been received. Figure 6 presents the notification body of the service request.

```
{
    "serviceUrl":"/performance-analysis/KPI-1",
    "serviceMethod":"GET",
    "serviceBody":{},
    "serviceParams":{},
    "serviceQuery":{}
}
```

Figure 6: SM notification

The PU which is the core of the FM generates a SPARQL query to validate service.

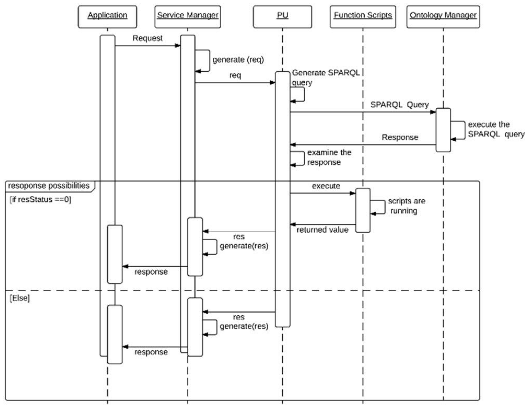


Figure 7: Run time work flow

The validation of the services is combined within the KBS. As highlighted in the previous subsection, the user defines all the services which the platform will serve. Therefore, the service validation can be extracted from response of the SPARQL query.

```
PREFIX MESF:<http://www.fast.tut.fi/MESF#>
SELECT ?FunScr ?resStatus ?resBody ?s
WHERE {
    ?s MESF:method "$ServiceMethod".
    ?s MESF:url "$ServiceUrl".
    ?s MESF:hasFunctionalScript ?FunScr.
    ?s MESF:resStatus ?resStatus.
}
```

Figure 8: validation SPARQL scripts

Figure 8 shows the SPARQL query which is used for validating the request. As shown, the SPARQL query returns the service, functional script, response status and response body of the provided service URL and service method. The result of the query could return empty result in the case where the URL and method are not matching. The empty result leads for *not found* response. On the other hand, the non-empty result means that the requested service is registered in the platform. depends on the *resResult* datatype property, the platform takes its action. For instance, if the *resStatus* datatype equals 0, then the PU responds to the requested service with the result of the functional script that the ontology manager provided. Otherwise, the response is directly provided by the result of the *resBody* of the ontology manager's results. An example of the implementation is illustrated in the next subsection.

D. KPI case-study

In order to demonstrating the usage of the platform, an example of deploying the Performance Analysis MES function is illustrated in this subsection. In this scenario case, the performance analysis MES function performs the following activities:

1. It measures continuously the exploitation ratio of a resource in the PHL as presented in (2).
2. It serves a GET method for retrieving the utilization efficiency KPI which is defined in ISO 22400-2:2014 standard.

It is important to note that the device uses eScop PHL event API template which is shown in Figure 9 in this scenario case.

$$T_{busy} = \sum T_{Idle \rightarrow Busy \rightarrow Idle} \quad (1)$$

$$KPI_{exploitation} = \frac{T_{total}}{T_{busy}} \times 100\% \quad (2)$$

Where,

$T_{Idle \rightarrow busy \rightarrow Idle}$: the period that the resource has been in the busy state.

T_{busy}: the summation of the periods which the resource has been in busy state.

T_{total}: the total period of time since the system is running.


```
{
    "id":"stateChanged",
    "resourceID":"Robot_A",
    "lastEmitted":"1466679151934",
    "payload":{
        "v":"IDLE",
        "q":"good",
        "t":"1466679179807"
    }
}
```


Figure 9: PHL events format as provided by eScop

As appears in Figure 9, the received event form the PHL follows the eScop templates for event. The id and resourceId represent the event ID and the event publisher ID respectively. The lastEmitted provides information about the last emitting of the event. The format of the time in milliseconds since 00:00:00.00 01/01/1970. The specific information of the event is held in the payload. In this manner, the v key maps the value of the resource stat, q for the quality of the value and t for the current time once the event is issued.

To do so, the MES functions should subscribe to the resource state change event in the PHL. In this context, the platform requests from the device catalogue (deviceCatalogue datatype for Configurations class in Figure 10) the specific URL of the subscription. The platform subscribes in PHL by providing the notification URL for the event (eventListenerUrl datatype of Configurations class in Figure 10).

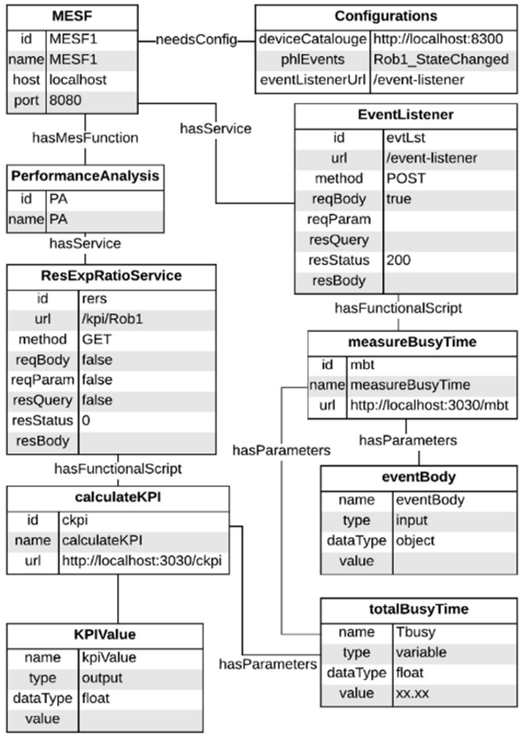


Figure 10: Performance analysis case scenario

With reference to this, the platform performs the functional script which is shown in Figure 10. Each time the platform receives an event with *stateChanged* id and the status is *IDLE.* The functional script performs the arithmetic equations 1 and 2. The OM.setParameter and OM.getParameter are embedded functions for manipulating parameters in KBS. It has to be noted that the functional scripts are in JavaScript language because the platform has been built using Node.JS[7].

```
if(event.id == 'stateChanged' && event.payload.status == 'IDLE')
{
    var busyPeriod = event.lastEmit - event.payload.time;
    var busyPeriods = OM.getParameter(Tbusy) + busyPeriod;
    var T_total = OM.getParameter('T_total');
    OM.setParameter('Tbusy',busyPeriods);
    OM.setParameter('KPIValue',100*T_toal/T_busy);
}
```

Figure 11: measureBusyTime functional script

Then the platform responds with the KPI value once the GET KPI service is invoked. The service uses calculateKPI functional script (see Figure 12).

```
return OM.getParameterValue(parameterId);
```

Figure 12: calculateKPI functional script

[7] https://nodejs.org/en/

IV. Discussion

The application of the proposed approach for the implementation of OKD-MES system is expected to enhance system capabilities. As was already mentioned, one of the primary goals was to increase the flexibility of the system. Usage of the proposed framework allows the dynamic system reconfiguration as the functional requirements and corresponding components are being bound in the runtime. The dependency injection mechanism outlined in the OKD-MES concept was improved and implemented in the presented platform. In combination with the ontology based configuration and communication in the system it enables "cooperation without coordination". As result independent, community driven evolution of the platform functionality is enabled. Considering that the design of the system is based on the open and widely accepted web standards, is technology agnostic and is enabling the proper isolation of abstraction levels some of the technical and social challenges of the platform acceptance are addressed and possible developer community is broadened. The modularity of the system encompasses the small steps migration from the conventional MES solutions to OKD-MES and possible further organic improvement of the systems.

The benefits of the presented approach cost some drawbacks and introduce some challenges. One of the most important drawbacks is the increased demand in computational resources to maintain the performance of the system. As any loosely-coupled system, current, OKD-MES platform has communication and configuration overhead, comparing to the tightly-coupled analogues. Furthermore, the late binding based on the ontology increases its complexity, making the process more resource demanding. Another challenge is the security of the system. Usage of widely accepted web standards and internet based communication leads to an increased amount of vulnerability points in the system. Besides the technical security, there is a wider challenge to overcome the "digital angst" – overall scepticism towards the web, which is especially strong in the established domains such as manufacturing. Finally, to exploit all the benefits of the proposed platform there is a need to modify the paradigm in the development of surrounding systems. For example, the controllers in the manufacturing lines should expose more metadata about themselves, and provide the functionality of the higher level of abstraction.

In opinion of the authors, the advantages of the proposed approach are addressing emerging needs in growing factory information systems, while some of the drawbacks and challenges are showing a trend to be resolved by the advance of technology and overall digitalization in all domains of human life.

V. Conclusion

The article has described an architecture for implementing MES systems. The presented architecture is expected to serve the manufacturing systems through MES functionalities. The platform showed an easy configurability and flexibility in terms of setting up by the user. Moreover, it addressed the genericity of serving different manufacturing systems. Since the platform relies on Node.jsplatform, the installation is expected to simple and fast.

Future work should address development of relevant business models to support OKD-MES principles as migration towards knowledge-driven approaches would require rethinking established industrial practices. The change would need the actions on all the levels of OKD-MES architecture, starting from the controller devices in charge of industrial enjoinment to the higher level information systems, where the shift from data bases to knowledge bases should be performed. This would require development of new methodologies to include proved methods and powerful and easy-to-use tools. However, the authors believe that the development of such tools and methods will be growing and supported due to broad availability of developers and experts working with the web standards, which are also in the core of the presented architecture. Also, well-established tools are available form general software engineering discipline. Those tools can be easily adopted in the field of industrial automation.

References

[1] B. S. Iarovyi, W. M. Mohammed, A. Lobov, B. R. Ferrer, and J. L. M. Lastra, "Cyber – Physical Systems for Manufacturing Execution Systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1142 – 1154, 2016.

[2] L. Fumagalli, S. Pala, M. Garetti, and E. Negri, "Ontology-based modeling of manufacturing and logistics systems for a new MES architecture," in *APMS 2014, Part I, IFIP AICT 438*, 2014, pp. 192–200.

[3] E. Negri, L. Fumagalli, M. Macchi, and M. Garetti, "Ontology for Service-Based Control of Production Systems," in *APMS 2015, Part II, IFIP AICT 460*, 2015, pp. 484 – 492.

[4] MESA International, "MESA White Paper #39: MESA Model Evolution," 2011. [Online]. Available: https://services.mesa.org/ResourceLibrary/ShowResource/73b23d9e-133e-456b-b844-d7ba5ff8278a. [Accessed: 21-Jul-2016]

[5] R. G. Qiu and M. Zhou, "Mighty MESs - State-of-the-Art and Future Manufacturing Execution Systems," *IEE Robotics & Automation Magazine*, no. March, pp. 19–40, 2004.

[6] M. Gonia, "Using Manufacturing Execution Systems (MES) to Track Complex Manufacturing Processes," in *IEEE/SEMI Int'l Electronics Manufacturing Technology Symposium*, 2004, pp. 1–3.

[7] V. Marik and D. McFarlane, "Industrial Adoption of Agent-Based Technologies," *Intell. Manuf. Control*, no. February, pp. 27–35, 2005.

[8] Z. Yu, Q. Fan, and N. Gao, "The Study of Cost Management and Control Based on Knowledge Management for Process Industry MES," in *8th International Conference on Control, Automation, Robotics and Visions*, 2004, no. December, pp. 568–573.

[9] T. Sauter, "The continuing evolution of integration in Manufacturing Automation," *IEEE Industrial Electronics Magazine*, pp. 10–19, 2007.

[10] X. Wang, L. Sun, Q. Meng, and W. Cao, "Study of MES for Cement Industry," in *Proceedings of the IEEE International Conference on Automation and Logistics*, 2007, pp. 1278–1282.

[11] J. Shaohong and Q. Meng, "Research on MES Architecture and Application for Cement Enterprises," in *IEEE International Conference on Control and Automation*, 2007, pp. 1255–1259.

[12] J. Yang, "Manufacturing Execution System for Knitting Industry," in *International Conference on Smart Manufacturing Application*, 2008, pp. 553–558.

[13] J. Hua, H. Sun, T. Liang, and Z. Lei, "The Design of Manufacturing Execution System Based On RFID," in *Workshop on Power Electronics and Intelligent Transportation System*, 2008, pp. 8–11.

[14] M. Younus, L. Hu, F. Yuqing, and C. P. Yong, "Manufacturing Execution System for a Subsidiary of Aerospace Manufacturing Industry," in *International Conference on Computer and Automation Engineering*, 2009, pp. 208–212.

[15] M. Younus, C. Peiyong, L. Hu, and F. Yuqing, "MES Development and Significant Applications in Manufacturing -A Review," in *International Conference on Education Technology and Computer (ICETC)*, 2010, pp. 97–101.

[16] T. Berners-Lee, "WWW: Past, Present, and Future," *Computer*, vol. 29, no. 10, pp. 69–77, 1996.

[17] M. P. Papazoglou and W. J. Van Den Heuvel, "Service oriented architectures: Approaches, technologies and research issues," *VLDB J.*, vol. 16, no. 3, pp. 389–415, 2007.

[18] F. Jammes and H. Smit, "Service-oriented paradigms in industrial automation," *Ind. Inform. IEEE Trans. On*, vol. 1, no. 1, pp. 62–70, 2005.

[19] F. Jammes, H. Smit, J. L. M. Lastra, and I. M. Delamer, "Orchestration of Service-Oriented Manufacturing Processes The advent of SOA," in *Proceedings of the 10th IEEE International Conference on Emerging Technologies and Factory Automation, Vol. 1*, 2005, pp. 617–624.

[20] Ondřej Severa and Roman Pišl, "REST-Enabled Physical Devices in Service Oriented Architecture," in *OPEN KNOWLEDGE-DRIVEN MANUFACTURING & LOGISTICS, THE ESCOP APPROACH*, S. Strzelczak, P. Balda, M. Garetti, and A. Lobov, Eds. Warsaw University of Technology Publishing House, Warsaw 2015, pp. 341–354.

[21] R. Davis, H. Shrobe, and P. Szolovits, "What is a Knowledge Representation?," *AI Mag.*, vol. 14, no. 1, pp. 17–33.

[22] B. Motik and R. Rosati, "Reconciling description logics and rules," *J. ACM*, vol. 57, no. 5, pp. 1–62, Jun. 2010.

[23] E. Negri, L. Fumagalli, M. Garetti, and L. Tanca, "Requirements and languages for the semantic representation of manufacturing systems," *Comput. Ind.*, vol. 81, pp. 55–66, 2016.

[24] A. Yue, W. Liu, and A. Hunter, "Approaches to constructing a stratified merged knowledge base," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 2007, pp. 54–65.

[25] I. Astrova, N. Korda, and A. Kalja, "Storing OWL Ontologies in SQL Relational Databases," *Eng. Technol.*, vol. 1, no. 4, pp. 167–172, 2007.

[26] A. L. Opdahl, B. Henderson-Sellers, and F. Barbier, "Ontological analysis of whole–part relationships in OO-models," *Inf. Softw. Technol.*, vol. 43, no. 6, pp. 387–399, May 2001.

[27] T. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *Int. J. Hum.-Comput. Stud.*, vol. 43, pp. 907–928, 1995.

[28] N. Guarino, "Formal Ontology and Information Systems," *Form. Ontol. Inf. Syst. Proc. FOIS'98*, no. June, pp. 3–15, 1998.

[29] A. Giovannini, A. Aubry, H. Panetto, M. Dassisti, and H. El Haouzi, "Ontology-based system for supporting manufacturing sustainability," *Annu. Rev. Control*, vol. 36, no. 2, pp. 309–317, 2012.

[30] C. Schlenoff, R. Ivester, D. Libes, P. Denno, and S. Szykman, *An Analysis of Existing Ontological Systems for Applications in Manufacturing and Healthcare*. 1999.

[31] M. Garetti, L. Fumagalli, a. Lobov, and J. L. Martinez Lastra, "Open automation of manufacturing systems through integration of ontology and web services," *IFAC Proc. Vol. IFAC-Pap.*, pp. 198–203, 2013.

[32] M. Garetti, L. Fumagalli, and E. Negri, "Role of Ontologies for CPS Implementation in Manufacturing," *MPER - Manag. Prod. Eng. Rev.*, vol. 6, no. 4, pp. 26–32, 2015.

[33] M. Garetti and L. Fumagalli, "P-PSO Ontology for Manufacturing Systems," *INCOM 2012 - Inf. Control Manuf. Conf.*, vol. 14, no. 1, pp. 449–456, 2012.

[34] F. Giunchiglia, F. Farazi, L. Tanca, and R. De Virgilio, "The Semantic Web Languages," in *Semantic Web Information Management*, Springer-Verlag Berlin - Heidelberg, 2010, pp. 25–38.

[35] C. Legat, C. Seitz, S. Lamparter, and S. Feldmann, "Semantics to the Shop Floor: Towards Ontology Modularization and Reuse in the Automation Domain," in *Proceedings of the 19th IFAC World Congress*, 2014, pp. 3444–3449.

PUBLICATION
III

Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing

Borja Ramis Ferrer, Wael M Mohammed, Mussawar Ahmad, Sergii Iarovyi, Jiayi Zhang, Robert Harrison, Jose Luis Martinez Lastra

Knowledge and Information Systems, 2021, 63(6), 1271-1304
Doi: 10.1007/s10115-021-01558-4

Knowledge and Information Systems (2021) 63:1271–1304
https://doi.org/10.1007/s10115-021-01558-4

REGULAR PAPER

Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing

Borja Ramis Ferrer[1] · **Wael M. Mohammed**[1] · **Mussawar Ahmad**[2] · **Sergii Iarovyi**[3] · **Jiayi Zhang**[2] · **Robert Harrison**[2] · **Jose Luis Martinez Lastra**[1]

Received: 20 December 2017 / Revised: 5 March 2021 / Accepted: 10 March 2021 / Published online: 3 April 2021

Abstract

The literature on the modeling and management of data generated through the lifecycle of a manufacturing system is split into two main paradigms: product lifecycle management (PLM) and product, process, resource (PPR) modeling. These paradigms are complementary, and the latter could be considered a more neutral version of the former. There are two main technologies associated with these paradigms: ontologies and databases. Database technology is widespread in industry and is well established. Ontologies remain largely a plaything of the academic community which, despite numerous projects and publications, have seen limited implementations in industrial manufacturing applications. The main objective of this paper is to provide a comparison between ontologies and databases, offering both qualitative and quantitative analyses in the context of PLM and PPR. To achieve this, the article presents (1) a literature review within the context of manufacturing systems that use databases and ontologies, identifying their respective strengths and weaknesses, and (2) an implementation in a real industrial scenario that demonstrates how different modeling approaches can be used

✉ Wael M. Mohammed
wael.mohammed@tuni.fi

Borja Ramis Ferrer
ramis@ieee.org

Mussawar Ahmad
mussawar.ahmad@warwick.ac.uk

Sergii Iarovyi
sergii.iarovyi@kalmarglobal.com

Jiayi Zhang
jiayi.zhang@warwick.ac.uk

Robert Harrison
robert.harrison@warwick.ac.uk

Jose Luis Martinez Lastra
jose.martinezlastra@tuni.fi

1 FAST-Lab., Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland

2 University of Warwick, Coventry, UK

3 Kalmar Global, Tampere, Finland

for the same purpose. This experiment is used to enable discussion and comparative analysis of both modeling strategies.

Keywords Ontology · Database · Comparison · Data modeling · Product lifecycle management

1 Introduction

The advent of computer science and information communication technologies (ICT) in diverse fields such as manufacturing, healthcare and smart cities has improved the manner in which information is created and exchanged between multiple stakeholders [1]. Furthermore, paradigms such as service-oriented architecture (SOA) [2] and cloud computing (CC) [3] can be implemented in order to permit the remote access, storage and manipulation of resources. These can be physical or cyber resources, which are, in turn, mapped to different physical elements such as industrial equipment or measuring devices. This is achieved through the modeling and management of data. More precisely, any domain that can be described within a data model may use models as engineering artifacts for different purposes, e.g., simulation, inference, monitoring and control.

Nowadays, in the case of manufacturing systems, databases seem to be the most common technology used by organizations in order to represent, store and share information [4] with databases currently being used for product lifecycle management (PLM) and product, process and resource (PPR) data modeling [5,6]. Databases are widespread in industry and are more established than ontologies as a means for representing system knowledge [7]. Nevertheless, the design of ontologies for this domain has, in the last decade, gained momentum—particularly in academia. This is evidenced by the increase in published research on this matter in various research portals [8].

Selection of data modeling approach should depend on the needs and desire of end users. However, such decisions are frequently made by software engineers who tend to select a solution according to their knowledge and experience, tending toward those that they are comfortable with. The authors of this paper consider there to be a lack of comparative studies that offer sufficient means for deciding between the employment of ontologies and databases, particularly for those unfamiliar with the former. Some of the few existing studies can be found in [9,10]. However, such works need to be supported by research that provides representative examples to enable a comparison of the capabilities of both technologies within the context of the requirements placed upon them. It is important to state that other knowledge representation (KR)-based solutions, such as production rules or frames, are used in industry; this work considers that, based on contemporary trends, the prominent technologies to be compared are ontologies and databases.

This paper aims to present a qualitative and quantitative comparison between ontologies and databases, demonstrating some of the capabilities of each technology when facing the same issue in the context of manufacturing systems. The main contribution of this article is to provide a study that permits an assessment of the strengths and weaknesses of both technologies for a specific domain—manufacturing systems—which requires robust technologies for storing, accessing and updating data dynamically, at runtime. To achieve this, this research presents a concrete and industrial use case whereby a manufacturing system is described within different semantic models to be accessed and updated. Further, this paper addresses real-world benefits of both data modeling technologies and discusses a set of research questions that need to be answered.

The rest of the paper is structured as follows: Sect. 2 presents a literature and industrial practices review within the scope of this research. More precisely, the review contains an introduction to data modeling and management as well as the definition of ontologies and databases, including a classification and introduction of their main principles. Finally, there is also a description of state-of-the-art transformation solutions. Section 3 describes the methodology that has been followed in order to achieve the reported results. Section 4 presents the test environment that has been designed for the required implementation of this research. Section 5 presents the use case that has been applied to obtain both qualitative and quantitative results for the comparison between ontologies and databases. Section 6 presents and discusses the results with Sect. 7 concluding the article and identifying future directions.

2 Literature and industrial practices review

2.1 Data modeling and management

A vast amount of data are currently generated throughout the product realization process—from design, through to process planning, and then on to manufacturing system design and engineering. Prior to the emergence of the present-day engineering software, models were limited to those instantiated within the physical world, such as mock-ups and prototypes for testing products and processes, as well as some digital versions such as spreadsheet-based calculations to predict costs and cycle times [11]. However, the amount of complexity in a typical manufacturing system has increased due to the emergence of more sophisticated technologies (both within products and manufacturing systems), requiring the expertise of multiple domains for realization. In addition, there is an ongoing paradigm shift toward mass customization and product personalization [12,13]. These factors are the cause of an exponential rise in the amount of data that are now generated by the manufacturing industry, with sources ranging from customer requirements to production systems and the supply chain [14].

Although much can be gleaned from these data, it is necessary for it to be managed and analyzed, to allow maximum value to be extracted from it. Some data are generated from physical systems during operation; however, a considerable amount of data are also generated through the modeling of products and systems. This is done in order to support design activities and to understand the interactions of components—carrying out simulations to predict performance, and visualization to communicate requirements and the desired outcome.

To help manage these data, the key paradigm, recognized as state of the art by the industry, is PLM. PLM manages the storage and exchange of information supporting design and engineering, and integrates it with business processes [5]. PLM is envisioned to allow stakeholders to make data- or information-driven decisions throughout the lifecycle of the product. The implementation of PLM is intended to create a so-called unbroken "digital thread" that prevents the loss of information and ensures that the data are an up-to-date, truthful representation of reality. PLM software acts as a hub or platform that brings together a suite of engineering tools and business processes. Relational database management systems (RDBMS) are a core part of all major, existing, PLM platforms and are renowned for their scalability and stability [15]. However, with the exponential rise in the volume and type of data that now need to be managed by such systems, the efficacy of RDBMS is called into question, particularly within the context of adaptability, expressiveness, interoperability and extensibility [10,16,17]. To continue to support the industry, it is vital for some form of PLM to continue to exist; however, with the increasing complexity and demands on such systems, it is necessary to consider whether new data management and modeling techniques are required.

2.2 Ontologies

2.2.1 What is an ontology?

The word "ontology" has different meanings depending on the context. Firstly, there is the philosophical discipline which is an uncountable noun written as "ontology," which deals with nature and the structure of "reality" [18]. Aristotle dealt with this subject and defined ontology as the "science of being." Unlike the scientific ontology, this branch of metaphysics focuses on the nature and structure or reality independent of how this information would be used.

Contrastingly, the use of ontologies in this research stems from the field of computer science, where it refers to a type of information object. An ontology is a form of KR and is defined by Gruber [19] as "an explicit specification of a conceptualization" while Borst [20] extends this definition to "formal specification of shared conceptualization." Ontologies are a form of KR for a given domain that uses formal semantics and can be used to arrange and define a concept hierarchy, taxonomy and topology.

Ontologies can be accessed for querying and/or modification purposes and they can be implemented using several semantic languages [21,22]. Resource description framework (RDF)-based languages remain dominant, using XML as the syntax option for writing expressions. RDF-based models (e.g., RDF graphs) are sets of triples composed of a subject, a predicate and an object. The Ontology Web Language (OWL) [23] is a description language that extends RDF with cardinality constraints, enumeration and axioms, enabling the creation of richer and more accurate models. OWL comprises three sublanguages: OWL Lite, OWL DL and OWL Full, in order of increasing expressivity, respectively. OWL 2 extends OWL with additional features, including extended datatype support and annotation capabilities. However, OWL remains the prevalent ontology language, with a large number of supporting editors.

The information from OWL models can be queried using an RDF-based query language such as SPARQL [24]. In addition, SPARQL Update [25] can be used for retrieving and updating ontological models. Rule-based languages such as the Semantic Web Rule Language (SWRL) [26] can be employed within ontologies. These rules are defined on top of such ontological models, as presented in [27]. Through the use of rules and RDF triples, semantic reasoning engines can infer implicit knowledge and validate the consistency of a model [28,29].

2.2.2 Types of ontologies

There are different types of ontologies, as reported in [30], with two main criteria that are used to categorize them: the level of formalization, and the level of specificity. In the former, there exist "lightweight" and "heavyweight" ontologies, while in the latter, there exist "foundational," "core," and "domain" ontologies.

– Levels of ontological formalization: Lightweight ontologies are based on simple taxonomies with simple parent child relationships between concepts [31]. Examples of this type of ontology include WordNet [32], and a number of international standards within the context of product data management, such as STEP [33]. This type of ontology has limited concept constraints such that their semantics are insufficient to support interoperability, i.e., to integrate different domain models [34]. To address this, particularly for the STEP format, the ONTOSTEP ontology was developed, which addressed the lack of logical formalism of EXPRESS so that reasoning and semantic operability could be realized [35]. Thus, heavyweight ontologies describe concepts, relationships and logic constraints for automatic prediction and logical inference.

– Levels of ontological specification: Foundational ontologies aim to cover the semantics of “everything” and therefore cover the semantic base for any given domain. Examples of foundational ontologies include DOLCE [36], and the Basic Formal Ontology (BFO) [37]. The concepts in foundational ontologies are generic and as a result are often too broad to be used in a practical engineering context. Core ontologies are limited in the literature and sit at a level of specificity between foundational and domain ontologies. The objective of core ontologies is to cover a set of semantics that are shared across multiple domains [38]. As a result, they lend themselves to reuse and are of particular importance within the context of interoperability. Domain ontologies have the greatest level of specificity and, due to their focus and distinct semantics, interoperability between domain ontologies is challenging. Within the context of supporting manufacturing system lifecycles, it is therefore incumbent on the domain ontology development team to identify domain touchpoints and to ensure that links and mappings exist between the relevant concepts.

2.2.3 Ontology development methodologies

As a result of over two decades of development and learning, a number of methodologies have evolved to support the development of ontological models from the modeling process through to implementation and use. In 1994, the U.S. Air Force defined an ontology method to structure semantic information modeling called IDEF5 [39]. An ontology acquisition process was developed based on five basic steps [40].

1. Organizing and Scoping of Project: The structure and content of the project is described in this part and the main objectives of ontology development are clearly specified.
2. Data Collection: The raw data are defined and classified to enable the development of the ontology and the data collection methods are summarized from different domains.
3. Data Analysis: This part is used to analyze the existing data material to establish an initial ontology for knowledge engineers and the domain developer.
4. Initial Ontology Development: By developing prototype ontologies, ontology classes, properties, attributes and relationships are refined and given detailed specifications.
5. Ontology Refinement and Validation: This phase integrates the known information with the ontology. Through a refinement procedure, ontologies are summarized in specification form for evaluation by domain experts.

Based on the IDEF5 methodology, [41] a documentation stage is added to standardize the ontologies and to support the foundation for future ontology development. METHONTOLOGY introduced the iterative development of an ontology, with a focus on maintenance [42]. Reusing knowledge from existing ontologies forms part of a seven-step guide for ontology creation by Noy and McGuinness [43]. Other than the knowledge reuse aspect, the method remains similar to what is proposed by [40]. An important conclusion derived by Noy and McGuinness is that there is no single correct ontology for a given domain, despite following a common methodology. Determining whether the "right" ontology has been created can only be done by using it in the application for which it was developed [44].

2.3 Databases

2.3.1 What is a database?

The concept of the database appeared following development of direct-access memory, i.e., immediately after stored computer data became feasible. The term database appeared in the early 1960s, and since that moment, multiple database implementations have emerged [45]. The term database in computer science is understood as a structured collection of data [46]. This collection includes several kinds of objects, such as schemas, tables and queries, that permit the representation of data to enable it to be interpreted and reused by computer systems and by humans.

As the deployment, across multiple domains, of IT-based solutions for managing and storing data has grown over the last few decades, specific types of databases have been selected and implemented, according to the requirements of the field of application. The authors suggest the following non-exhaustive expectations for a database in the modern production environment:

1. Databases are expected to be medium sized, i.e., smaller than global social network databases occupying the data centers around the world, but large enough to accommodate all relevant production information.
2. Database models are expected to be of average complexity. Due to standardization of the production processes and components, the allowed abstraction level can be relatively high.
3. Database models are expected to be stable. This is because changes made to adequately designed database models are only made in relation to significant changes in the production.
4. Databases are expected to ensure data consistency, i.e., corrupt data should be spotted early, while the failure is still recoverable.
5. Databases are expected to provide high data throughput, accessibility and robustness. Basically, a database should not be a bottleneck of the production process.

The authors of this research work claim that, in the manufacturing domain, databases should, at least:

1. be medium-sized models in order to easily manage, access and update them;
2. allow the description of static schemas, not affected by a highly dynamic number of requests;
3. ensure data robustness and consistency;
4. permit the processing of multiple requests from/to different data providers and consumers in parallel.

2.3.2 Types of databases

There are many ways to classify different kinds of databases as they can be differentiated according to their structure, contents or application area. These characteristics affect interrelated concepts such as data storage, organization and access. Data storage typically refers to the number of levels of abstraction between the data and its representation in computer memory. The levels of abstraction add functionality to the database but increase memory usage and, generally, the access time. The data can be stored in its binary form directly in the database (DB) program memory, on a disk as a file in binary format, in DB-specific format, or even in plain or marked-up text format. Furthermore, the stored data can be present in a single place in memory or can be replicated across clusters.

Different organization approaches influence the storage and access options as well as the performance and applicability of certain techniques for representing data. In turn, represented data affects several aspects such as consistency, synchronization, redundancy and robustness. Among the access options, the most common ones include direct access as well as several querying languages, such as SQL [47], NoSQL [48] and some customized and/or proprietary ones, such as Hyper Text SQL (HTSQL).[1] One of the main objectives of this manuscript is to provide a qualitative and quantitative comparison between ontologies and databases. This research considers relational SQL, NoSQL and graph databases [45], which can be mapped to RDF-based models.

From the first databases that emerged in the 1960s, such as the network-based CODASYL [49] or the hierarchy-based IMS [50], to the recent models, many different types of database have been designed and implemented for diverse applications. It is important to note that database engines are capable of handling specific types of databases. For instance, Oracle[2] allows, as a primary data model, a relational database management system (DBMS). However, some database engines permit additional secondary data models or even multi-model functionality, i.e., processing different database types as primary data models. For example, while Oracle permits document store and key-value store as a secondary data model, the OrientDB[3] allows the implementation of document store, graph DBMS and key-value store.

2.3.3 Database development methodologies

There are many methodologies that database designers and developers may follow in order to create coherent and consistent data models [51–53]. As there are many different types of databases that can be developed, it is not feasible to find a methodology that supports the creation of any kind of data model, covering all the required steps.

Nevertheless, there are many common steps regardless of the database type being created. Fundamentally, the development of a database model starts with a decision on terminology for certain concepts such as entities, relationships, attributes and constraints. This convention of terms is similar to the fourth step presented in the basic step list for designing ontologies in the previous section. Following this step, is it necessary to check for any redundancy in the model for simplification purposes [54].

2.4 Previous work on comparison of different approaches for data modeling

A number of works have been published that present some level of comparison between ontologies and databases. In some cases, only a passing comment is made, while others delve deeper. In order to demonstrate the value and contribution of this work, the authors present what has already been discussed in this area alongside the remaining questions.

Ontologies differ from a database approach as their focus is on the preservation of meaning to facilitate interoperability, whereas the main purpose of a database schema is to store and

[1] http://htsql.org/.

[2] https://www.oracle.com/database/index.html.

[3] http://orientdb.com/orientdb/.

query large data sets [9]. One of the most comprehensive reviews on the topic was presented in [10], which aimed to clarify the differences and similarities between ontologies and databases. Similar points were also raised in [55]. A summary of the conclusions made concerning the differences are as follows:

– Design approach: Databases are created from scratch for a specific purpose, whereas ontologies may be created by reusing existing ontologies. Although ontologies can also be created from scratch to be used for a specific purpose, their inherent dependence on semantics facilitates reuse for unforeseen applications—unlike databases.
– Manner of KR: Databases relies on closed world assumption (CWA), which means that the assumption is that the model represents complete information. This has the consequence that what is not known to be true, must be false. Ontologies, however, use an open world assumption (OWA), whereby if a query does not return a result, the interpretation would be that the information is unknown.
– Syntax: Databases utilize entity-relationship diagrams, which represent the logic of the database, whereas ontologies are expressed in languages with which you can describe logics. By extending this notion, semantic features are the underlying foundation of ontologies, but are unimportant for databases.

There are also some similarities: the expressiveness of the respective tools resembles each other to some degree (classes $\ncong$ tables, properties $\ncong$ attributes, and axioms $\ncong$ constraints). Thus, we conclude that the key differences are derived from how the respective tools are used: databases are for storing large data sets, while ontologies are focused on integrating semantic data or exchanging information between heterogeneous systems. An example of exchanging data between heterogeneous software is present within PLM [56,57]. As such, it has been proposed that ontologies can make databases entirely redundant within this context. This is because the conceptual model is stored together with the instances. Additionally, when transforming the conceptual model associated with a database to physical and logical models, there is an associated semantic loss [58].

In [59], a framework for representing functional knowledge within ontologies to retain design intent is presented. The authors explicitly decide not to use databases specifically because they have been known to hinder the reuse of documents due to the lack of semantic constraints for functional knowledge. Within the context of the work, the authors define a semantic constraint as a restriction that allows the description of a model that complies with the conceptualization committed by the author.

Research that presented a comparison of databases and ontologies through implementation within a medical data management system concluded that SPARQL, querying triple stores, retrieved instance data via the Virtuoso Universal Server faster than SQL, querying a relational database [16]. Comments were also raised concerning the usability of SPARQL versus SQL, whereby the former adhered to a clearer standard which was not always the case for SQL-capable systems. A further advantage of the ontological approach was its flexible schema that could be extended without comprehensive system redesign. On the other hand, the OWA offers no constraint validation, requiring implementation of this functionality in the application layer.

Finally, in the scope of this research, Bizer and Schultz presented the Berlin SPARQL Benchmark (BSBM) [60], which is a study of the querying performance on a variety of different RDF and SQL-based stores via SPARQL and SPARQL-to-SQL queries. The authors performed a set of queries where the data size and the number of clients (representing the number of the end users) changed in order to add realism to the conducted tests. As a result, the SPARQL-to-SQL rewriters slightly outperformed the RDF as the data set increased. It is important to highlight that the authors did not discuss the low-level specification of different technologies used. An analysis of their experiment infrastructure would provide a benchmark for comparing the results presented in this research work.

2.4.1 Transformation tools

Moving on from the comparison above, it is clear that there are some areas where the respective data modeling methods are complementary. A number of transformation tools have been developed to allow the benefits of the respective approaches to be exploited. Such tools enable the sharing and reuse of knowledge structures to support domain experts in addressing the integration and analysis of existing data sets. Relational database-based conversion tools serve as a method to facilitate ontology development by reducing development lead times, examples include DB2OWL, RDB2Onto and OWL2DB [61].

- DB2OWL: DB2OWL is a conversion tool that can automatically generate ontologies from relational databases via mapping database tables and description logic using OWL DL language [62]. Based on their algorithm, database concepts are translated to a related ontology component. For example, tables will be classes in ontology description; columns and rows are represented by properties and instances; the relation in database schema is relationship between domain ontologies. The advantage of this tool is that it can automatically generate records for logging ontology mapping processes including (1) each corresponding description for the ontology component, (2) the conceptual relationship between ontology and database, and (3) the mapping history of instances and attributes [63]. However, this tool depends only on a particular case or database table, and current databases only support Oracle and MySQL due to limited metadata. In addition, data mapping cannot occur across different databases to generate one ontology.
- RDB2Onto: The automatic generation of ontologies usually focuses on mapping a relational database with ontology concepts, such as, such as DB2OWL, D2R and R2O [64]. RDB2Onto is an SQL-query-based RDF/OWL translation tool that can transfer existing data to ontology templates using only SQL queries. In order to analyze XML schema in an ontology template, data are merged to an ontology data format. This tool is developed in the Java environment using the Sesame and Jena library, which support SPARQL to connect an ontology with a MySQL database, but it can also be used for any other relational database. The advantage of this solution is its simplicity and ease of operation via a visual user interface [65]. RDB2Onto also provides the ability to customize instances and create decision-making rules through an ontology library. Unlike DB2OWL, this approach cannot directly generate database instances to ontology. Furthermore, the main components of this tool are the OWL Builder and the OWL Writer, which cannot preserve the ontology structural constraints. Thus, this tool does not support reasoning tasks for extending ontology with predication of rules.
- Others: There are solutions that permit transformation from OWL to relational databases [66]. In fact, the work reported in [66] describes the main principles for mapping OWL elements to relational database schemas within a specific tool, based on the OWL2DB transformation algorithm. Furthermore, a qualitative comparison between similar transformation solutions is provided. The research works compared are, predominantly, the ones reported in [67–72]. The aforementioned articles demonstrate that the mapping between ontology and database models is feasible and must be taken into account in environments that employ both types of modeling approach. However, OWL2DB focuses on a one-to-one class relationship and breadth-first search method. As a result, the per-

Table 1 Research questions about ontologies and databases

#	Research question	Requirements
1	What is the effect on performance (as a set of characteristics) as the volume of data within a given model is incrementally increased?	Investigate the differences on processing when using each technology. This should be tool-agnostic and focus on the performance differences between fundamentals of ontologies and databases
2	How could database models and ontologies interact with each other in a future scenario?	Investigate the employment of ontological models in conjunction with databases to support complex and demanding needs of ICT-based platforms
3	What is the perceived difference in effort for implementing and maintenance of a database versus an ontology for common applications?	Evaluate the required effort to create script, model and resources of each solution. In addition, research about the effort of modification and maintenance of models at different phases (e.g., design and operation)

formance of this tool is limited by the transformation algorithm. Depending on different cases, this tool may not create all the relationships between tables or classes. Further, knowledge can only be transformed in terms of OWL Lite syntax and a part of OWL DL syntax.

3 Open questions and a methodology for comparing digital data modeling technologies

The review of literature and industrial practices in the scope of PLM and PPR data modeling led to the discovery of a set of unanswered questions. As shown in Table 1, this section presents a set of research questions (RQs) and required research actions to address them.

These RQs are the starting point of this research. The following steps have been performed to compare different data modeling technologies:

1. Model and environment design: This consists of the selection of the main methods and tools for designing similar data models to describe and control the same system using different data modeling technologies (ontologies and databases). The decisions and the final environment that was selected for this research are described in the following section; the decision has three core aspects: data collection, applications and tools, and test environment.

2. Implementation: This concerns the implementation of the test environment and the data models. The completion of this step provides an experimental setup that permits the assessment of different features of the data modeling technologies. In relation to RQ3, this step illustrates some of the aspects of model implementation that require management.
3. Test and compare: This is the final step of the methodology. It involves testing both technologies to obtain results for analysis and discussion. The testing of each model demonstrates directly the effects on performance, which is the concern of RQ1. In addition, the discussion of the experimental tests leads to the identification of potential synergies between databases and ontologies.

4 A test environment for the lifecycle engineering databases and ontology comparison

As a representative case in the scope of PLM and PPR data modeling, this research work implements a means of data collection for retrieving information from a discrete assembly line. Principally, the objective is to collect random events that are triggered by industrial controllers located in such a line. These events describe the status of each machine in the line. More precisely, an orchestrator engine has been designed in order to produce one variation among the available products. While the line is producing the products, the orchestrator records all triggered events. Over a period of 12 hours, the line produced 100,000 random events. The authors of this research believe that the randomness and unconstrained nature of the data are representative of a real manufacturing environment. In fact, the randomness is generated by the nature of the manufacturing systems, where events can be triggered once a change occurs. For example, a change can include pallet position, machine status, operations feedback or safety alarms. The event collection routines are not linked to the current process or status of the production system.

Besides the data collection, the research methodology requires the selection of the applications that are to be used for the comparison. The goal is to exploit the same tools and frameworks for each kind of modeling technology to ensure a valid comparison. The selection study took more effort than was expected, since the available support for both technologies is significantly different. This contrast originated from several factors; for example, the life span of the technology was the major factor, since the two technologies have many years of difference in terms of maturity. Besides that, the level of usage and maturity also plays an important role in terms of technical and programming support. The authors use the Java development environment[4] with similar libraries, due to the availability of frameworks such as SQL[5] provided by Java, which suits each of the technologies being compared. In addition, the Java Microbenchmarking Harness (JMH)[6] framework is used for measuring the operation-to-time ratio for both implemented technologies. Within this setup, each technology to be compared had similar programs for making the required benchmarking test. The MySQL[7] data store has been used to implement the database store, and Jena ARQ[8] has been used to store the data for the ontology model.

4 https://www.java.com/en/.

5 https://docs.oracle.com/javase/7/docs/api/java/sql/package-summary.html.

6 http://tutorials.jenkov.com/java-performance/jmh.html.

7 https://dev.mysql.com/.

8 https://jena.apache.org/documentation/query/.

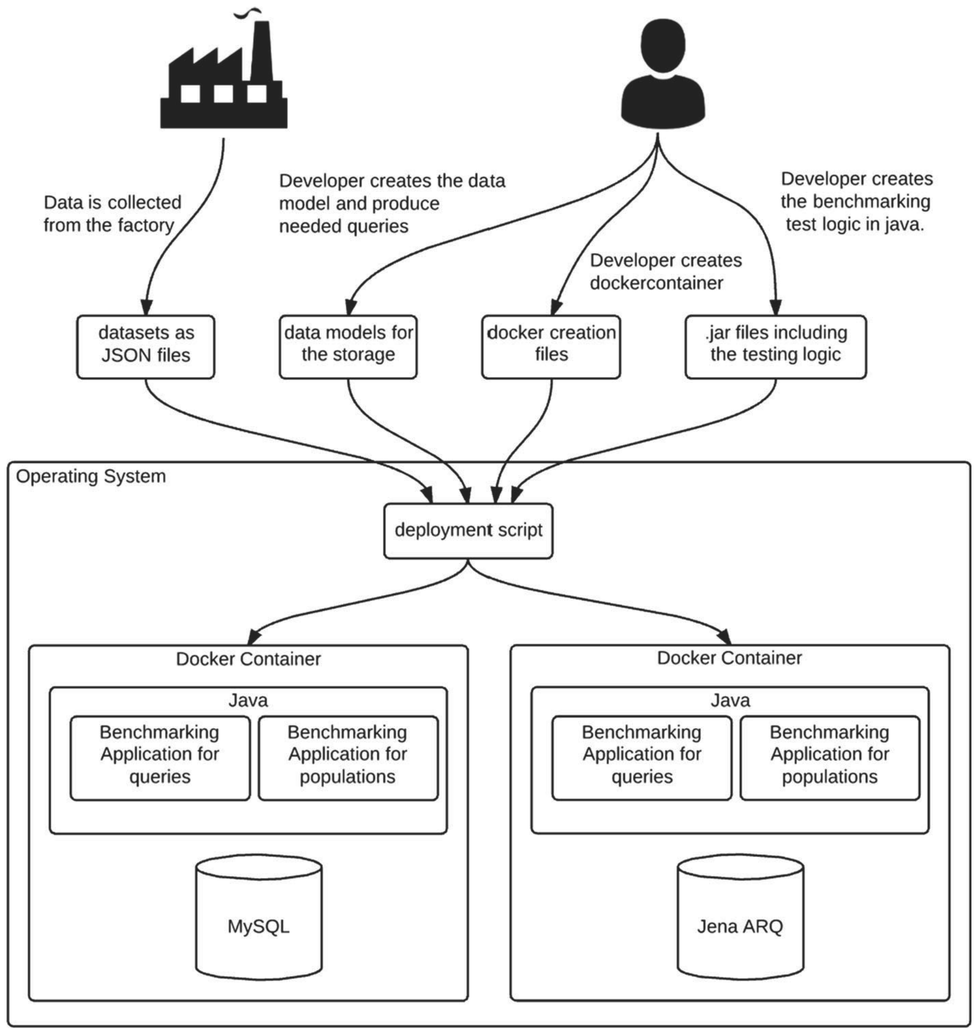


Fig. 1 Deployment environment for database and knowledge-based technology validation

Finally, the decision on the deployment environment has ensured similar impact due to computational resources such as central processing unit (CPU) capabilities, random access memory (RAM) size and background services. The objective is to execute both benchmarking applications on the same machine with the same operating system (OS) conditions and background services. Furthermore, it is important to deploy both applications without them affecting or interfering with each other. As these requirements could be achieved by employing virtual machines or containers, "Docker"[9] containers have been built based on Linux Ubuntu images for the deployment of the benchmarking tests. As shown in Fig. 1, both tests are deployed on similar Docker images—the difference being the data store, since the technology is different.

[9] https://www.docker.com/.

Fig. 2 A completed mobile phone from the FASTory line

5 Lifecycle engineering models in manufacturing: a use case

5.1 The FASTory line

The FAST-Lab (Future Automation Systems and Technologies Laboratory)[10] FASTory line is a production line that demonstrates the assembly process of mobile phones by drawing different variants on sheets of paper that are located on special pallets. Up to three components are drawn: the frame, screen and keyboard. Each of the mobile parts may be drawn in any one of three different colors and three different models. This means that the line can produce 81 different mobile models and 729 different mobile variants, taking into account different color options. Figure 2 shows the FASTory line and an example of a finalized product.

The FASTory assembly line contains a workstation (WS1) for loading/unloading papers to/from the pallets using a SCARA robot. Another workstation (WS7) is used for loading/unloading pallets from the assembly line served by a human operator. Ten workstations are used for drawing purposes. These workstations are identical and are able to draw all mobile models with different colors. All workstations include a segment of the central transport system, which is based on a belt conveyor. All workstations used for drawing operations have a path for pallets to bypass the workstation if it is operating—reducing the possible delays or traffic in the overall production process. This is appreciable from Fig. 3, which shows the interface of a FASTory line web-based simulator. In addition, each conveyor is divided into multiple zones that have one presence sensor to detect the presence of the pallet, one stopper to stop the pallet and an RFID reader for pallet recognition. Each red-filled triangle of four represents a different stopper, each located at a different zone of the conveyor. Each drawing workstation has up to five conveyor zones, while WS7 and WS1 have only four zones.

The FASTory line has evolved during the implementation of several European projects, such as the eSonia[11] and eScop[12] projects. Some of the tasks performed during the eScop project made it possible to create a remotely accessible virtual replica of the production line to support the project developers. This virtual replica is referred to as the FASTory Simulator. In this research work, the FASTory Simulator[13] is used to collect the event logs for the comparison tests of ontologies and databases, with it being the system description container.

10 https://www.tuni.fi/en/research/fast-lab#expander-trigger-field-group-members.

11 https://artemis-ia.eu/project/18-esonia.html.

12 https://cordis.europa.eu/project/id/332946.

13 http://escop.rd.tut.fi:3000/fmw.

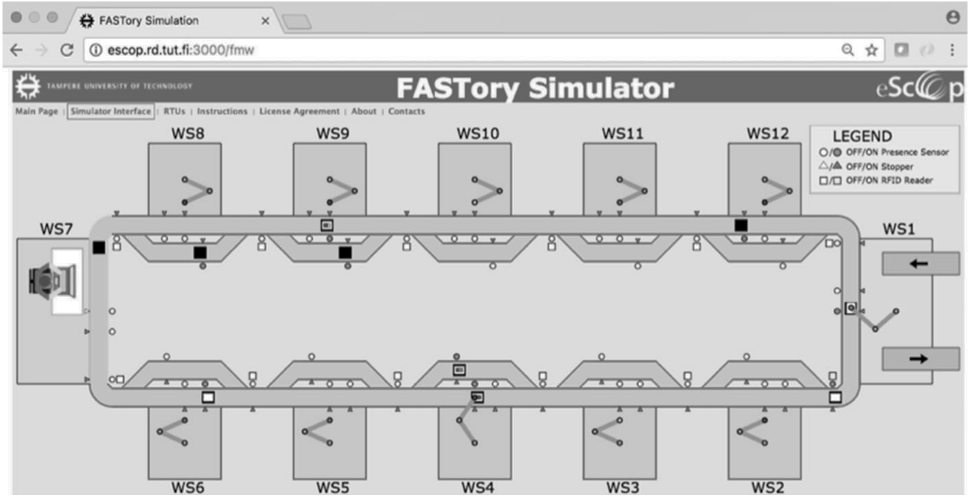


Fig. 3 FASTory layout shown in the FASTory web-based simulator

5.1.1 Collecting data in the FASTory line

To achieve the aim of collecting data from the FASTory Simulator, a web-service-enabled orchestrator has been designed. This engine consists of two main blocks: the JobExecuter and the Logger. As depicted in Fig. 4, the orchestrator is an application that runs on a normal personal computer (PC) and is connected to the FASTory network through an Ethernet socket. Figure 4 shows that a switch has been used to connect to different remote terminal units (RTUs) which, in turn, are connected to different devices (robots and conveyors) of the FASTory line. The RTUs are devices profile for web services (DPWS)-enabled devices that permit the description of service operations that can be executed in order to control the performance of the robots and conveyors. In Fig. 4, the RTUs are labeled according to the type of device and number of web services that they control, i.e., ROB1 RTU denotes an RTU connected to the robot located at WS1 and CNV12 RTU denotes an RTU connected to the conveyor located at WS12.

At the initialization phase, the orchestrator subscribes to each event in the FASTory line. This subscription allows the Logger block to be notified whenever any change occurs in the line. The Logger then creates a JSON object to store all the notifications. Hourly records are stored for each day. In the experiment performed for this research, up to 106,154 events have been collected over a total of 12 operating hours. The JSON-formatted records allow parsing for further analysis.

The JobExecuter (JE) block is capable of managing a simple production process. The production process tested in this research experiment requires the participation of all workstations. This scenario provides several events of different types, sent from multiple senders. Figure 5 shows all possible event types that could be generated, including their main information. As depicted, each event type (ET) includes three principal objects: timestamp, event (which in turn contains id and senderId), and finally the payload object. It is the payload object that categorizes the ET. While ET1 includes a payload object with palletID, ET2 also contains the recipe item. Additionally, ET3 includes color information.

The difference between the ETs is originated by each event sender. Each type of event is linked to specific operations that are executed in the line. ET1 originates from CNV RTUs

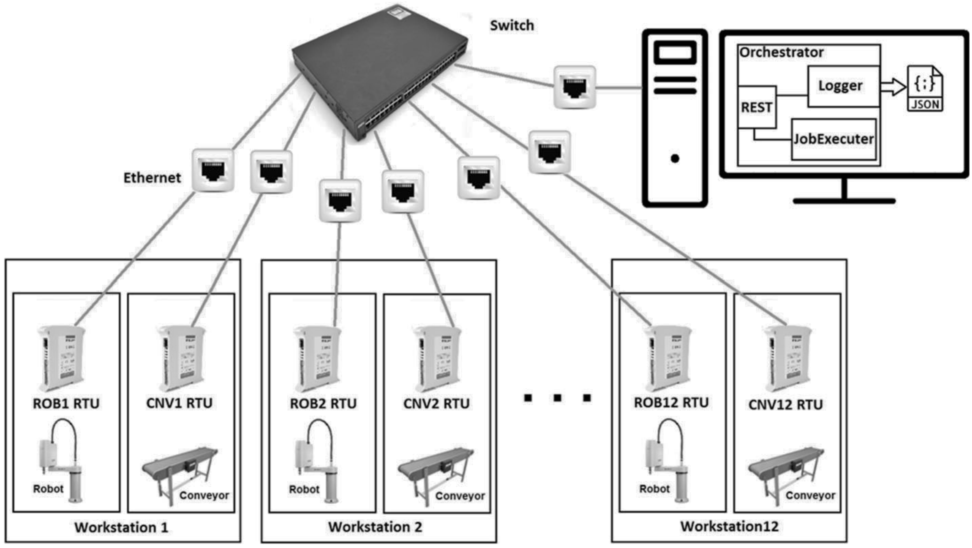


Fig. 4 Data collection orchestrator integration

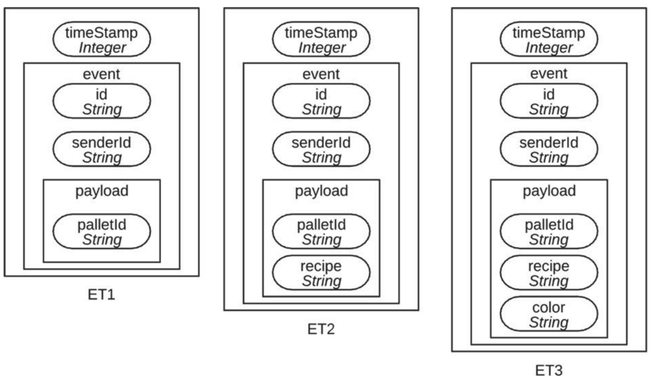


Fig. 5 Information included in each ET

whenever a pallet moves to a new conveyor zone (i.e., a new position). The first operation linked to ET1 is Z_CHANGED. In addition, ET1 is issued when executing PAPER_LOADED and PAPER_UNLOADED operations, which notify the load or unload of papers in WS1. Secondly, ET2 is linked to the DRAW_START_EXECUTION operation, which is executed for starting a drawing with any robot from WS2 to WS6 and from WS8 to WS12. Finally, ET3 is linked to the DRAW_END_EXECUTION, which is triggered once a robot finishes the drawing process. Figure 6 shows an example of the JSON format of an ET3.

5.2 An ontology-based modeling approach for the FASTory line

This subsection presents the ontology model designed to describe the FASTory line. In the UML class diagram presented in Fig. 7, the model that this research employs for storing, retrieving and reasoning information generated from FASTory events via ontologies is described.

```
{
            "timeStamp":1501750807464,
            "event":{
            "id":"DRAW_END_EXECUTION",
            "senderId":"ROB2",
            "payload":{
                    "palletId":"1501749696163",
                    "recipe":"4",
                    "color":"RED"
            }
            }
}
```

Fig. 6 ET3 JSON format example

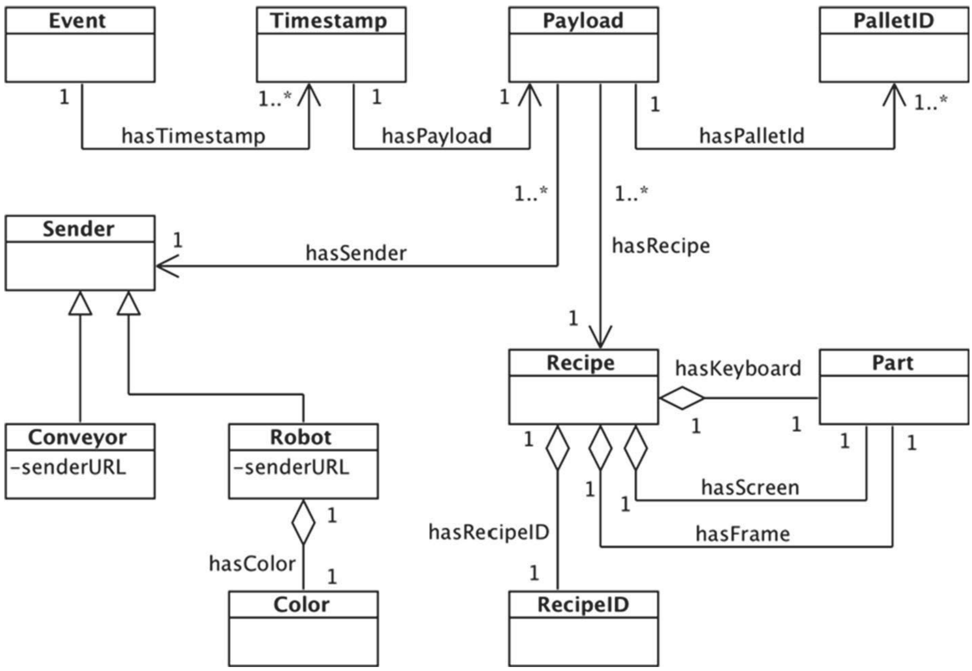


Fig. 7 FASTory ontology model represented within the UML class diagram

Figure 7 illustrates that the ontology is composed of 11 classes. Besides the depicted range and domain of the object properties, the model includes the datatype property senderURL, which is used for describing the URL of event senders, i.e., robot and conveyor RTUs. In order to demonstrate the implementation of the presented model, Fig. 8 shows the model in Protégé.[14] Protégé has been used as the ontology editor at the design phase to create the model and to perform both consistency tests and execution of queries in order to validate the model.

As presented in Fig. 8, the model includes certain instances by default. There is one instance for each robot and conveyor as well as the senderURL datatype property value, which is a string indicating the sender URL. Furthermore, all robots are linked to the color RED, as the default color that any robot of the FASTory line will use for drawing operations. This research requires the population of different models in order to evaluate each modeling

[14] https://protege.stanford.edu/.

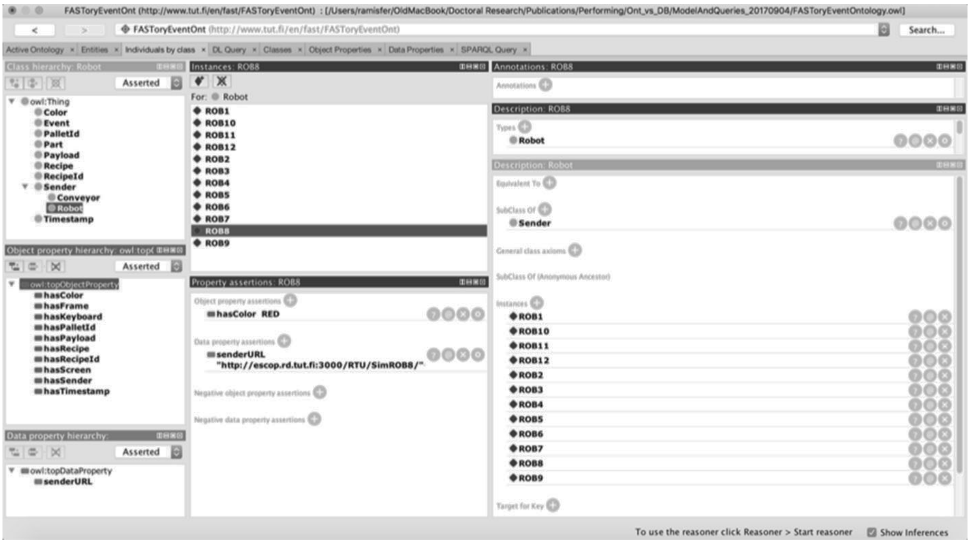


Fig. 8 FASTory model seen via the Protégé user interface

```
PREFIX ont: <http://www.tut.fi/en/fast/FASToryEventOnt#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
INSERT DATA
{
ont:%EVENT_ID% rdf:type ont:Event.
ont:%TIMESTAMP% rdf:type ont:Timestamp.
ont:%EVENT_ID% ont:hasTimestamp ont:%TIMESTAMP%.
ont:%PAYLOAD% rdf:type ont:Payload.
ont:%TIMESTAMP% ont:hasPayload ont:%PAYLOAD%.
ont:%PALLET_ID% rdf:type ont:PalletId.
ont:%PAYLOAD% ont:hasPalletId ont:%PALLET_ID%.
ont:%PAYLOAD% ont:hasSender ont:%SENDER_ID%.
ont:%PAYLOAD% ont:hasRecipe ont:%RECIPE%.
ont:%PAYLOAD% ont:hasColor ont:%COLOR%.
}
```

Fig. 9 SPARQL query for ET3 population

approach. The population of each model has been done in the environment described in Sect. 4. Figure 9 shows, as an example, the query that permits populating events of type 3 (ET3), as presented in Fig. 6. This illustrates the structure of the implemented update queries.

Since the ontology population is about updating the model, the executed queries are SPARQL Update queries, which are usable within RDF-based models, as described in Sect. 2. It is important to mention that the words in between "%" characters are variables, to be replaced by the Java code in order for a query to be fully executable. These variables are taken from each incoming event for the model population.

Besides similar queries for populating the ontology with ET1 and ET2 events, this research required the design and implementation of SPARQL SELECT queries for retrieving information that is useful for supporting RQ1 and RQ3. This is then presented to aid discussion and to demonstrate the results and required efforts during the research.

Table 2 Data type events

Path	Data type	Mandatory
timeStamp	TIMESTAMP	Yes
event.id	VARCHAR[20]	Yes
event.senderId	VARCHAR[10]	Yes
event.payload.palletId	CHAR[13]	Yes
event.payload.recipe	VARCHAR[5]	No
event.payload.color	VARCHAR[5]	No

5.3 A DB-based modeling approach for the FASTory line

An open-source database system named PostgreSQL[15] is used for DB-based modeling. The data modeling for this research employed generic RDBMS and does not exploit the specific features provided by PostgreSQL, being instead representative of SQL technology in general. The RDBMS's features include the ability to store and to query the data in JSON format—similar to other document-oriented DBs. Storing the event data in JSON format may decrease the querying performance of a DB, but does significantly simplify the design process, as the message itself can persist in the DB. There are several steps in the design of a relational DB. Firstly, the data to be stored in the system should be classified to data primitives. Secondly, the data should be organized across DB tables and connected via relations.

As described in Sect. 5.1.1, the events generated by FASTory line are similar in structure and share multiple similar data parts. These events must therefore have an associated timestamp, type, sender, pallet, and may have recipe and color in the model. The timestamp can be directly mapped to the TIMESTAMP data primitive to allow advanced operations with the column. The event id, sender id, recipe and color are varying length strings, while the pallet id is a constant length string. All the fields shown in Table 2 are present in all the events with the exceptions of color and recipe.

The next step involves defining the tables and the relations between them. The structure of the tables depends mainly on the data structure, considering aspects such as the relation between fields in the data and which parts of the data are continuously updated. The design of the queries that are needed could affect the table's structure where some SQL commands depend on the technology employed. As a result, the developer needs to find the balance between query performance and the nature of data for constructing tables in the database. In some cases, the exact set of expected queries to be run on the database is known in the design phase; this can also be true of industrial DB deployments. This issue becomes more apparent with the contemporary trend for a more iterative approach. As a result, the process of DB design becomes more complicated, as not all the requirements are available and future changes cannot always be anticipated.

Hence, the most simple and straightforward approach for the case described in the paper would be to place all the data in the same table. Such an approach provides a reasonable structure for available data, since only a few fields could be skipped in the events in the system. In addition, such an approach should deliver a good performance for the expected queries. The structure for such a table is shown in Fig. 10.

[15] https://www.postgresql.org/.

Fig. 10 Structure of the table

```
CREATE TABLE events (
  id SERIAL PRIMARY KEY,
  ts TIMESTAMP NOT NULL,
  action_id VARCHAR(20) NOT NULL,
  sender_id VARCHAR(10) NOT NULL,
  pallet_id CHAR(13) NOT NULL,
  recipe VARCHAR(5),
  color VARCHAR(5)
);
```

Fig. 11 Organization for query population

```
INSERT INTO events (ts,
    action_id,
    sender_id,
    pallet_id,
    recipe,
    color)
VALUES (?, ?, ?, ?, ?, ?);
```

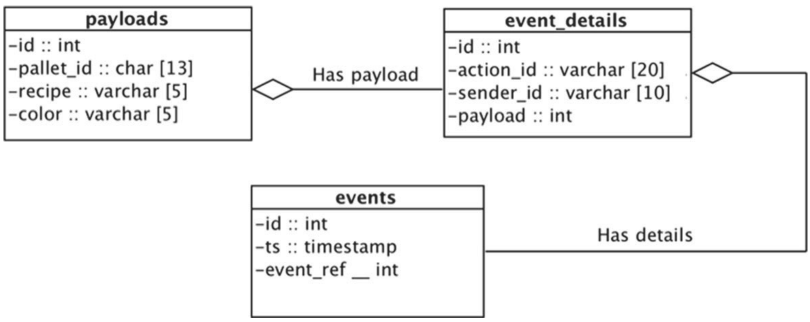


Fig. 12 DB model tables

For such a structure, the population queries are to be organized as shown in Fig. 11. During the population phase, all question marks are to be replaced with the proper values following the same order in the insert command.

Another approach to data modeling used in this research comprises splitting the data according to the content into three separate tables: one defining event payloads, one defining other event description, and one connecting the events to the timestamps. The events table should include a reference to an entry of event details table, which in turn should include a reference to the payload entry. The separation of the data into three tables, shown in Fig. 12, makes the queries more complicated, which leads to higher execution overhead in some cases.

The creation of the three tables is presented in Fig. 13. Once created, the tables are populated by the query as shown in Fig. 14. In a similar manner to the single-table structure, the question marks are replaced by proper values to form the raw data during the population phase.

6 Results and discussion

This section presents the querying benchmark tests performed on different data models (DB and ontology), which have been populated as described in Sects. 5.2 and 5.3. More precisely, three query types have been generated in order to compare the performance of the two data

```
CREATE TABLE payloads (
  id SERIAL PRIMARY KEY,
  pallet_id CHAR(13) NOT NULL,
  recipe VARCHAR(5),
  color VARCHAR(5)
);

CREATE TABLE event_details (
  id SERIAL PRIMARY KEY,
  action_id VARCHAR(20) NOT NULL,
  sender_id VARCHAR(10) NOT NULL,
  payload INT NOT NULL REFERENCES payloads(id)
);

CREATE TABLE events (
  id SERIAL PRIMARY KEY,
  ts TIMESTAMP NOT NULL,
  event_ref INT NOT NULL REFERENCES event_details(id)
);
```

Fig. 13 Creation of DB tables

```
WITH payload AS (
     INSERT INTO payloads (pallet_id, recipe, color)
     VALUES (?, ?, ?)
     RETURNING id
), event_dtls AS (
     INSERT INTO event_details (payload, action_id, sender_id)
     SELECT id, ?, ?
     FROM payload
     RETURNING id
)
INSERT INTO events (ts, event_ref)
     SELECT ?, id FROM event_dtls;
```

Fig. 14 Query to be executed for population

modeling technologies. Since the tests depend on the technology implementation, this may affect the overall results. Nevertheless, the tools employed for these experiments are among the most frequently used by data model designers, and so the results are reflective of common implementation performance.

The design of the three queries provides similar functionality for both DB and knowledge base (KB, ontology) implementations. These queries are presented in Tables 3, 4 and 5.

- The first query counts the number of triggered events for a given event ID.
- The second query counts the number of products that were produced using a specific pallet, which is filtered using the desired pallet ID.
- The third query returns all events that have been triggered in a specific period by giving the start and end timestamps.

A performance analysis of both data modeling technologies is achieved through the execution of the aforementioned queries. In the case of DBs, as discussed in the previous section, two different data models have been designed in order to study the effect of data structure on the performance. The first model considers the data to be stored as one single table, whereas the second approach stores the data in three tables (payloads, event_details and events). This difference in the database structure requires different queries for each functionality. Table 3 presents the three types of query for a single-table database. Although it is simpler to build a

Table 3 DB single-table queries

Query type	Query statement

Count events types	`SELECT events.action_id,`
	`COUNT(events.action_id)`
	`FROM events;`
	`GROUP BY events.action_id;`

Count products made on pallet	`SELECT events.pallet_id,`
	`COUNT(events.pallet_id)`
	`FROM events`
	`WHERE events.action_id ='DRAW_END_EXECUTION'`
	`GROUP BY events.pallet_id;`

Count events in time scope	`SELECT events.sender_id,`
	`COUNT(events.sender_id)`
	`FROM events`
	`WHERE events.ts`
	`BETWEEN '2017-08-03 11:00:00.000'::timestamp`
	`AND '2017-08-03 11:15:00.000'::timestamp`
	`GROUP BY events.sender_id;`

query for a single table, this could decrease the flexibility should the data structure change. The query "Count events types" allows the counting of the number of events for a certain event ID. It returns the list of event_id and the count of appearance for each event. The second row shows the count of the products that have been in a certain pallet. Finally, the third query counts the number of events within a given time window.

Similar to the single-table data structure, the three-table DB structure uses the same queries but with some changes, as shown in Table 4.

The queries for the KB are presented in Table 5; it is apparent that the structure of the queries is somewhat different from the DB queries. This is due to the different syntax of the languages that are used for querying the distinct data models. Nonetheless, the queries maintain, at some level, the same structure since they are used to extract the same information.

As mentioned in Sect. 4, the JMH framework has been used to create benchmarking tests. This framework allows programmers to examine the performance of Java routines. These tests measure the performance of a Java routine by executing the routines successively within a preconfigured period, which is known as iteration. Next, the programmer configures the framework to run the iteration several times. Depending on the configuration, the programmer can allocate some of the iterations to the warmup iteration. During the measurement, the JMH returns the number of times that the routine is executed during the specified period for each iteration. Then, once the framework completes all benchmarking tests, the average value of each benchmark across all iterations is obtained. A higher number indicates a better performance of the tested routine.

To achieve a satisfactory and relevant result for the objectives of this research, the JMH framework has been configured to run ten warmup iterations and ten benchmarking iterations taking one second for each iteration—this means that the unit for the results are operations per second (Ops/s). For the test routines, four benchmarking tests have been created for each model. Three routines execute the queries listed in Tables 3, 4 and 5. In addition, another

Table 4 DB three-table queries

Query type	Query statement

ount events types	`SELECT DISTINCT action_id, COUNT(action_id)`
	`FROM event_details;`
	`GROUP BY action_id;`

Count products made on pallet	`SELECT payloads.pallet_id, COUNT(payloads.pallet_id)`
	`FROM event_details, payloads`
	`WHERE event_details.action_id = 'DRAW_END_EXECUTION'`
	`AND payloads.id = event_details.payload`
	`GROUP BY payloads.pallet_id;`

Count events in time scope	`SELECT event_details.sender_id, COUNT(event_details.sender_id)`
	`FROM events, event_details`
	`WHERE events.ts`
	`BETWEEN '2017-08-03 11:00:00.000'::timestamp`
	`AND '2017-08-03 11:15:00.000'::timestamp`
	`AND event_details.id = events.event_ref`
	`GROUP BY event_details.sender_id;`

routine executes the populating routine for the DB and KB models in Figs. 9, 11, and 14, correspondingly. These benchmarking tests have been deployed in Docker containers to avoid disturbances from different OS processes. For this purpose, three Docker containers have been created and deployed, one for each model. For the single-table and three-table DB models, the measurements were similar but a slightly better performance is shown by the three tables. It is noticeable that there are, in both cases, drops in performance that could be related to the DB query cache feature.[16] In this context, MySQL caches the results of SELECTED queries by default in order to improve the performance. Due to conditions such as the complexity of the query, size of data, size of the results and hardware capability, this cache can be flushed or cleared automatically by the optimizer to prevent any possible errors. On the other hand, the Population benchmarking test showed greater variation from normal performance, as illustrated in both Figs. 15 and 16.

In the ontology model case, the performance is different from that of the DB, as presented in Fig. 17. During testing of the three queries routines, the warmup iteration showed incremental performance improvement until it reached a saturation value. Afterwards, the warmup continued at a steady performance. This steady performance was present in the benchmarking measurements as well. The dramatic difference between the performance results of ontologies compared with DBs is due to several technological and tool-based factors. The main reason for such a difference could be the buffering of the queries in the Jena ARQ,[17] since KB do not repeat the existing instance. This allows for caching of the paths between nodes. Unlike the query tests, the population was similar to the DB population benchmarking results.

After finishing all the benchmarking tests, the JMH framework presented the overall results, which are presented in Fig. 18 and Table 6. As shown, the results show a very distinct

[16] https://dev.mysql.com/doc/refman/5.5/en/query-cache.html.

[17] https://docs.oracle.com/database/122/RDFRM/rdf-suipport-for-apache-jena.htm#RDFRM246.

Table 5 KB queries

Query type	Query statement

Count events types	`PREFIX ont: <http://www.tut.fi/en/fast/FASToryEventOnt#>`
	`SELECT DISTINCT ?Event (count(?Event) as ?Count`
	`WHERE {`
	`?Event ont:hasTimestamp ?Timestamp.`
	`}`
	`GROUP BY ?Event`

Count products made on pallet	`PREFIX ont: <http://www.tut.fi/en/fast/FASToryEventOnt#>`
	`SELECT ?Pallet (count(?Pallet) as ?FinishedProducts)`
	`WHERE {`
	`ont:DRAW_END_EXECUTION ont:hasTimestamp ?Timestamp.`
	`?Timestamp ont:hasPayload ?Payload.`
	`?Payload ont:hasPalletId ?Pallet.`
	`}`
	`GROUP BY ?Pallet`

Count events in time scope	`PREFIX ont: <http://www.tut.fi/en/fast/FASToryEventOnt#>`
	`PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>`
	`SELECT ?Sender (count(?Event) as ?Count)`
	`WHERE {`
	`?Event ont:hasTimestamp ?Timestamp.`
	`?Timestamp ont:hasPayload ?Payload.`
	`?Payload ont:hasSender ?Sender.\n”`
	`Filter (STR(?Timestamp) >=`
	`“http://www.tut.fi/en/fast/FASToryEventOnt#1501747200000”).`
	`Filter (STR(?Timestamp) <=`
	`“http://www.tut.fi/en/fast/FASToryEventOnt#1501748100000”)).`
	`}`
	`GROUP BY ?Sender`

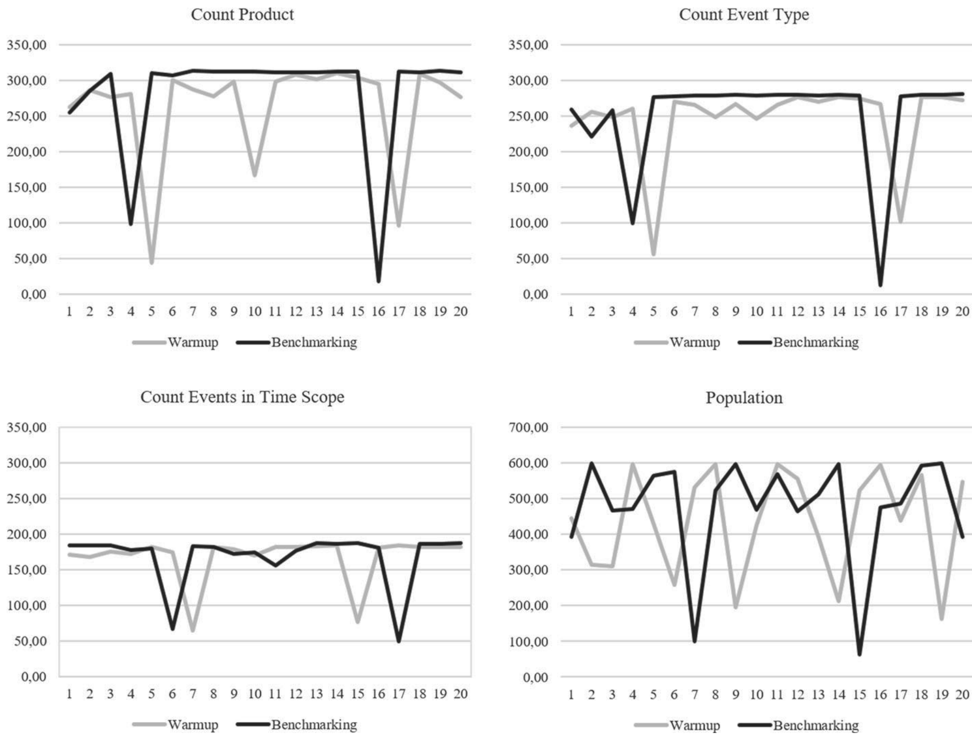


Fig. 15 Single-table DB benchmarking measurements

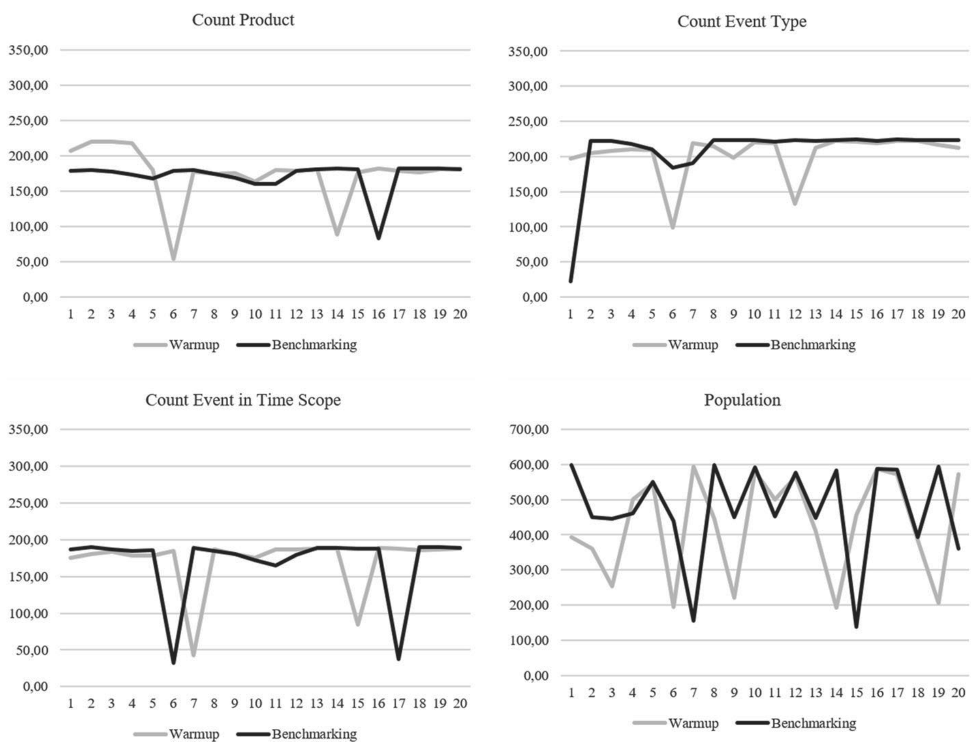


Fig. 16 Three-table DB benchmarking measurements

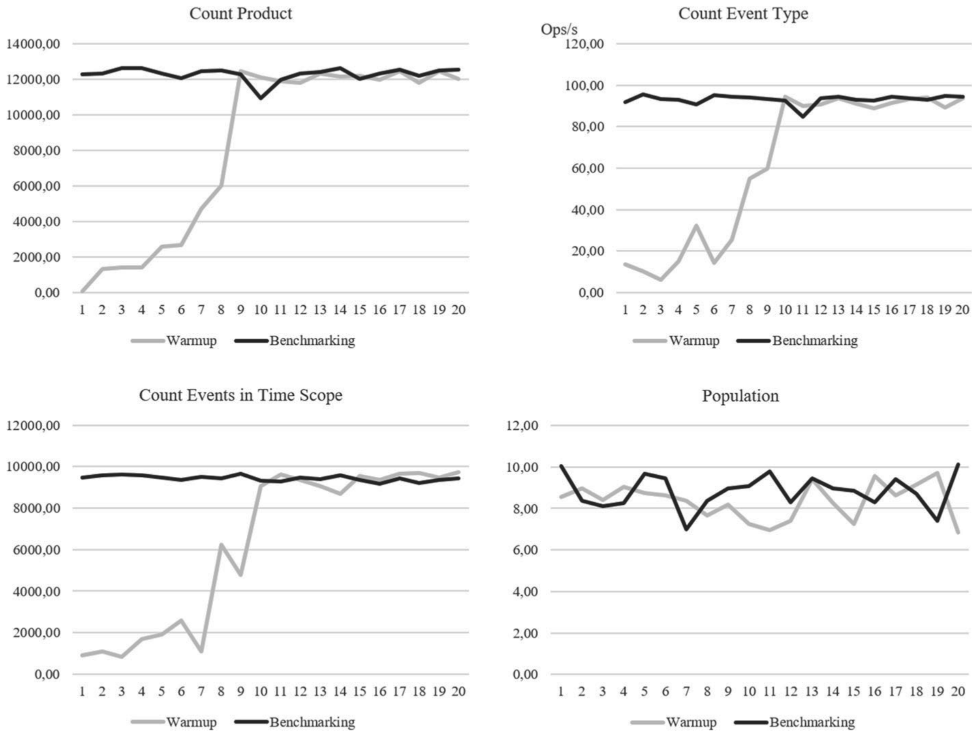


Fig. 17 KB benchmarking measurements

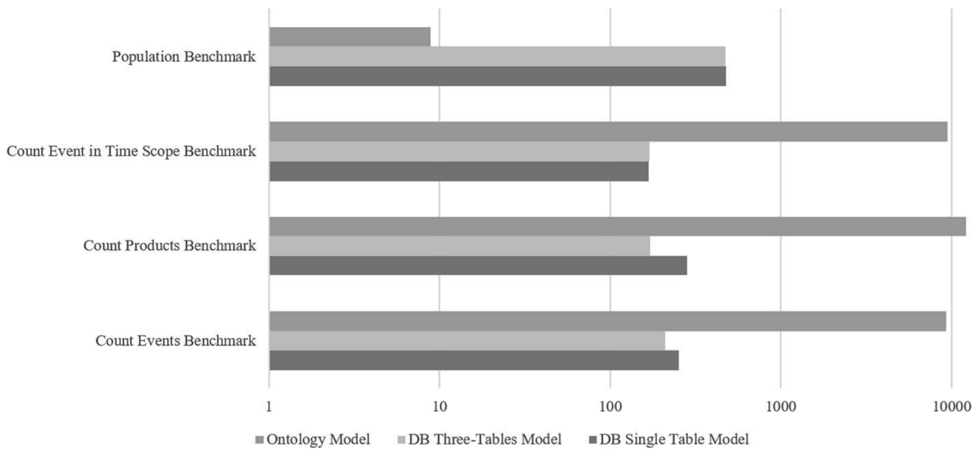


Fig. 18 Overall results for all benchmarking tests

performance difference between updating and querying the models. In the case of population benchmarking, the DB shows much better performance, with an average difference of 200 Ops/s. This difference could be generated by the nature of the technology, since the DB inserts data directly without extra mapping operations as the DB table is constructed. In contrast, the ontology update operation requires a mapping process in order to not duplicate the instances of the classes.

Table 6 Overall average benchmarking results

Test name	DB single-table model [Ops/s]	DB three-table model [Ops/s]	KB [Ops/s]
Count events benchmark	251.727	208.406	931.421
Count products benchmark	282.039	171.624	12291.206
Count event in time scope benchmark	168.307	169.658	9437.583
Population benchmark	475.498	473.017	8.831

Furthermore, the ontology model had the upper hand in the data retrieval process with a difference of approximately 9000 Ops/s. This difference could be caused by the caching feature, provided by the JENA ARQ, which caches the path between nodes for a querying operation—providing better performance. In this research, these parameters were kept at the default configuration settings to represent a common client trying to use the tool directly from the box. It is important to highlight that these tests depend on the search engine for each technology; therefore, the specific indexing algorithm should be further analyzed. These results can be compared with other research that addressed the same problem; however, the comparison can be unfair for some technologies since the tests and experimental techniques might vary, with different configurations and parameters. As an example, while the BSBM in [60] used an HTTP interface to connect with the JENA TDB; this research interfaces with the JENA TDB using the Java API (application programming interface) directly. This could affect the test results, where HTTP services can add latency to the system.

This section has presented the experimental results that have enabled a performance comparison of DB and ontology technologies. On the performance level, the tests intend to eliminate the technology effects, as close as possible, by following the same test conditions and using the same OS. However, as the vision was to try to deploy these tests with the most commonly used tools available in order to reflect the real-world scenario, this has been a challenging experiment. Instead of comparing both technologies, it is important to highlight that both technologies may work side by side—each providing unique features for the user. As an example, from the tests, ontologies should perform better than DBs at querying processes, which makes it a more reasonable choice to be used as a knowledge provider. On the other hand, DBs show more consistent performance for both querying and updating, which makes them suited for use as a data store. In addition, an ontology-based model permits more rich representation of the data. This fact suggests that ontologies would be a better choice for applications that require reasoning and for inferencing implicit knowledge of the data model.

With regard to RQ2, the boundaries between databases and knowledge-base technologies must be investigated. Due to the involvement of the authors of this research in several EU projects, they have experience of the evolution of both data modeling technologies (DBs and ontologies) with new concepts. As an example, the Cloud Collaborative Manufacturing Networks (C2NET)[18] project includes both technologies within the same solution, in order to exploit different features of each one. In this context, the C2NET project provides key functionalities for the smart and medium-sized enterprises (SMEs) on the enterprise resource planning (ERP) level in the well-known automation pyramid described in the ISA-95 standard [73]. These features, which are provided as web services in the C2NET project, include i) optimization, ii) monitoring and iii) assessment of production, delivery and logistics plans. Besides this, the C2NET platform also allows the companies to interact with other companies

[18] https://cordis.europa.eu/project/id/636909.

in the same supply chain, acting as a network for the exchange of information and facilitating communicating through the web.

In regard to its architecture, and related to the synergy of databases and ontologies, the C2NET platform employs both technologies. The data collected by the C2NET platform from SMEs—potentially ERP or factory shop-floor data—pass through a transformation process where it is homogenized with the schemas or standards of the C2NET platform data. The transformation is applied within the ontology technology since each company can provide the data in a different schema or format. The C2NET platform then uses the database technology to manage the transformed data before it is utilized by the aforementioned features.

Similar to C2NET, and as described in Sects. 5.2 and 5.3, both DB- and KB (ontology)-based approaches may work in conjunction and support each other since each one might have it is own specialized role to play. Although it is discussed in the following section, at first glance it can be argued that the knowledge base provides more flexibility and adjustability for the data format, whereas the database provides better performance and robustness to the system.

7 Conclusions and further work

The core objective of this research was to compare two data modeling approaches that are used in the context of PLM and PPR: ontologies and databases. In order to achieve this, the authors explored the literature to identify what work existed in this area. The knowledge gap that was identified gives rise to a limited comparison of the two technologies for common applications with limited quantitative and qualitative analyses. To address this gap, three RQs were synthesized:

– RQ1 focused on understanding how the data modeling approaches performed as data volume increased. This is important because databases are currently used extensively in industrial environments, handling large volumes of data, and it is necessary to understand how the ontological approach compares. The authors worked to create data models in a way which enabled a fair comparison with benchmarks that presented data on the following: event counting, product counting, event-in-time counting, and data model population. The results found that the ontology performed more than three times better in the event counting benchmarks, and orders of magnitude better than databases in all other test—apart from the population test. It is proposed that this is due to the fact that when an instance is created in an ontology, the respective mappings must also be created based on rules, which is not necessarily the case for databases (accounting for the poor performance of ontologies in executing population tasks). By comparison, once the instance exists, it is much easier to access and manipulate it in an ontological model than in a database due to the benefits that these mappings provide. In addition, it is important to keep in mind the effect of the tools used for such a study. The tool itself might play an important role since each tool is supported by optimization algorithms to enhance the performance. This could be the subject of further study to gain insight into the potential optimization process.
– RQ2 considered the idea that, ultimately, databases and ontologies have been developed for two quite distinct purposes and it is therefore necessary to understand how these respective technologies may complement each other. Drawing on experience from previous projects (e.g., [6,74,75]), the authors describe a scenario where myriad standards are homogenized using an ontological model and the data then instantiated within a database.

This type of complementary working addresses an environment where there may be a need to realize interoperability between heterogeneous software. Such an environment is typical of a manufacturing system. The results do demonstrate, however, that the population rate of an ontological model gives cause for concern in a high-data-volume environment. Given the significant differences in performance between ontologies and databases, it would be of value to investigate if the high-speed data instantiation of databases could be brought into a system where the high-speed querying of ontologies could be exploited.
– RQ3 aimed to examine the maintainability of databases in comparison with ontologies, but was not directly addressed in this work. This question was included as the authors appreciate the need to examine the respective technologies holistically and therefore require a lifecycle assessment from system design, through to implementation and then reconfiguration—this is true for both physical and digital systems. This article lacks a study that evaluates the maintenance efforts, although the authors do shed some light on the creation of the respective models. To address this, the authors are working on a further piece of research to introduce a change to system requirements and to assess the efforts required to realize them. This proposed study will consider the hypothesis that the perceived difference in effort for implementing and maintaining the data models will depend significantly on the engineer's familiarity with the respective technology, even though design tools can abstract users from fully understanding the model syntax.

This research concludes that ontologies and databases should not replace, but rather complement each other. The experiments show that both technologies present differences in their performance and that the decision for using one instead of the other will depend on the implementation and application. Nevertheless, it may be seen that the experiments presented do not allow the full exploitation of ontologies, due to the low expressivity of event information. To address this, the authors will further increase the complexity of event content—enabling the demonstration of other features such as implicit knowledge inference.

In summary, the work presented in this research contributes significantly to the body of knowledge by:

– Developing a methodology for comparing data modeling approaches.
– Quantitatively comparing ontological models and databases with a view to understanding how data volume affects performance.
– Considering how databases and ontologies may complement each other in the future and the scenarios in which they exist in a system whereby their whole is greater than the sum of their parts.

Acknowledgements The research work by Wael M. Mohammed is supported by the Doctoral School of the President call 2018 of the Tampere University of Technology. Also, the authors gratefully acknowledge the support of the graduate school funding of Tampere University and the UK EPSRC under the iCASE Ph.D. studentship (1377735) in carrying out this work.

References

1. IEC (2021) IEC White Paper Future Factory:2015 | IEC Webstore | water automation, water management, smart city. https://webstore.iec.ch/publication/23389
2. Delamer IM, Lastra JLM (2006) Service-oriented architecture for distributed publish/subscribe middleware in electronics production. IEEE Trans Ind Inform 2(4):281–294. https://doi.org/10.1109/TII.2006.885188
3. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M (2010) A view of cloud computing. Commun ACM 53(4):50–58. https://doi.org/10.1145/1721654.1721672
4. Liu C, Ding J, Toprac AJ, Chai T (2014) Data-based adaptive online prediction model for plant-wide production indices. Knowl Inf Syst 41(2):401–421. https://doi.org/10.1007/s10115-014-0757-8
5. Stark J (2011) Product lifecycle management: 21st century paradigm for product realisation, 2nd edn. Springer, London. https://doi.org/10.1007/978-0-85729-546-0
6. Ahmad M, Ferrer BR, Ahmad B, Vera D, Martinez Lastra JL, Harrison R (2018) Knowledge-based PPR modelling for assembly automation. CIRP J Manuf Sci Technol 21:33–46. https://doi.org/10.1016/j.cirpj.2018.01.001
7. Verborgh Ruben, Sande Miel Vander (2020) The Semantic Web identity crisis: in search of the trivialities that never were. Seman Web J 11(1):19–27
8. Ramis Ferrer B, Ahmad B, Vera D, Lobov A, Harrison R, Martínez Lastra JL (2016) Product, process and resource model coupling for knowledge-driven assembly automation. Automatisierungstechnik. https://doi.org/10.1515/auto-2015-0073
9. Martinez-Cruz C, Blanco IJ, Vila MA (2012) Ontologies versus relational databases: are they so different? A comparison. Artif Intell Rev 38(4):271–290. https://doi.org/10.1007/s10462-011-9251-9
10. Sir M, Bradac Z, Fiedler P (2015) Ontology versus database. IFAC-PapersOnLine 48(4):220–225. https://doi.org/10.1016/j.ifacol.2015.07.036
11. Schmidt D (2006) Guest editor's introduction: model-driven engineering. Computer 39(2):25–31. https://doi.org/10.1109/MC.2006.58
12. Michalos G, Makris S, Papakostas N, Mourtzis D, Chryssolouris G (2010) Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. CIRP J Manuf Sci Technol 2(2):81–91. https://doi.org/10.1016/j.cirpj.2009.12.001
13. Fogliatto FS, da Silveira GJC, Borenstein D (2012) The mass customization decade: an updated review of the literature. Int J Prod Econ 138(1):14–25. https://doi.org/10.1016/j.ijpe.2012.03.002
14. Lee J, Lapira E, Bagheri B, Kao Ha (2013) Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lett 1(1):38–41. https://doi.org/10.1016/j.mfglet.2013.09.005
15. Crescenzi V, Fernandes AAA, Merialdo P, Paton NW (2017) Crowdsourcing for data management. Knowl Inf Syst 53(1):1–41. https://doi.org/10.1007/s10115-017-1057-x
16. Kilintzis V, Beredimas N, Chouvarda I (2014) Evaluation of the performance of open-source RDBMS and triplestores for storing medical data over a web service. In: 2014 36th annual international conference of the IEEE Engineering in Medicine and Biology Society, IEEE, Chicago, IL, pp 4499–4502. https://doi.org/10.1109/EMBC.2014.6944623
17. Lin HK, Harding JA, Shahbaz M (2004) Manufacturing system engineering ontology for semantic interoperability across extended project teams. Int J Prod Res 42(24):5099–5118. https://doi.org/10.1080/00207540412331281999
18. Staab S, Studer R (eds) (2009) Handbook on ontologies. International handbooks on information systems, 2nd edn. Springer, Berlin. https://doi.org/10.1007/978-3-540-92673-3
19. Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5(2):199–220. https://doi.org/10.1006/knac.1993.1008
20. Borst P, Akkermans H, Top J (1997) Engineering ontologies. Int J Hum Comput Stud 46(2–3):365–406. https://doi.org/10.1006/ijhc.1996.0096
21. Zdravković M, Panetto H, Trajanović M, Aubry A (2014) Explication and semantic querying of enterprise information systems. Knowl Inf Syst 40(3):697–724. https://doi.org/10.1007/s10115-013-0650-x
22. Kalibatiene D, Vasilecas O (2011) Survey on ontology languages. In: Grabis J, Kirikova M (eds) Perspectives in business informatics research. Springer, Berlin, pp 124–141
23. Bechhofer S, van Harmelen JHF (2021) OWL web ontology language reference. https://www.w3.org/TR/owl-ref/
24. Eric Prud hommeaux AS (2021) SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query/
25. Paula Gearon AP Alexandre Passant (2021) SPARQL 1.1 Update. https://www.w3.org/TR/sparql11-update/

26. Horrocks I, Patel-Schneider P, Boley H, Tabet S, Grosof B, Dean M (2007) SWRL: a semantic web rule language combining oWL and ruleML. W3C Member submission 21
27. Puttonen J, Lobov A, Lastra JLM (2013) Maintaining a dynamic view of semantic web services representing factory automation systems. In: 2013 IEEE 20th international conference on web services, pp 419–426. https://doi.org/10.1109/ICWS.2013.63
28. Efthymiou K, Sipsas K, Mourtzis D, Chryssolouris G (2015) On knowledge reuse for manufacturing systems design and planning: a semantic technology approach. CIRP J Manuf Sci Technol 8:1–11. https://doi.org/10.1016/j.cirpj.2014.10.006
29. Delamer IM, Lastra JLM (2006) Ontology modeling of assembly processes and systems using semantic web services. In: 2006 4th IEEE international conference on industrial informatics, pp 611–617. https://doi.org/10.1109/INDIN.2006.275631
30. Usman Z (2012) A manufacturing core concepts ontology to support knowledge sharing. thesis, Loughborough University, http://articles/thesis/A_manufacturing_core_concepts_ontology_to_support_knowledge_sharing/9524204/1
31. Borgo S, Leitão P (2007) Foundations for a core ontology of manufacturing. In: Sharman R, Kishore R, Ramesh R (eds) Ontologies: a handbook of principles, concepts and applications in information systems, integrated series in information systems. Springer, Boston, pp 751–775. https://doi.org/10.1007/978-0-387-37022-4_27
32. Miller GA (1995) WordNet: a lexical database for English. Commun ACM 38(11):39–41. https://doi.org/10.1145/219717.219748
33. Pratt MJ (2001) Introduction to ISO 10303: the STEP standard for product data exchange. J Comput Inf Sci Eng 1(1):102–103. https://doi.org/10.1115/1.1354995
34. Dartigues C, Ghodous P, Gruninger M, Pallez D, Sriram R (2007) CAD/CAPP integration using feature ontology. Concurr Eng 15(2):237–249. https://doi.org/10.1177/1063293X07079312
35. Krima SI, Barbau R, Fiorentini X, Rachuri S, Sriram RD (2009) OntoSTEP: OWL-DL Ontology for STEP. NIST Interagency https://www.nist.gov/publications/ontostep-owl-dl-ontology-step, last Modified: 2017-02-19T20:02-05:00
36. Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L (2002) Sweetening ontologies with DOLCE. In: Gómez-Pérez A, Benjamins VR (eds) Knowledge engineering and knowledge management: ontologies and the semantic web. Lecture notes in computer science. Springer, Berlin, pp 166–181. https://doi.org/10.1007/3-540-45810-7_18
37. Ontology BF (2021) Basic Formal Ontology (BFO) | Home. http://basic-formal-ontology.org/
38. Deshayes L, Foufou S, Gruninger M (2007) An ontology architecture for standards integration and conformance in manufacturing. In: Tichkiewitch S, Tollenaere M, Ray P (eds) Advances in integrated design and manufacturing in mechanical engineering II. Springer, Dordrecht, pp 261–276. https://doi.org/10.1007/978-1-4020-6761-7_18
39. Li Q, Chen YL (2009) Ontology capture methods. In: Li Q, Chen YL (eds) Modeling and analysis of enterprise and information systems: from requirements to realization. Springer, Berlin, pp 227–243. https://doi.org/10.1007/978-3-540-89556-5_12
40. Peraketh B, Menzel C, Mayer R, Fillion F, Futrell M (1994) Ontology capture method (IDEF5). Armstrong Laboratory, Arlington
41. Uschold M, Gruninger M (1996) Ontologies: principles, methods and applications. Knowl Eng Rev 11(2):93–136. https://doi.org/10.1017/S0269888900007797
42. Fernández-López M, Gómez-Pérez A, Juristo N (1997) METHONTOLOGY: from ontological art towards ontological engineering. In: Proceedings of the ontological engineering AAAI-97 Spring symposium series, Facultad de Informática (UPM), Stanford University, EEUU. http://oa.upm.es/5484/
43. Noy N, Mcguinness D (2001) Ontology development 101: a guide to creating your first ontology. Knowl Syst Lab 32
44. Fokou G, Jean S, Hadjali A, Baron M (2017) Handling failing RDF queries: from diagnosis to relaxation. Knowl Inf Syst 50(1):167–195. https://doi.org/10.1007/s10115-016-0941-0
45. Angles R, Gutierrez C (2008) Survey of graph database models. ACM Comput Surv 40(1):1–39. https://doi.org/10.1145/1322432.1322433
46. Elmasri R, Navathe S (2016) Fundamentals of database systems, 7th edn. Pearson, Hoboken
47. Gulutzan P, Pelzer T (1999) SQL-99 complete, really: an example-based reference manual of the new standard. R&D Books, Lawrence
48. Sadalage PJ, Fowler M (2012) NoSQL distilled a brief guide to the emerging world of polyglot persistence. Addison-Wesley, Upper Saddle River
49. Taylor RW, Frank RL (1976) CODASYL data-base management systems. ACM Comput Surv 8(1):67–103. https://doi.org/10.1145/356662.356666

50. IBM (2005) IMS - Introduction - History of IMS: Beginnings at NASA. https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.imsintro.doc.intro/ip0ind0011003710.htm
51. Yannakoudakis EJ (1988) Database design methodology. In: Yannakoudakis EJ (ed) The architectural logic of database systems. Springer, London, pp 135–161. https://doi.org/10.1007/978-1-4471-1616-5_6
52. Teorey TJ, Yang D, Fry JP (1986) A logical design methodology for relational databases using the extended entity-relationship model. ACM Comput Surv 18(2):197–222. https://doi.org/10.1145/7474.7475
53. Teorey T, Lightstone S, Nadeau T, Jagadish HV, Safari OMC (2011) Database modeling and design, 5th edn. Morgan Kaufmann, Burlington
54. Gani A, Siddiqa A, Shamshirband S, Hanum F (2016) A survey on indexing techniques for big data: taxonomy and performance evaluation. Knowl Inf Syst 46(2):241–284. https://doi.org/10.1007/s10115-015-0830-y
55. Konstantinou N, Spanos DE, Mitrou N (2008) Ontology and database mapping: a survey of current implementations and future directions. J Web Eng (JWE) 7:1–24
56. Choi SS, Yoon TH, Noh SD (2010) XML-based neutral file and PLM integrator for PPR information exchange between heterogeneous PLM systems. Int J Comput Integr Manuf 23(3):216–228. https://doi.org/10.1080/09511920903443234
57. Gunpinar E, Han S (2008) Interfacing heterogeneous PDM systems using the PLM Services. Adv Eng Inform 22(3):307–316. https://doi.org/10.1016/j.aei.2007.08.009
58. El Kadiri S, Kiritsis D (2015) Ontologies in the context of product lifecycle management: state of the art literature review. Int J Prod Res 53(18):5657–5668. https://doi.org/10.1080/00207543.2015.1052155
59. Kitamura Y, Koji Y, Mizoguchi R (2006) An ontological model of device function: industrial deployment and lessons learned. Appl Ontol 1(3–4):237–262
60. Bizer C, Schultz A (2009) The Berlin SPARQL benchmark. Int J Semant Web Inf Syst 5(2):1–24. https://doi.org/10.4018/jswis.2009040101
61. M Li, X-Y Du, S Wang (2005) Learning ontology from relational database. In: 2005 International conference on machine learning and cybernetics, vol 6, pp 3410–3415. https://doi.org/10.1109/ICMLC.2005.1527531
62. Cullot N, Ghawi R, Yetongnon K (2007) DB2OWL: a tool for automatic database-to-ontology mapping. In: Proceedings of the fifteenth Italian symposium on advanced database systems, p 494
63. Jayakumar P, Shobana P (2014) Creating ontology based user profile for searching web information. In: International conference on information communication and embedded systems (ICICES2014), pp 1–6. https://doi.org/10.1109/ICICES.2014.7033893
64. Barrasa J, Corcho O, Gomez perez A (2004) R2O, an extensible and semantically based database-to-ontology mapping language. In: Proceedings of the 2nd workshop on semantic web and databases(SWDB2004). Springer, pp 1069–1070
65. Laclavík M (2006) RDB2Onto: relational database data to ontology individuals mapping. Information and Knowledge Oriented Technologies Group
66. Ho LTT, Tran CPT, Hoang Q (2015) An approach of transforming ontologies into relational databases. In: Nguyen NT, Trawiński B, Kosala R (eds) Intelligent information and database systems. Lecture notes in computer science. Springer, Cham, pp 149–158. https://doi.org/10.1007/978-3-319-15702-3_15
67. Gali A, Chen CX, Claypool KT, Uceda-Sosa R (2004) From ontology to relational databases. In: Wang S, Tanaka K, Zhou S, Ling TW, Guan J, Yang D, Grandi F, Mangina EE, Song IY, Mayr HC (eds) Conceptual modeling for advanced application domains. Lecture notes in computer science. Springer, Berlin, pp 278–289. https://doi.org/10.1007/978-3-540-30466-1_26
68. Eder J, Haav HM, Kalja A, Penjam J (eds) (2005) Advances in databases and information systems: 9th East European conference, ADBIS 2005, Tallinn, Estonia, 12–15 September 2005. In: Proceedings. Information systems and applications, incl. Internet/Web, and HCI. Springer, Berlin. https://doi.org/10.1007/11547686
69. Trinkunas J, Vasilecas O (2015) A graph oriented model for ontology transformation into conceptual data model. Inf Technol Control. https://doi.org/10.5755/J01.ITC.36.1.11832
70. Lv Y, Xie C (2012) An ontology-based approach to build conceptual data model. In: 2012 9th international conference on fuzzy systems and knowledge discovery, pp 807–810. https://doi.org/10.1109/FSKD.2012.6234141
71. Lee J, Goodwin R (2006) Ontology management for large-scale enterprise systems. Electron Commer Res Appl 5(1):2–15. https://doi.org/10.1016/j.elerap.2005.08.003
72. Saccol DB, Andrade TC, Piveta EK (2011) Mapping OWL ontologies to relational schemas. In: 2011 IEEE international conference on information reuse integration, pp 71–76. https://doi.org/10.1109/IRI.2011.6009523

73. Commission IE (2021) ANSI/ISA-95.00.03-2013 enterprise-control system integration—part 3: activity models of manufacturing operations management. https://www.isa.org/products/ansi-isa-95-00-03-2013-enterprise-control-system-i
74. Ramis Ferrer B, Martinez Lastra JL (2017) Private local automation clouds built by CPS: potential and challenges for distributed reasoning. Adv Eng Inf 32:113–125. https://doi.org/10.1016/j.aei.2017.01.007
75. Iarovyi S, Ramis B, Xiangbin X, Sampath A, Lobov A, Lastra JLM (2015) Representation of manufacturing equipment and services for OKD-MES: from service descriptions to ontology. In: 2015 IEEE 13th international conference on industrial informatics (INDIN), pp 1069–1074. https://doi.org/10.1109/INDIN.2015.7281883

Dr. Borja Ramis Ferrer received the B.Sc. degree in electrical engineering from the Universidad de las Islas Baleares, Islas Baleares, Spain, in 2011 and the M.Sc. degree (with distinction) in Factory Automation from Tampere University of Technology (TUT), Tampere, Finland, in 2013. In addition, he obtained the Dr. Sc. (Tech.) degree in Automation from TUT, in 2018. With more than 8 years of experience in R&D&i in Industrial Informatics, Knowledge Representation (Ontology) and Embedded Systems for Factory Automation, Dr. Ramis contributed to many EU projects. He has (co)authored over 50 scientific papers and his research interests include the deployment of knowledge-based and cyber-physical systems in the fields of Automation, Robotics and Industrial Informatics. Since 2019, he is working with cybersecurity solutions for factories of the future and smart environments.

Wael M. Mohammed is a Doctoral Researcher at Tampere University. He received a M.Sc. in Automation Engineering from Tampere University of Technology in 2017 and a B.Sc. in Mechatronics Engineering from university of Jordan in 2010. As he is pursuing a Doctoral Degree in Engineering Sciences; Mr. Mohammed's research interests include Robotics, Digital Twins, Knowledge-Based Reasoning Engines and Factory Automation. In addition, Mr. Mohammed has been involved in writing proposals for research and innovation project funded by the EU commission. In 2010, he worked as a research assistant in the Production Engineering Department at Tampere University of Technology. Then, in 2011, Mr. Mohammed worked as a head of the technical department in the Traffic Management System project at Etihad Alafandi L.L.C. in Saudi Arabia.

Mussawar Ahmad has a background in Mechanical Engineering and a Ph.D. in Industrial Automation from the University of Warwick. He is currently a Senior Research Engineer at the Manufacturing Technology Centre in the Automation and Robotics team. His current research interests are around the use of knowledge-based to realize the adoption and deployment of intelligent automation systems.

Sergii Iarovyi is a Software Engineer at Kalmar. He received a M.Sc. degree (with Distinction) in Electro-mechanics from the National Technical University "KhPI," Kharkiv, Ukraine, in 2011 and the M.Sc. degree in Factory Automation from Tampere University of Technology, Tampere, Finland, in 2014. He has worked at FAST laboratory, Tampere University of Technology as a Project Researcher. His research interests are in the application of semantic web services, cyber-physical systems, and enterprise integration for factory automation.

Jiayi Zhang is a Software Developer at WMG. He received his Ph.D. in Automation Systems from the Department of WMG, University of Warwick. He received his B.E. degree in Software Engineering and M.Sc. degree in Computing, respectively, from Tianjin Normal University (China) and University of Teesside (UK). His main research interests are semantic ontology predictive modeling, intelligent manufacturing, and knowledge-based system integration. He has worked as an IT engineer, participating in different manufacturing projects related to software development technologies, ontology modeling design, and semantic data analysis. Currently, he is researcher and software developer of the Automation Systems Group at WMG. He has been working closely with the automotive industry on several collaborative research projects focusing on data integration system, automated ontology generation, and audio emotion analysis. His current projects include Smart Seat Manufacturing (Realtime process data analysis), Lear Smart Manufacturing (ProductProcessResource data integration), vueOne auto code generation.

Robert Harrison is Professor of Automation Systems and Head of the Digital Technologies Directorate at WMG. Having joined University of Warwick in March 2013, he was previously Director of Automation Systems at Loughborough University. Current research includes ISCF Manufacturing Made Smarter SIMPLE, DIALOG and DAMPO projects, EU DIGIMAN, and numerous contract research and HVM Catapult projects, with BAE Systems, Lear Corporation, Atlas Copco, and others. Between 2013 and 2018, he led the EPSRC Knowledge Driven Configurable Manufacturing project. From 2005 onwards, his Automation Systems Group were partners on several large EU projects related to serviceoriented distributed systems, including SOCRADES, RIMACS, Arrowhead, and IMCAESOP. Robert received a Royal Academy of Engineering Global Research Award in 2004 with visiting research roles at Schneider Electric and Ford. He was a Research Fellow at NUS Singapore from 1993 to 1995 and Ford Research Fellow from 1996 to 2002. He holds a PhD in automation from Loughborough University.

Jose Luis Martinez Lastra joined Tampere University of Technology in 1997 and became University Full Professor in 2006. His research interest is on applying Information and Communication Technologies to the field of Factory Automation. Prof. Lastra leads the FAST-Lab with the ultimate goal of seamlessly integrating the knowledge of humans and machines. Prof. Lastra has co/authored over 300 scientific papers and holds a number of patents in the field of Industrial Informatics and Automation. He serves as an Associate Editor of the IEEE Transactions on Industrial Informatics, and he served a Technical Editor of the IEEE/ASME Transactions on Mechatronics.

PUBLICATION
IV

Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications

Wael M Mohammed, Rodolfo E Haber, Jose L Martinez Lastra

Machines, 2022, 10 (10), 861
Doi: 10.3390/machines10100861

machines

Article

Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications

Wael M. Mohammed [1,*], Rodolfo E. Haber [2] and Jose L. Martinez Lastra [1]

[1] FAST-Lab, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 600, 33014 Tampere, Finland
[2] Center for Automation and Robotics (CSIC-UPM), 28500 Madrid, Spain
* Correspondence: wael.mohammed@tuni.fi; +358-50-447-9119

Abstract: The rapid emerging technologies in various fields permitted the creation of simulation tools. These tools are designed to replicate physical systems in order to provide faster, cheaper and more detailed illustrative analysis of the physical system. In this regard, the concept of digital twins has been introduced to generally define these simulation tools. In fact, and according to the creator of the digital twin term Micheal Grieves, a digital twin is defined as a physical system, a digital replica of the physical system and information flow between the former parts. This definition is simple and generic for describing digital twins and yet, holistic. This broad definition creates a challenge for developers who target the development of such applications. Therefore, this paper presents a paradigm for architecting digital twins for manufacturing processes. The approach is inspired by the definitions of the ISA95 standard and the onion concept of computer applications to create multi-layer and multi-level concepts. Furthermore, and to satisfy the different required features by industries, the approach considers a multi-perspective concept that allows the separation of the digital twin views based on functionality. This paradigm aims at providing a modular, scalable, reusable, interoperable and composable approach for developing digital twins. Then, an implementation of the approach has been introduced using an ontology-based system and the IEC61499 standard. This implementation has been demonstrated on a discrete manufacturing assembly line.

Keywords: digital twins; Industry 4.0; systematic method; simulation; generic approach; industrial application

Citation: Mohammed, W.M.; Haber, R.E.; Martinez Lastra, J.L. Ontology-Driven Guidelines for Architecting Digital Twins in Factory Automation Applications. *Machines* **2022**, *10*, 861. https://doi.org/10.3390/machines10100861

Academic Editor: Zhuming Bi

Received: 24 August 2022
Accepted: 19 September 2022
Published: 26 September 2022

1. Introduction

Simulation applications have become necessary tools for engineers and systems' developers to maximize the performance of these systems in recent years. These tools provide low-cost, safe and effective approaches for testing, validating and monitoring various systems [1]. In addition, simulation tools or simulators are intensively employed in training operators and human workers on expensive machines and equipment and critically sensitive processes and operations. Examples of simulators include airplanes, Rubber-tiered Gantries (RTG)s, construction cranes, manufacturing and production systems, robotics, and many others. Historically, the creation of the concept of simulation is unknown. Nonetheless, the concept of simulation has been reported in the 18th century when the midwife Angélique du Coudray created a full-scale replica of an obstetrical mannequin for practicing the parturition process [2]. At that time, the simulation was only performed in the physical world. In recent decades, simulation tools have become more effective as computer and visual technologies continue prospering. For instance, newly developed immersive technologies such as Virtual Reality (VR) and Augmented Reality (AR) have been introduced in many fields to provide a better experience for the user during the training on working closely with robots [3]. This evolution is surely driven by the advancement in the computer science and communication domains.

In the manufacturing and industrial domain, the concept of simulation has been recently introduced in the term Digital Twins (DT) for products and processes. The term Digital Twins was firstly used by Michael Grieves in 2002 [4]. According to J. David et al. in [4], the initial definition of DTs by Michael Grieves in his white paper [5] considers the DT as a virtual representation of a product from a Product Lifecycle Management (PLM) perspective. Furthermore, the authors in [4] indicated that Mr. Grieves illustrated the DT as a three-component system: the product, the virtual representation of the product, and the virtual-physical connection between the product and the virtual replica. This definition has been debated and modified by several scholars according to the publications in [6–9]. For processes, the concept of DT is similar to the product DT but the focus is on the execution and the processes rather than the lifecycle.

The mainchallenge that faces developers and researchers when studying the concept of DTs is the definition of a digital twin itself. For instance, are DTs considered to be simulation units or monitoring systems? Do DTs require 3D visualization interfaces? Do DTs require human-friendly interactions? These questions and many others formulate the purpose of DTs. In this regard, this paper presents a paradigm for architecting Digital Twins for manufacturing processes. The main focus in this paper is ignoring the precise definition of the DT and considering it as a data consumer and information generator as highlighted in [5,10–12]. After accepting that, this generic definition can be tailored to suit the application. In more detail, the main contributions of this paper include:

- Conducting research for analyzing the techniques and the technologies that are adapted for developing and building digital twins using academic and commercial solutions as information sources.
- Reusing the available standards and techniques in manufacturing, computer science and industrial management domains for reshaping a generic paradigm for DTs.
- Presenting a guideline or paradigm for architecting a digital twin.
- Presenting an implementation example for architecture.

The rest of the article is structured as follows: Section 2 presents a literature review on related topics. More precisely, it is divided into a review of academic results and commercial solutions. In addition, a brief review is focused on DTs in the context of modern aspects such as Industry 4.0. Section 3 presents the generic paradigm and illustrates the facts behind the selected approach. The fourth section introduces the systematic methods of building a DT. Then, Section 5 depicts a use case using an educational and research assembly line. The Section 6 provides a discussion and the authors' view on the concept. Finally, Section 7 concludes the paper and provides the possible future work.

2. Digital Twins: Literature and Commercial Solutions Review

In order to understand the techniques and methods of developing and building digital twins, a review must be conducted. Thankfully, several surveys and systematic reviews have been provided by the scientific community. In addition to scientific outcomes, the commercially available digital twins represent invaluable references for this research. Moreover, the available standards, such as the ISO-23347, include knowledge regarding the development of such systems. Therefore, this section aims at highlighting the methods and techniques that are reported in the literature regarding the creation of digital twins.

2.1. Literature Review on Designing and Developing Digital Twins

The usage of the term "Digital Twin" has been increasing rapidly in the academic community with the manufacturing domain as the major contributor [4,7,13,14]. This leap in the research of DTs is mainly related to the demand by the industrial sector and due to the advances in the Internet and Communication Technologies (ICT) domain [15,16]. In other words, the high computational capabilities, the highly interconnected devices, and the efficient data analysis techniques permit the development and deployment of such intelligent applications.

Therefore, several approaches have been introduced for building DTs. This is clear in the well-presented article by Chiara et al. [17]. As the paper discusses the application of DT in manufacturing, the authors conducted an adequate analysis on the DT for determining the direction of the current research for developing DTs. Among several aspects, the paper focuses on the used software, the nature of the user interface, the communication protocols and the provided services. The majority of the compared DTs share similar features which include data acquisition, 3D modeling, real-time monitoring, and analysis for optimization. This means the development of a DT must commonly satisfy these features which are aligned with the generic definition of DTs. Some examples on that, Zhang et al. [18] present a dynamic scheduling application based on digital twin agents for scheduling and optimizing workshop tasks. Fang et al. in [19] present research on scheduling job shop tasks using digital twins.

Resman et al. in[20] present an approach for developing an architecture model for planning manufacturing processes. In fact, the aforementioned examples employ AI-based techniques for building digital twin tools that allow optimization and scheduling of processes and tasks. This is considered to be one of the most important utilizations of a digital twin.

Looking further for more academic research on building DTs, different methods, techniques, and strategies have been utilized for developing and deploying activities. As an example, Klemenetina et al. in [21] identify four major building blocks for a DT. These blocks include the Physical Entity platform, Data Management platform, Service Platform and Virtual Entity platform. Further, Tekinerdogan and Verdouw in [22] propose a pattern-based digital twin architecture, where each pattern consists of structure to define the components, and the dynamic to define the interactions. Using semantics technology, Li et al. presented an approach for monitoring robot interactions using a semantic-based digital twin [23]. Mattila et al. in [24] presented an approach for building a digital twin using the ROS, Gazebo and Twinbase frameworks. Moreover, as the digital twin topic has attracted many researchers in recent years, several attempts were seen to organize the development of DTs. As an example, Hendrik et al. proposed a taxonomy for building digital twins in [25]. This taxonomy helps the developers to pinpoint their efforts based on the requirements of the digital twin. Another published research suggests a six-layered architecture for developing digital twins [26]. Additionally, Resman et al. presented a five-step approach for developing a data-driven digital twin [27].

These different approaches and techniques are driven by the usage of the digital twin itself. As an example, the term Cognitive Digital Twins (CDT) involves the digital twins with human-like capabilities such as perception and reasoning. In fact, the term "cognitive twin" is relatively new with simultaneous works reported in [28,29]. Abburu et al. [30] presents a conceptual definition and initial implementation in the COGNITWIN software toolbox aiming at cognitive capabilities for optimal operations and maintenance of process equipment and assets. Likewise, Lu et al. [31] presented an approach enriched with augmented semantic capabilities for identifying the dynamics of virtual model evolution, promoting the understanding of interrelationships between virtual models and enhancing the decision-making based on DT in the frame of H2020 project FACTLOG. Recently, Rozanec et al. [32,33] drew attention to the strategy for knowledge graph modeling to construct actionable cognitive twins focused on capturing specific knowledge related to production planning and demand forecasting in a manufacturing plant, whereas Li et al. [34] put the attention on a unified ontology modeling approach based on GOPPRR (graph, object, point, property, role, relationship) for co-simulation.

Looking at the number of published papers regarding digital twins, it is evident that DTs are an important technology for both research and commercial use. With the majority of the presented research work tending to address different industries and systems, they still agree on the basics of a digital twin. Mainly, a digital twin is a computer application where it will require data management and proper interfaces. The main difference appears in the usage and employment of the digital twin. As an example, if a digital twin is used to optimize operations, then it will require an optimization solver, and if a digital twin is designed to simulate the behavior of the physical system, then it will require a modeling feature. However, the majority of the reviewed articles present an ad hoc solution for a specific problem which provides important insights into the specific application. This limitation is key for the objectives of this research.

2.2. Commercial Solutions for Building Digital Twins

The commercial solutions are critically important in such a study as these solutions are featured as realistic, build-to-fit the customer needs, easy to use, and reliable products [6]. In addition, the commercial solutions hint the researchers towards the direction and the trends of the future in terms of research and innovation activities. Therefore, this section highlights some features of commercially available products. Some of these products are marketed as digital twins. While some are marketed as IoT platforms, simulators, suits, studios or even programming languages [35]. Regardless of the naming, all the studied tools can be listed under the generic definition of a digital twin as long as the main concept of having a representation of a physical system that consumes data and generates information stands. In this matter, Table 1 presents some well-known use cases of using the Digital Twin concept in a commercial manner.

Table 1. Commercially available solutions for building DTs.

Vendor	**Brief Description**
General Electric [36]	General Electric (GE) targets the concept of DT based on the application area. According to [37], Dr. Colin Parris presents the interest of GE in three main areas: assets, network and process. For each area, GE provides a set of applications in order to form and build the digital twin. As an example, for the Assets Digital Twin, GE employs the Assets Performance Management and for the Network digital twin, the applications named ADMS and GIS are employed.
IBM Digital Twin Exchange [38]	IBM provides an open marketplace for asset owners and end users to exchange assets and build digital twins based on shared data. In this regard, the platform is open for anyone to introduce models and data to purchase or use.
PTC Digital Twin [39]	Like IBM, PTC provides the customer with a marketplace that contains more than 130 tools. These tools can be utilized for building DTs. In addition, PTC provides guidance and development kits for building new tools that specifically suit their customers. Regarding the application domain, PTC considers DT to be beneficial in five key sectors. These sectors include: Corporate/CXO, Product Engineering, Sales and Marketing, Manufacturing Operations, and Customer and Technician Services [40].
Microsoft Azure Digital Twin [41]	Azure is a cloud-based platform from Microsoft. It is marketed as an IoT cloud platform for data acquisitions, modeling and estimation. According to the [42,43], the Azure DT allows the customer to build a virtual model based on the IoT data. Afterwards, these models are connected with each other to form the DT.
Ansys Twin Builder [44]	Ansys is a corporation specializing in developing simulation tools for various industries and business sectors. According to [45,46], Ansys Twin Builder exploits predefined modules for building DTs. In addition, it allows integration with third-party applications such as Azure for data collection and modeling purposes.
SAP SE [47]	The solution provided by SAP mainly addresses the resources, products and assets. This solution is marketed as a network of digital twins as described in the white paper [48]. It can provide simulation and real-time estimation of products during the lifecycle. In addition, it provides a modifiable interface for flexible interaction with the end user.
Oracle [49]	Oracle's DT is based on an IoT platform that permits data and information interconnectivity. According to Oracle, the implementation of a DT includes three main pillars. These pillars include Virtual Twins where devices are emulated, Predictive Twins where the data is analyzed, and Twin Projections where whole systems are simulated based on the analysis that is created by the Predictive Twins.
Bosch GMBH [50]	Bosch is specialized in developing a building Digital Twin that consumes IoT data from sensors that are scattered in buildings using Bosch devices. According to Bosch, the Digital Twin is built using the Microsoft Azure IoT platform, and it employs the semantic technology for knowledge reasoning.
Emerson [51]	The Digital Twin provided by Emerson holds features such as an automation system, vendor independence, selective fidelity, open architecture, and cloud ready [52]. In fact, the DT is utilized in safety, training, Knowledge transfer, environmental, regularity, and optimization applications.
ABB [53]	ABB provides several solutions that form digital twins based on the end user needs. In this regard, ABB supports the customers with DTs for the design, system integration, diagnosis, and prediction activities and applications. For example, the virtual commissioning DT for discrete manufacturing, virtual drive tuning DT, and the predictive maintenance DT for vessels.

Table 1. *Cont.*

Vendor	Brief Description
MATLAB/ Simulink [54]	Even though they are known to be a programming language and/or a tool for mathematicians and engineering, MATLAB and Simulink are capable of providing virtual representation on a physical system for the purpose of testing. This is evident in [55] as there are two main methods reported for building a digital twin. The first method employs a data-driven approach exploiting the deep learning tools in MATLAB. The second method involves a Simulink block network. The latter is known as a physics-based approach.
COMSOL Multiphysics ® [56]	Like MATLAB, COMSOL Multiphysics is a multi-disciplinary simulation platform. This platform utilizes mathematical models of the physical systems for the simulation process. The models are flexibly added as an add-on where the user decides what is needed for the application and domain. In addition, the platform provides a very useful interface including visualization of the components based on the simulation parameters.
NVIDIA Omniverse™ Enterprise [57]	Omniverse is a state-of-the-art platform that allows developers and designers to build and simulate systems with a high level of realism [58]. The platform is based on Pixar's Universal Scene Description and it uses the NVIDIA RTX™ technology. The Omniverse platform includes five key elements: Nucleus is the database and collaboration engine, Connect is the data connections plugins engine, Kit is the Software Development Kit (SDK), Simulation is a set of realistic models that allows the user to select or create, and finally, RTX Renderer creates the high realistic simulations for users and developers.
Visual Components [59]	Visual Components offers 3D manufacturing simulation tools for the manufacturing domain. These tools offer several functionalities such as factory layout configuration, process modeling, statistical analysis, shopfloor connectivity, and robotics simulation among other features. In fact, Visual Component also provides compatible adapters that can work with the omniverse of NVIDIA.
Tecnomatix® by Siemens [60]	Tecnomatix is an industry-driven digital twin for manufacturing applications by Siemens. The solution provides features such as virtual commissioning, human-centered design and planning, plant simulation, robotics programming, statistical analysis, planning and processes optimization, assembly simulation, and shopfloor and layout configuration. This digital twin can be customized and scaled thanks to its development environment.

The search for commercial DTs can be considered a challenging task as information is kept from the public. In addition, several vendors and companies market their existing solutions as digital twins which requires extra investigation. However, there are four main groups or categories that appeared to form in terms of commercially available solutions regarding the methods for building and using a digital twin. The first group is the cloud-based platforms, which are also titled IoT platforms. For instance, this group includes Amazon, Google, Microsoft and Oracle cloud systems. This category is based mostly on the data-driven approach where the user uploads data or connects IoT devices that publish the data to the platform. Then, a set of predefined or accustomed functions and operations are introduced in order to model, analyze, and process the data. In addition, this category provides easy-to-use AI solutions for customers. Finally, the end user maps the results with web services for utilization purposes. Overall, this category is highly flexible and generic for end users to utilize in several domains. Hence, several solutions are built using these categories such as ICONICS's building digital twin [61] and DOOSAN's sustainable energy production digital twin [62].

The second group includes vendor-specific digital twins. This category consists of the companies that provide industrial systems, devices, controllers, and hardware that requires tools to simulate the operations. As an example, ABB's RobotStudio, Siemens's TIA Portal, Omron's ACE Software, and the majority of robotics and PLCs vendors. This category is very accurate and detailed in terms of simulating the systems. In addition, it provides the needed visualization for the end user to allow the best interaction. Nonetheless, the DTs in this group may suffer from being very limited only to the vendor's own equipment. In other words, a digital twin in this group will hardly include a model of equipment from another vendor. As an example, the libraries of the TIA portal by Siemens only include Siemens PLC devices or the libraries of RobotStudio by ABB include only ABB robots and controllers. To overcome this limitation, the users may use standards and protocols for communicating between digital twins from different vendors. This practice might be complicated and will increase the time for preparing a functional digital twin of the production system. Thus, it is advisable to reduce multi-vendor setup in the factories.

The third category is the application-specific Digital Twins. This category is focused on a certain application or domain. For instance, SolidWorks, SAP, ICONICS, Visual Components and Tecnomatix by Siemens. The DTs in this category are dedicated to solving definite problems in a specific domain. As an example, the Visual Components Digital Twin concept includes several simulation applications that can solve different problems in manufacturing such as layout configuration, process modeling and virtual commissioning [63]. Looking at Siemens's Tecnomatix, the digital twin includes very sophisticated models that allows statistical analysis, optimization of processes and virtual commissioning. Adding to that, this digital twin includes a development environment for end users which increases the flexibility and interoperability. In fact, and as this category of digital twins is built to fit specific applications, industrial digital twins can be found mostly in this category. Finally, the fourth group is the generic graphics and simulation group. This group includes solutions such as Blender [64] and Autodesk's Maya [65]. These solutions are mainly used in film making and game making to develop highly realistic scenes and visual effects using its built-in physics engines and modeling tools. This group is not usually exploited in industrial applications as it lacks connectivity to the real world. Nonetheless, it can create a decent simulation of systems, especially in mechanical and civil engineering. Furthermore, and as an advancement, NVIDIA took these concepts to the industrial sector with its digital twin platform Omniverse. As reported in [66], BMW and NVIDIA collaborated to build a highly realistic and accurate digital twin of an automotive factory with the inclusion of human workers. This puts NVIDIA's Omniverse platform as one of the most sophisticated solutions that can be used to develop digital twins.

Overall, there are several business models that are implemented in marketing and operating these digital twins. Most importantly, and it was very noticeable in the commercially available solution, these digital twins use and follow industry standards which allows interoperability and exchangeability of tools and applications. This feature gives the developers room to develop their own applications and thus, exploit these solutions to its maximum.

2.3. Review on Digital Twins Standards

There is no doubt that the concept of digital twins is highly linked with various technologies and standards as illustrated in [67]. According to the authors, the digital twin combines standards from topics such as Physics Entities, Virtual Entities, Data, Connection and Services. Each topic contributes to the concept of digital twins as the use case needs. For instance, if the digital twin connects to real-time data, then it will require following data collection standards that allow such a service. Furthermore, the industry domain may also contribute to or influence the digital twin solution. As an example, developing a digital twin for the marine industry will require the solution to adhere to some standards and concepts of marine life in general. Similarly, the case of manufacturing, mining, military, etc.

This interrelation with other technologies and domains triggered the issue of ISO 23247 Automation systems and integration—Digital twin framework for manufacturing standard [68]. This standard consist of four parts. The first part presents an overview and terminology related to digital twins. The second part describes the reference architecture. The third part presents the relation with the manufacturing elements. The fourth part presents the information exchange concepts and methodologies. Figure 1 depicts the reference model of the digital twin in the ISO 23247. As presented in the figure, the reference model consists of four main entities. The first entity is the user entity which includes all the needed components for interfacing with the digital twin and manufacturing systems. The second entity is the digital twin entity which includes all applications, services, operations, and management components. This is the core entity that builds the models and the simulation acts.

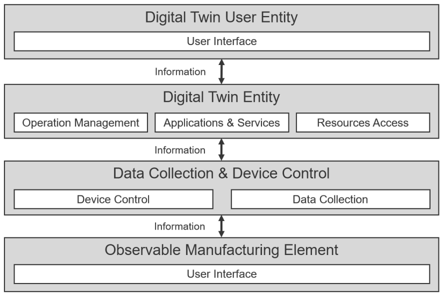


Figure 1. Digital twin framework based on the description in the ISO 23247 standard. [Source: Own figure based on the International Organization for Standardization (ISO) illustration in [68]].

Then, the third entity is the data collection and control entity. This entity is the module where manufacturing elements are connected to collect data from shopfloor devices. Finally, the fourth entity is the observable manufacturing elements entity which includes the physical components. For better understanding, the National Institute of Standards and Technology published use case scenarios of implementing digital twins based on the ISO 23247 [69]. In addition to that, the ISO published a dedicated standard for the visualization elements in automation system under industrial data topic is titled as ISO/TR 24464:2000 Automation systems and integration—Industrial data—Visualization elements of digital twins [70]. This standard focuses only on the visualization elements to be shared between the physical and virtual replicas. ISO is also developing two standards related to digital twin topics. The first one, the ISO/IEC AWI 30172 Digital Twin-Use cases, includes description based on the use case [71] and the second one, the ISO/IEC AWI 30173 Digital twin—Concepts and terminology, on generic concepts and terminology [72].

3. Generic Paradigm for Architecting Digital Twins

Following the review in the previous section, the development of a digital twin depends on several factors including the domain where the DT will be utilized, the understanding of the physical system behavior, the availability of the data, the required level of details in simulation, and the required human interfaces and manner of interactions. Thus, this section presents a generic paradigm of architecting digital twins for manufacturing processes. The main objective of this generic paradigm is to support developers and engineers with adequate guidance for architecting the digital twin regardless of the application or the technologies. In addition, this generic paradigm must be applicable and compatible with different platforms, programming methods and techniques, operating systems, and hardware devices. Consequently, developers will be able to reuse some concepts of the generic paradigm fully or partially based on the application's need.

Generally, the paradigm realizes the digital twin from three points of view. Firstly, the digital twin is a computer application that is deployable on a centralized, decentralized or distributed computing system that is able to interact with other entities and applications using services and proper human interfaces. Secondly, the digital twin is a replica of the physical part of a Cyber-physical System (CPS). Thirdly, the digital twin is a system with one or more perspectives that allow simulating a physical system using virtual replicas. Following these three points of view, the generic paradigm for developing digital twins is a combination of three concepts: *multi-layer* , *multi-level*, and *multi-perspective*.

3.1. Multi-Layer Concept

Developing computer applications may implement well-defined approaches such as Model-View-Control (MVN), Object-Oriented Architecture (OOA), or Multi-Agent System (MAS) among others. These approaches help the developers in architecting the applications based on the problem needs. As an example, the MVC approach helps the developers to separate the application blocks or modules based on the functionality. Meanwhile, OOA is mainly used in hierarchical architecture where instances are dynamically created and linked with a parent instance. The MAS approach is mainly used in distributed system architecture where agents are permitted to communicate with each other as needed. Another approach for building applications is the onion architecture [73,74]. The onion architecture was developed by Jeffery Palermo [75] which is an implementation of the concept of the clean architecture by Robert C. Martin [76,77]. The onion architecture is a systematic clean method for architecting computer applications by using onion-like layers where the core or the model of the application is in the center and the application interfaces, such as Graphical User Interfaces (GUIs) and Application Programming Interfaces (APIs), at the outside layer of the onion. Some layers can be planned in between the inner and the outer layers if needed. Inspired by the onion architecture, the *multi-layer* concept envisions the digital twin as a computer application with three major layers: *model*, *signals*, and *interfaces*. Figure 2 depicts the multi-layer concept. The *Model* layer represents the core part of the application which includes the logic and behavior of the application. The *Signals* layer includes the data and information management and transformation. Finally, the *Interfaces* layer includes the HMIs and APIs. These interfaces include the human–machine interface and the machine–machine interface. The *signals* layer acts as a linking layer between the interfaces and the core of the digital twin.

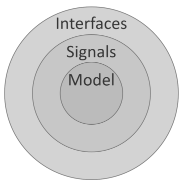


Figure 2. Multi-layer concept following the onion architecture. [Source: Own figure].

3.2. Multi-Level Concept

Following the second consideration of the generic paradigm, a DT must mimic the physical part of a CPS. In industrial applications, Industrial Cyber-physical Systems (ICPS) can be projected on the ISA95 standard (ANSI/ISA–95.00.01–2000) [78] as depicted in Figure 3. As highlighted in the figure, the cyber part is colored in blue and includes the Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MES) levels. Meanwhile, the physical part is presented in red and includes the production process at level 0 and manipulation and sensing at level 1 of the factory shopfloor. The transition between the cyber and physical parts or worlds occurs at the control and supervisory level (Level 2) of the factory shopfloor. As shown in the figure, the data and information flow is defined following the purpose and features of each level. In this regard, the ERP provides the schedules to the MES level based on the incoming orders and the available resources. At this level, the planning, scheduling and optimization of the production processes are conducted. Then, the MES transfers these schedules into optimized plans and then issues commands to the control and supervisory level to accomplish these plans. The control and supervisory level makes actions accordingly at the manipulation level. These actions affect the production process. As a consequence of that, the sensors in Level 1 detect the changes and send the reactions to the control and supervisory level. The control level provides these reactions as feedback to the MES level. Finally, the MES reports to the ERP the production status.

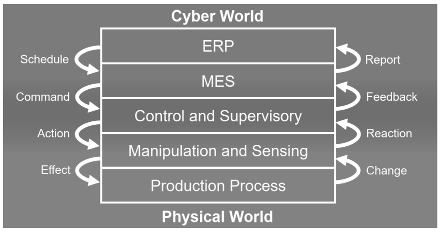


Figure 3. Projection of the CPS concept on the ISA95 levels. [Source: Own figure].

In the context of digital twins for industrial applications, the factory shopfloor levels (Level 0–*Level 2*) represents the physical part that can be twined. As depicted in Figure 4, and being inspired by the ISA95 standard, the *Multi-level* concept of this generic paradigm includes three levels. *Level 2* is the logic level where all the logic of the production system is included. This level imitates the ISA95-*Level 2* where the control and the supervisory happens. *Level 0* is the level where all the physics of the production system is included. Similar to the ISA95-*Level 0*, this level aims at modeling the production process. Moreover, in addressing multi digital twin situation, *Level 0* is the location where the digital twins can be linked in the physical space. Finally, *Level 1* is the level where the transformation between the logic and the physics occurs. This level represents the manipulation and sensing technologies. Moreover, and as presented in the figure, during production, the MES and the ERP are normally connected to the shopfloor resources. For the digital twin, two main operational modes are presented. These operational modes are activated once the digital twin is developed and deployed. The *On-line* mode when the DT is connected in real-time to the shopfloor resources. In this mode, the digital twin acts as a monitoring system for the physical twin where live streams of data (shown in gray arrows) feed the DT. Meanwhile, the *Off-line* mode is used in simulating the shopfloor resources. In this mode, the DT will be connected to the existing ERP and MES systemsfor testing and validation of plans and operations.

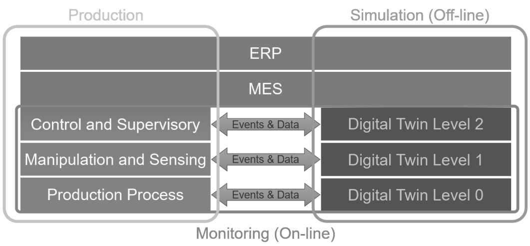


Figure 4. Multi-level concept and the DT operational modes. [Source: Own figure].

By combining the concept of *multi-layer* and the *multi-level*, a digital twin will contain 9 blocks as depicted in Figure 5. For a generic design of a digital twin, the 9-block architecture allows modeling and simulating any production system. As depicted in the figure, the *Model* layer includes all the modeling, logic, behavior and physics simulation components. At level 2, *M2* includes the logic of the control and supervisory. The *M1* block will contain the models of the sensors and the actuator systems. The *M0* will include the models and physics of the process. Then, the *Signals* layer includes all the operations and transformations on the data that will be shared between the *Model* and the *Interfaces* layers. This layer allows the developers of the digital twin to include any needed operations of the data and the information. In this regard, the data and the information can be either generated by the *Model* layer and need to be prepared for the *Interfaces*, or visa-versa, received by the interfaces and need to be delivered to the model layer.

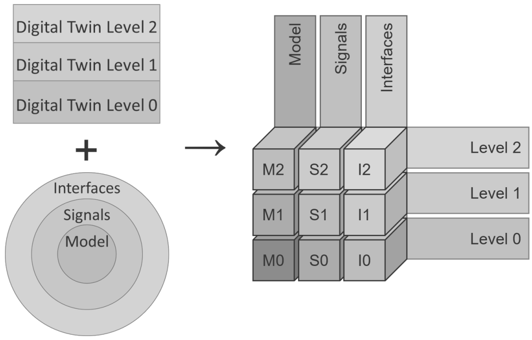


Figure 5. The 9-block concept. [Source: Own figure].

Following that, the *S2* block will contain the transformation components of the control and supervisory. Meanwhile, the *S1* will contain the information transformation components for the actuation and sensing level, and *S0* will include the information transformation components for the process level. Finally, the Interfaces layer is where the communication occurs with human using HMIs and with machines using APIs. The *I2* is dedicated to the interfaces regarding the control level. Similarly, the *I1* is designed for interfaces of the actuation and sensing level, and *I0* for the interfaces of the process level. It is important to mention that the interfaces allow access to inner data and information that is not possible to access in the real system. As an example, in a production system, a motor can be controlled by a drive. This drive includes power electronics that create the sinusoidal signals for the motors. These signals often include the undesired harmonics. Thus, the drives use filters to clean the signals from the unwanted harmonics. In real systems, it is very difficult to study the effect of these harmonics on the other systems in the production systems due to a lack of access. However, in digital twins, this can be replicated and the interfaces layer can allow access to any inner values that the user of the digital twin is interested in.

3.3. Multi-Perspective Concept

Returning to the third point of view for a generic paradigm, a digital twin must provide multiple perspectives to fulfill the different possible applications. In this regard, the *Multi-perspective* concept considers a complete digital twin as concentric cylinders as depicted in Figure 6a. As described before, the levels (in orange) are represented by vertically stacked cylinders and the layers (in blue) are concentric cylinders. To understand the multi-perspective concept, a look from the top of the digital twin paradigm shows different sectors as presented in Figure 6b. Each sector (in green) focuses on a different perspective of the digital twin. As an example, a digital twin can be used for cost estimation and simulation. At the same time, it can be also used for simulating energy consumption or resource utilization. Following the multi-perspective concept, as shown in Figure 6, three perspectives will be needed to address the cost, energy consumption and the resources' utilization use cases. In addition, such a concept allows scaling or extending digital twin usage to address modern or newly developed technologies. As an example, the above-described digital twin can have a new perspective regarding the cognition aspect. This addition will allow the digital twin to be a cognitive digital twin.

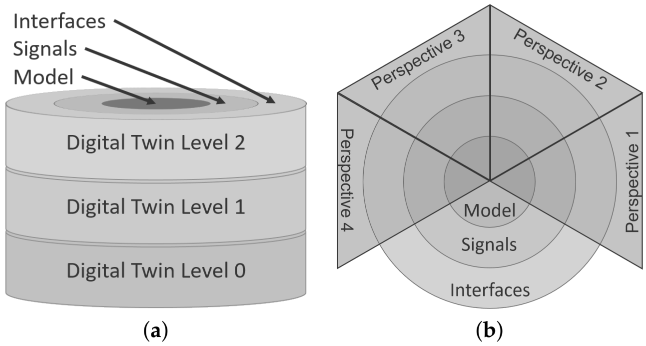


Figure 6. Multi-perspective concept. [Source: Own figure]. (**a**) Side view ; (**b**) Top view.

In summary, a generic digital twin will include several perspectives depending on the application and the purpose of the digital twin. Mainly, each perspective includes the 9-block architecture. Figure 7 depicts 3D visualization of the paradigm with different color for each perspective. The main objective of such a paradigm is guaranteeing *modular*, *scalable*, *reusable*, *interoperable* and *composable* digital twin architecture. These qualitative attributes can be mapped to the architecture as follows:

- *Modularity*: is the ability to build the DT using defined and interchangeable modules. These modules contribute to the flexibility of the overall system. In the context of the generic paradigm, the 9-block approach allows the developer to interchange the blocks based on the need of the application.
- *Scalability*: is the ability to grow the system in terms of resources and features. This quality is presented in the possibility of adding several perspectives to the digital twin in order to increase its capabilities.
- *Reusability*: is the ability to reuse legacy or existing assets in building newer versions. In this regard, the 9-block approach allows the user to reuse previously-developed blocks in newer versions of the DT.
- *Interoperability*: is the ability to work and to be compatible with other components or systems regardless of the vendor or the developer. For the 9-block approach, the use of standards and protocols permits such quality where different systems and applications can communicate easily.
- *Composability*: it is the ability to reassemble and reconstruct a system from other systems and components. In the context of this research, the 9-block architecture allows the development of a system of digital twins where the DT Level 0 includes the physics that each DT uses. Then, by connecting each Level 0 using spatial computing concepts, these DTs can form a system of digital twins.

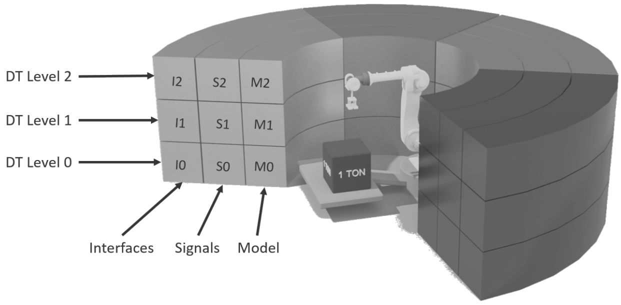


Figure 7. 3D visualization of the generic paradigm for architecting digital twins. [Source: Own figure].

4. The Approach of Architecting the Digital Twin

After establishing the skeleton of the DT using the 9-block concept, developers can start the process of architecting the DT. This process may include the selection of some aspects such as the communication protocol, the shared data, the components and applications, and the development techniques. In the domain of this research, and after defining the 9-block paradigm, architecture mainly refers to two parts. Firstly, the components that construct the digital twin modules, and secondly, the data types and protocols on how these components communicate with each other and with the external applications and services via defined APIs. In this regard, a knowledge-based system can be used to reason the architecture of the DT. As depicted in Figure 8, the reasoning process may require three knowledge bases. Firstly, the *Process description* (green) supports the reasoner with the production resources knowledge. As an example, this ontology may contain descriptions of the production resources such as robots and machinery, product, data and possible events. Secondly, the *Twinning specifications* (red) which include the rules for the reasoning engine, the needed perspectives for the digital twin, and the possible interfaces.

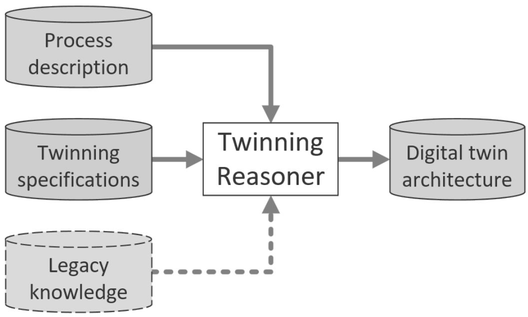


Figure 8. The knowledge-base system for reasoning the digital twin architecture. [Source: Own figure].

Thirdly, the *Legacy knowledge* (yellow) which includes manufacturing vocabulary and descriptions that are required to identify production resources such as robots or conveyor descriptions. In fact, many ontology models are developed and populated for this purpose such as the IOF-BFO developed by the Industrial Ontology Foundry [79–81], the Politecnico di Milano-Production Systems Ontology (P-PSO) and the Manufacturing Systems Ontology (MSO) [82], EU Vocabulary Ontology [83], and the Semantic Computing Research Group (SeCo) ontology [84]. Once these ontology models are utilized, the reasoning engine or as labeled in the figure *Twinning reasoner* will infer the architecture description as an ontology (blue). Figure 9 presets a graph that describes the ontology model of such a knowledge-based approach. This figure shows the relations between the factory and its production resources in the physical world and the digital twins and their components in the digital world. This graph follows the same colors scheme in Figure 8.

More precisely, and starting from the top, the *Production_Control* class represents the MES and the ERP systems that control the production process. As an example, this class can also include the concept of AI-based *Digital Agents*. This *Production_Control* class is linked to the *Production_Process* class via *manages* object property, and to the *Product* class via *monitors* object property. As the figure shows, a *Digital_Twin* simulates the production process or the product. In addition, the *Digital_Twin* virtually replicates the *Factory* and it can have multiple *Perspective* instances linked to it based on the DT specifications. The *Factory* is linked to the *Production_Process* via the property *conducts*. Further, the *Factory* class is linked to the *Product* via the property *produces*. As seen in the figure, the *Product* class and the related object properties are highlighted as this research focuses on the production processes of digital twins. The approach also considers that a factory comprises multiple instances of the *Production_Resource* as shown in Figure 9. The *Production_Resource* can be a robot, machinery or human worker that provides service or conducts an act in the production process. Additionally, the *Production_Resource* is linked to the *Level* class where the resource is located in terms of manufacturing hierarchy which is needed for the 9-block paradigm. Likewise, a digital twin comprises multiple *Digital_Twin_Resources*. This *Digital_Twin_Resource* class is linked to the *Production_Resource* via *virtually_replicates* object property. In addition, the *Digital_Twin_Resource* may contain several *Digital_Twin_Component*. Each instance of the class *Digital_Twin_Component* is linked to the *Model*, *Signal* and *Interface* classes via *located_at* object property. On the other side, a *Production_Resource* may consume or generates *Data*. Additionally, it can trigger or expect an *Event*. Finally, this approach uses the IEC61499 standard for building digital twin components. In this regard, the *Function_Block* class is linked with *Data* and *Event* via *is_input_to* and *is_output_of* objects properties and is represented as function blocks. As mentioned before, the *Legacy knowledge* (colored in yellow) defines manufacturing vocabulary. This ontology is added to the graph partially for illustration purposes.

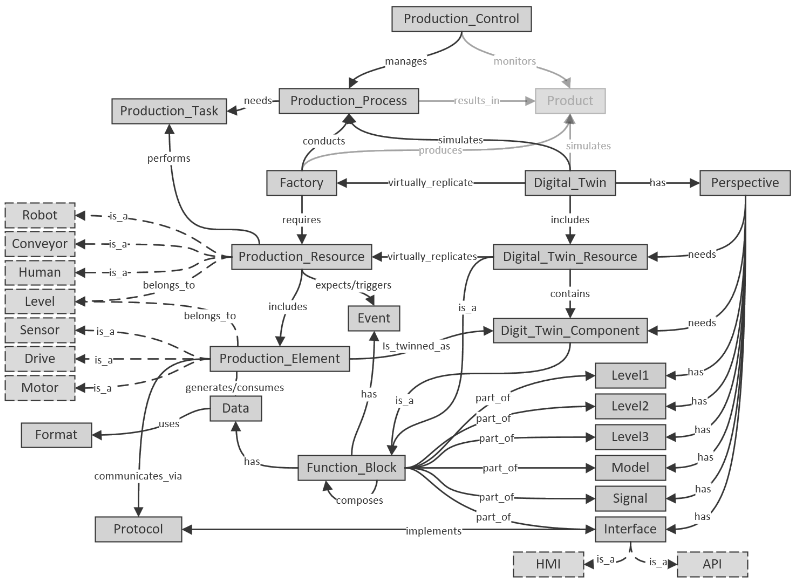


Figure 9. Digital Twin Architecture Ontology Model (DTAOM). [Source: Own figure].

5. Use Case Example

5.1. FASTory Use Case

To demonstrate the proposed approach, a production line with a discrete manufacturing use case is used. The production line is known as the FASTory assembly line where modular and sequential loop-like workstation configuration is used to mimic the production of electronic devices as depicted in Figure 10. The production system includes 10 identical workstations (labeled as WS2–WS6 and WS8–WS12) that are equipped with conveyors and different robots to perform the assembly processes. The assembly process mimics assembling mobile phones where robots draw different parts of the phone with different colors and different models to emulate the variation of the product. Workstation 1 and workstation 7 are used for pallets and material handling.

The research in [1] illustrates the physical system and the development of a simulator to help in developing applications for the production system. Furthermore, this use case was exploited in several research activities such as developing a generic MES system in [85], proposing a knowledge-based manufacturing system in [86], studying the industrial data lifecycle management in [87], and implementing the ISO-22400 KPIs in a virtual production systems in [88] among others. To reduce the complexity of demonstrating the approach in this paper, a single workstation will be used where the assembly process happens. As shown in the Figure 10, any drawing workstation, i.e., WS2, consists of two conveyors. The main conveyor is for passing the pallet to the work cell in order to draw the phone parts. Meanwhile, the by-pass conveyor allows the pallets to move past the workstation in case the robot is in a busy state. This configuration guarantees zero blockages due to pallet congestion. Five proximity sensors are attached to the conveyors to measure the presents of the pallet in the workstation. In addition, an RFID reader is located at the beginning of each workstation to identify each arriving pallet.

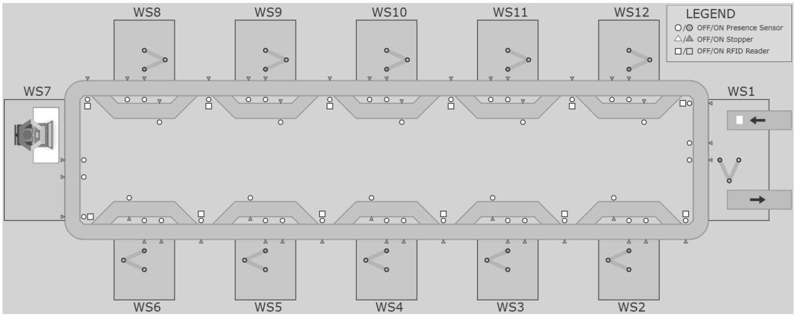


Figure 10. FASTory assembly line layout. [Source: Snapshot from the FASTory Simulator [1]].

Figure 11 depicts a graph description of the workstation process and resources. Each workstation is controlled by two Remote Terminal Units (RTUs) [89]. These RTUs provide connectivity to the MES and ERP systems using web services. Then, the Robot RTU controls the robot controller, which manipulates the robotic arm. The gripper is attached to the robotic arm, and it grasps the pen that has certain color to draw on the paper. This paper represents the product and is attached to a pallet. The pallet is transferred by the main conveyor if the robot is in an idle, stop or down state, and by the by-pass conveyor if the robot is in the busy state. These conveyors are moved by separate motors that are driven by separate drives. In addition, proximity sensors and an RFID are located at certain locations along the conveyor belts. Finally, these drives, sensors and the RFID are controlled and read by the Conveyor RTU.

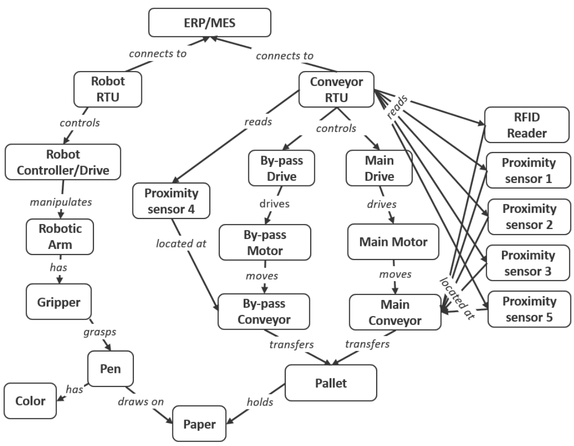


Figure 11. FASTory production resources description. [Source: Own figure].

5.2. Function Block-Based Digital Twin Architecture

For the FASTory use case, and following the graph in Figure 11, three main production resources can be listed including the robot, the main conveyor and the by-pass conveyor. Additionally, the proximity sensors, the RFID reader and the gripper can be added as production elements. Figure 12 depicts the populated model of the process description for the FASTory use case. The figure shows that the model includes four processes where three processes are performed by the robot and one process is performed by the transportation system.

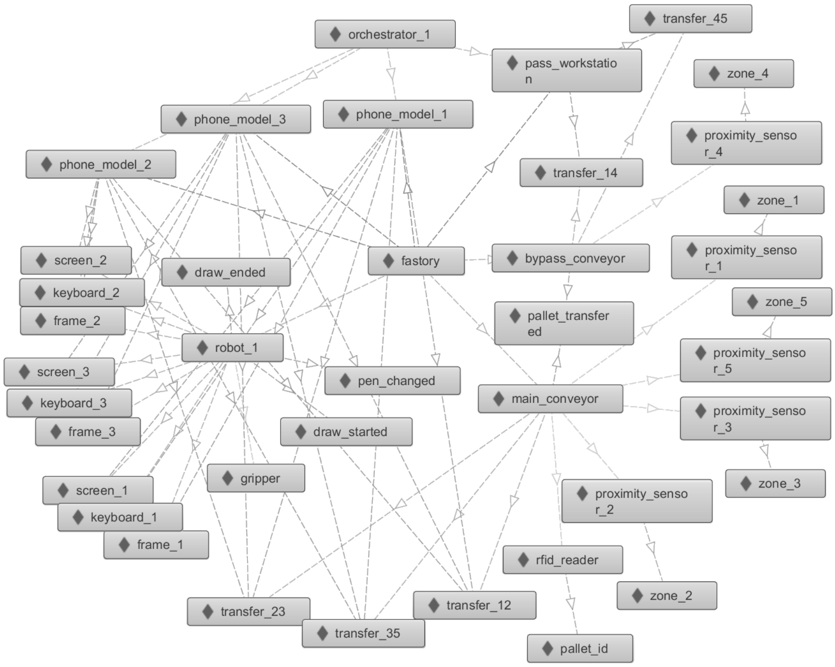


Figure 12. FASTory production description for drawing process. [Source: Own figure which is a snapshot from Protege Ontology Editor].

For the Data, six instances have been included in the use case as each zone includes the status of the proximity sensor that is related to. The *pallet_id* data instance is related to the RFID reader and it holds the pallet identifier code. For the events, four events have been added. These events include *draw_ended*, *draw_started*, *pen_changed*, which are triggered by the robot and *pallet_transferred* which is triggered by the conveyors.

Afterwards, the digital twin specifications can be introduced. For this use case, two perspectives can be created. The first one is the power consumption perspective. The second perspective can be resource utilization. Figure 13 depicts an example of a populated ontology for the digital twin specifications. As seen in the figure, each instance of the perspective class owns three instances of the model, three instances of the signals and three instances of the interface. These instances will be linked later with the digital twin components after the twinning operation. The third knowledge source is legacy knowledge. This knowledge includes all descriptions and definitions of the production resources. As an example for this use case, Figure 14 depicts a snippet of the ontology that describes the legacy knowledge of FASTory workstations as presented in Figure 11.

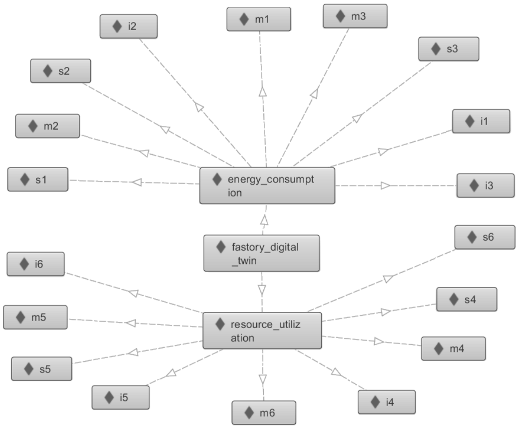


Figure 13. Twinning specifications for FASTory use case. [Source: Own figure which is a snapshot from Protege Ontology Editor].

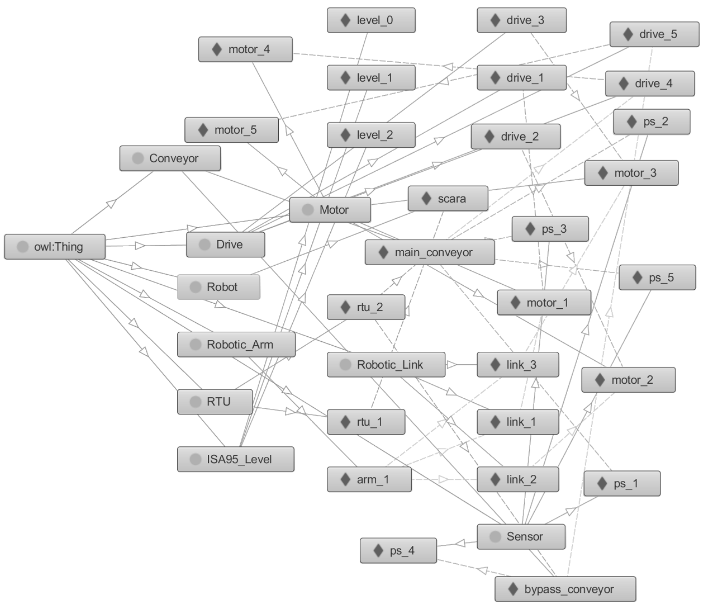


Figure 14. Snippet of the legacy ontology for the FASTory use case. [Source: Own figure which is a snapshot from Protege Ontology Editor].

Once these ontology models (seen in Figures 12–14) are populated, a set of rules are needed for the reasoning. These SWRL (Semantic Web Rule Language) rules are part of the *Digital Twin Architecture* ontology model (in blue in Figure 8). In this regard, Equation (1) is a rule that creates a *Digital_Twin_Resource* instance for each *Production_Resource*. Then, a datatype *type* is added to the *Digital_Twin_Resource* instances which highlight the virtual feature using the rule in Equation (2). Equation (3) can be used for linking the virtual and physical resources.

$$Production\_Resource(?x) \implies Digital\_Twin\_Resource(?x) \tag{1}$$

$$Digital\_Twin\_Resource(?x) \implies type(?x, "virtual - resource") \tag{2}$$

$$\begin{aligned} Production\_Resource(?x) \wedge \\ Digital\_Twin\_Resource(?x) &\implies virtually\_replicated(?x, ?x) \end{aligned} \tag{3}$$

Afterwards, an instance of *Function_Block* can be created for each instance of the *Digital_Twin_Resource* using the rule in Equation (4) . Additionally, the same rule links the function block instance with the digital twin resource instance using *is* object property. Equation (5) depicts the rule for adding an alias to the function block as a data property. Then rules in Equations (6) and (7) add the events to the function blocks.

$$\begin{aligned} Digital\_Twin\_Resource(?x) \wedge \\ swrl: makeOWLThing(?y, ?x) &\implies Function\_Block(?y) \wedge is(?x, ?y) \end{aligned} \tag{4}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Resource(?y) \wedge \\ name(?y, ?z) &\implies alias(?x, ?z) \end{aligned} \tag{5}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Resource(?y) \wedge \\ expects(?y, ?z) &\implies has(?x, ?z) \end{aligned} \tag{6}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Resource(?y) \wedge \\ triggers(?y, ?z) &\implies has(?x, ?z) \end{aligned} \tag{7}$$

Once the digital twin resources are created, the digital twin components can be created and linked. This can be seen in the rules in Equations (8)–(10).

$$Production\_Element(?x) \implies Digital\_Twin\_Component(?x) \tag{8}$$

$$Digital\_Twin\_Component(?x) \implies type(?x, "virtual - element") \tag{9}$$

$$\begin{aligned} Production\_Element(?x) \wedge \\ Digital\_Twin\_Component(?x) &\implies is\_twinned\_as(?x, ?x) \end{aligned} \tag{10}$$

Afterwards, the digital twin components are represented as function blocks. Equation (11) creates the function block, while Equation (12) adds the alias to the created instance of the function blocks. Then, Equations (13) and (14) link the data input and outputs of the function block

$$\begin{aligned} Digital\_Twin\_Component(?x) \wedge \\ swrl: makeOWLThing(?y, ?x) &\implies Function\_Block(?y) \wedge is(?x, ?y) \end{aligned} \tag{11}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Element(?y) \wedge \\ name(?y, ?z) &\implies alias(?x, ?z) \end{aligned} \tag{12}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Element(?y) \wedge \\ generates(?y, ?z) &\implies has(?x, ?z) \end{aligned} \tag{13}$$

$$\begin{aligned} Function\_Block(?x) \wedge \\ Production\_Element(?y) \wedge \\ consumes(?y, ?z) &\implies has(?x, ?z) \end{aligned} \tag{14}$$

The created instances of the function block represent the digital twin components and the digital twin resources. To compose the function blocks, the rule in Equation (15) is used to combine the digital twin components and resources using the function blocks.

$$\begin{aligned} Digital\_Twin\_Resource(?x1) \wedge \\ Digital\_Twin\_component(?x2) \wedge \\ contains(?x1, ?x2) \wedge \\ is\_a(?x1, ?x3) \wedge \\ is\_a(?x2, ?x4) \implies composes(?x4, ?x3) \end{aligned} \tag{15}$$

Lastly, the function components will be sorted to classify them using the 9-block concept. Initially, as presented in Figure 13, two perspectives are created for this digital twin. Each perspective has three layers and three levels. For the levels, the function blocks that belong to each level follow the associations of the production resources and production elements. This knowledge comes from the description of these instances which are represented in the legacy knowledge. Equation (16) depicts the rule for making this linkage.

$$\begin{aligned} Digital\_Twin\_Resource(?x1) \wedge \\ Production\_Resource(?x1) \wedge \\ Digital\_Twin\_component(?x2) \wedge \\ Production\_Element(?x2) \wedge \\ Function\_Block(?x3) \wedge \\ Function\_Block(?x4) \wedge \\ is\_a(?x1, ?x3) \wedge \\ is\_a(?x2, ?x4) \wedge \\ belongs\_to(?x1, ?x5) \wedge \\ belongs\_to(?x2, ?x6) \implies part\_of(?x1, ?x5) \wedge part\_of(?x2, ?x6) \end{aligned} \tag{16}$$

In regards to the layers, each instance included in the perspective will be linked to the three layers. This is achieved via the data property *needs*. The rules in Equations (17) and (18) present the reasoning of this connection.

$$\begin{aligned} Digital\_Twin\_Resource(?x1) \wedge \\ Function\_Block(?x2) \wedge \\ is\_a(?x1, ?x2) \wedge \\ Prespective(?x3) \wedge \\ needs(?x3, ?x1) \wedge \\ Model(?x4) \wedge \\ Signal(?x5) \wedge \\ Interface(?x6) \wedge \\ has(?x3, ?x4) \wedge \\ has(?x3, ?x5) \wedge \\ has(?x3, ?x6) \implies part\_of(?x2, ?x4) \wedge part\_of(?x2, ?x5) \wedge part\_of(?x2, ?x6) \end{aligned} \tag{17}$$

$$
\begin{aligned}
Digital\_Twin\_Component(?x1) \wedge \\
Function\_Block(?x2) \wedge \\
is\_a(?x1, ?x2) \wedge \\
Prespective(?x3) \wedge \\
needs(?x3, ?x1) \wedge \\
Model(?x4) \wedge \\
Signal(?x5) \wedge \\
Interface(?x6) \wedge \\
has(?x3, ?x4) \wedge \\
has(?x3, ?x5) \wedge \\
has(?x3, ?x6) \implies part\_of(?x2, ?x4) \wedge part\_of(?x2, ?x5) \wedge part\_of(?x2, ?x6)
\end{aligned} \quad (18)
$$

After the reasoning step, the architecture of the digital twin is persisted in *Digital Twin Architecture* ontology. This architecture includes the digital twin resources and the digital twin components as function blocks. Thus, the developer then needs to create the inferences function blocks in an FB editor. In this implementation, the development of the ontology and the reasoning has been performed using the Protege ontology editor [90] and Olingvo ontology editor [91]. As Figure 15 depicts, two main production resources have been identified, the robot, and transportation system which consists of the main conveyor and the by-pass conveyor. Following the reasoning flow, each one of these production resources will be represented in the digital twin world using a digital twin resource. These two resources combine several digital twin components which are reasoned via the aforementioned rules.

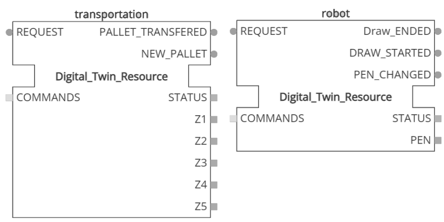


Figure 15. Two main production resources of the FASTory use case as digital twin resources. [Source: Own figure which is a snapshot from locally developed FB editor].

To simplify the presentation and satisfy the article's length, only the transportation system is selected. Looking in more detail at the transportation resource, a total of 16 digital twin components form the transportation digital twin resource as shown in Figure 16.

The *M2* block contains the RTU which controls the conveyor's tasks.Then, block *M1* contains the drives, motors and sensors that are related to the conveyor system. Finally, block *M0* contains the pulleys, the belts and the pallet.

At this point, the model layer has been populated, and the next stage is to look at the interface layer. As described in Section 3, the interface layer will depend on the digital twin perspectives. In this example, the digital twin has two perspectives, power consumption and resources utilization. As an example, from the power consumption perspective, the interface can include an HMI as a graph that shows the power in Watt units and APIs that use HTTP REST for publishing the consumption value. Figure 17 depicts the example for the needed function blocks for the interface and the signals.

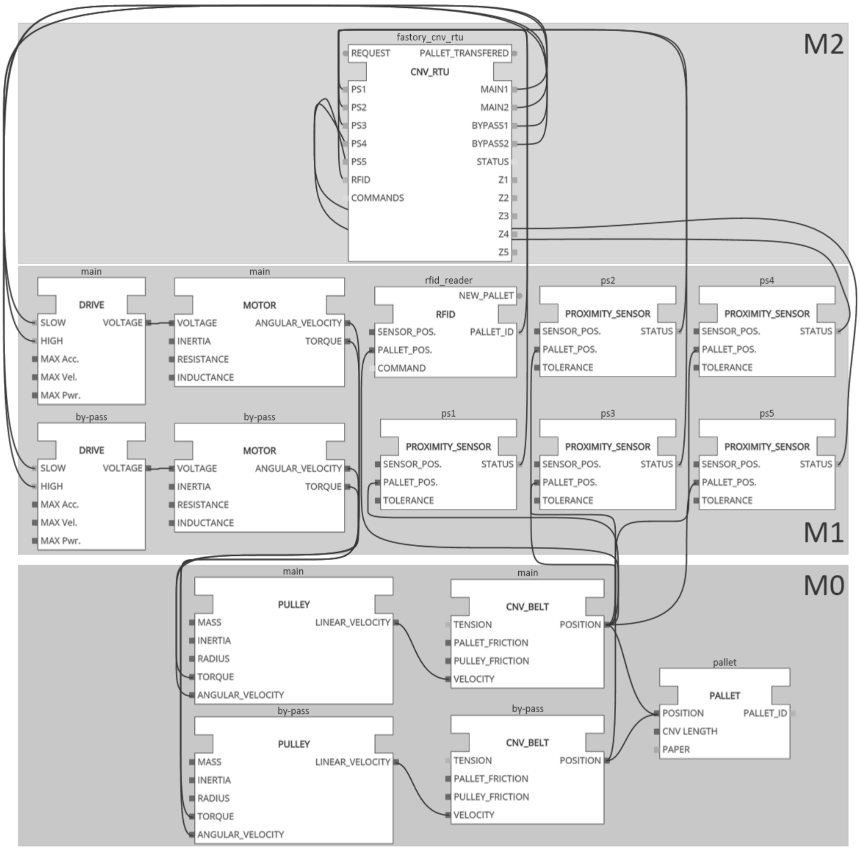


Figure 16. The composition of the transportation digital twin resource in the model layer. [Source: Own figure which is a snapshot from locally developed FB editor].

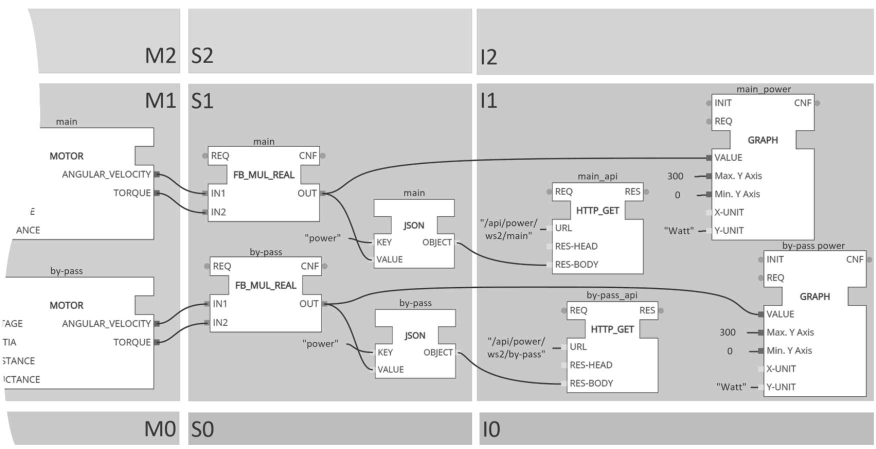


Figure 17. The signal and interface layers for the transportation digital resource in the FASTory use case. [Source: Own figure].

6. Discussion

The demonstrated approach in the previous section permits high flexibility for the developers to choose the desired components (i.e., function blocks) for the digital twin. The use case showed that the model layer is the core for building the digital twin. This layer depends mostly on the physical system descriptions including processes and equipment. Once the model layer is drafted, the interface layer can be developed to satisfy the objectives of the digital twin. This layer depends entirely on the perspectives of the digital twin and the features of the digital twin. Then, the crucially needed signal layer can be developed to connect the model layer with the interface layer. The three model levels were needed in building the digital twin for the transportation resource. However, only level 1 was needed in the signals and interfaces layers due to the fact that the shared information is at level 1. In this regard, the paradigm, which was presented in Section 3, can be implemented differently to satisfy the developers' needs. As an example, rather than having three levels for the interface, two levels can be introduced where one can be for the HMI and the other for the Machine–Machine Interface (i.e., APIs). Furthermore, the signals layer can be only one level where all conversions and calculations can be introduced. These variations are open to the developers to decide what is best for their use case.

For the qualitative attributes that are listed in Section 3, the FASTory implementation proved the expected outcomes. As an example, the *Modularity* attribute can be demonstrated by interchanging one of the function blocks, such as the controller or the drive, to a new different one. For the *Scalability*, the demonstrated use case showed the possibility of having two conveyors by replicating the associated components. The *Reusability* was evident in the use of the drive and the motor function blocks, where these two function blocks are from legacy implementation and are still possible to be reused. The *Interoperability* can be also guaranteed as the approach allows the user to use self-made function blocks or commercially available ones. In the demonstrated example, the system will work fine if the RTU is changed with a PLC or other type of controller from a different vendor. Lastly, the *Composability* is a feature of the IEC61499 standard which is included in this implementation. As an example, the transportation function block is composed of several function blocks.

The presented approach is still in its first versions and very well open for modifications and improvements. Thus, looking at the standard ISO 23247 and trying to find synergies with the presented approach is very well accepted. In this regard, both solutions dedicated separate modules for interfaces (user entity) and models (digital twin entity) which is important to guide the user in architecting such a complicated system. Both approaches addressed the information exchange as well. However, ISO 23247 did not dedicate a module for such an operation and left it open to the developers. While the presented approach considered these operations in the signal layer. The standard split the interfacing with the outside world into two entities (user and data) which might be necessary for more generic targeted domains. Overall, the presented approach appeared to be more specific for the manufacturing domain while the ISO 23247 standard targets a wider range of applications.

7. Conclusions and Future Work

Due to the versatility in exploiting them, digital twins b.ecame more essential in industrial applications. Adding to that, the benefits of saving energy, cost and time, and providing detailed insights about the behavior of systems prove to be very valuable. Thus, having a framework or paradigm will help the developers in creating such completed systems. In this regard, the conducted research provided guidelines or a paradigm for the developers to architect digital twins. The presented approach included multi-layer, multi-level and multi-perspective concepts for building a digital twin for manufacturing processes. Inspired by the ISA95 standard, the multi-level concept allows the user to allocate each component based on the level in the manufacturing pyramid. While the multi-layer concept is inspired by the onion architecture and helps the programmers and developers to draft the needed components based on the functionality. Finally, the multi-perspective concept allows the user to distinctly build the digital twin views or use cases. This combination of the concepts permits the creation of a modular, scalable, reusable, interoperable and composable digital twin as demonstrated in the paper. Then, to help illustrate the approach, an ontology-based implementation was presented. It was clearly seen that developing a digital twin requires substantial knowledge of the physical system that needs to be twined, the technical aspects and features of the development environment and the domain where the digital twin will be exploited. After the demonstration, a comparison with the ISO 23247 standard was drafted to find synergies between the two approaches. Hence, it was observed clearly that digital twins require three main aspects; modeling, interfacing and information exchange. These three main aspects were found in both approaches but with different implementations and conceptualizations. Overall, the differences between the approach can be seen in the targeted outcomes as the standard focuses on a wider range of applications while the presented approach in this paper targets manufacturing processes specifically. In fact, this comparison opens the doors for future work which may include addressing digital twinning for products and spatial computing. Moreover, future research may include extending the ontology concepts in this research to be compatible with the IOF vocabulary.

Author Contributions: Conceptualization, W.M.M. and J.L.M.L.; methodology, W.M.M.; software, W.M.M.; validation, W.M.M.; formal analysis, W.M.M.; investigation, W.M.M.; resources, W.M.M. and J.L.M.L.; data curation, W.M.M.; writing—original draft preparation, W.M.M.; writing—review and editing, W.M.M., R.E.H. and J.L.M.L.; visualization, W.M.M.; supervision, R.E.H. and J.L.M.L.; project administration, J.L.M.L.; funding acquisition, J.L.M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The research work by Wael M. Mohammed is supported by the Doctoral School of the President call 2018 of the Tampere University of Technology.

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

References

1. Mohammed, W.M.; Lobov, A.; Ferrer, B.R.; Iarovyi, S.; Lastra, J.L.M. A web-based simulator for a discrete manufacturing system. In Proceedings of the IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 6583–6589. https://doi.org/10.1109/IECON.2016.7793563.
2. Gelbart, N.R. *The King'S Midwife: A History and Mystery of Madame du Coudray*; University of California Press: Berkeley, CA, USA; London, UK, 1999.
3. Hormaza, L.A.; Mohammed, W.M.; Ferrer, B.R.; Bejarano, R.; Martinez Lastra, J.L. On-line Training and Monitoring of Robot Tasks through Virtual Reality. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki, Finland, 22–25 July 2019; pp. 841–846. https://doi.org/10.1109/INDIN41052.2019.8971967.
4. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002.
5. Grieves, M. Digital Twin: Manufacturing Excellence through Virtual Factory Replication. *White Paper* **2015**.
6. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2405–2415. https://doi.org/10.1109/TII.2018.2873186.
7. Barricelli, B.R.; Casiraghi, E.; Fogli, D. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* **2019**, *7*, 167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499.
8. VanDerHorn, E.; Mahadevan, S. Digital Twin: Generalization, characterization and implementation. *Decis. Support Syst.* **2021**, *145*, 113524. https://doi.org/10.1016/j.dss.2021.113524.
9. Guerra, R.H.; Quiza, R.; Villalonga, A.; Arenas, J.; Castano, F. Digital Twin-Based Optimization for Ultraprecision Motion Systems With Backlash and Friction. *IEEE Access* **2019**, *7*, 93462–93472. https://doi.org/10.1109/ACCESS.2019.2928141.
10. Grieves, M. Intelligent Digital Twins: The Role of AI and ML in the Future of Digital Twins Chief Scientist of Advanced Manufacturing Florida Institute of Technology 2021, *unpublished*. https://doi.org/10.13140/RG.2.2.17464.67849.
11. Mohammed, W.M.; Lastra, J.L.M. FASTory Digital Twin Data. *Data Brief* **2021**, *35*, 106912. https://doi.org/10.1016/j.dib.2021.106912.

12. Cheng, K.; Wang, Q.; Yang, D.; Dai, Q.; Wang, M. Digital-Twins-Driven Semi-Physical Simulation for Testing and Evaluation of Industrial Software in a Smart Manufacturing System. *Machines* **2022**, *10*, 388. https://doi.org/10.3390/machines10050388.
13. Nie, Z.; Cao, G.; Zhang, P.; Peng, Q.; Zhang, Z. Multi-Analogy Innovation Design Based on Digital Twin. *Machines* **2022**, *10*, 652. https://doi.org/10.3390/machines10080652.
14. Cruz, Y.J.; Rivas, M.; Quiza, R.; Haber, R.E.; Castaño, F.; Villalonga, A. A two-step machine learning approach for dynamic model selection: A case study on a micro milling process. *Comput. Ind.* **2022**, *143*, 103764. https://doi.org/10.1016/j.compind.2022.103764.
15. Villalonga, A.; Negri, E.; Biscardo, G.; Castano, F.; Haber, R.E.; Fumagalli, L.; Macchi, M. A decision-making framework for dynamic scheduling of cyber-physical production systems based on digital twins. *Annu. Rev. Control.* **2021**, *51*, 357–373. https://doi.org/10.1016/j.arcontrol.2021.04.008.
16. Khan, M.G.; Huda, N.U.; Zaman, U.K.U. Smart Warehouse Management System: Architecture, Real-Time Implementation and Prototype Design. *Machines* **2022**, *10*, 150. https://doi.org/10.3390/machines10020150.
17. Cimino, C.; Negri, E.; Fumagalli, L. Review of digital twin applications in manufacturing. *Comput. Ind.* **2019**, *113*, 103130. https://doi.org/10.1016/j.compind.2019.103130.
18. Zhang, J.; Deng, T.; Jiang, H.; Chen, H.; Qin, S.; Ding, G. Bi-level dynamic scheduling architecture based on service unit digital twin agents. *J. Manuf. Syst.* **2021**, *60*, 59–79. https://doi.org/10.1016/j.jmsy.2021.05.007.
19. Fang, Y.; Peng, C.; Lou, P.; Zhou, Z.; Hu, J.; Yan, J. Digital-Twin-Based Job Shop Scheduling Toward Smart Manufacturing. *IEEE Trans. Ind. Inform.* **2019**, *15*, 6425–6435. https://doi.org/10.1109/TII.2019.2938572.
20. Resman, M.; Pipan, M.; Simic, M.; Herakovic, N. A new architecture model for smart manufacturing: A performance analysis and comparison with the RAMI 4.0 reference model. *Adv. Prod. Eng. Manag.* **2019**, *14*, 153–165. https://doi.org/10.14743/apem2019.2.318.
21. Josifovska, K.; Yigitbas, E.; Engels, G. Reference Framework for Digital Twins within Cyber-Physical Systems. In Proceedings of the 2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS), Montreal, QC, Canada, 28 May 2019; pp. 25–31. https://doi.org/10.1109/SEsCPS.2019.00012.
22. Tekinerdogan, B.; Verdouw, C. Systems Architecture Design Pattern Catalog for Developing Digital Twins. *Sensors* **2020**, *20*, 5103. https://doi.org/10.3390/s20185103.
23. Li, X.; He, B.; Wang, Z.; Zhou, Y.; Li, G.; Jiang, R. Semantic-Enhanced Digital Twin System for Robot–Environment Interaction Monitoring. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. https://doi.org/10.1109/TIM.2021.3066542.
24. Mattila, J.; Ala-Laurinaho, R.; Autiosalo, J.; Salminen, P.; Tammi, K. Using Digital Twin Documents to Control a Smart Factory: Simulation Approach with ROS, Gazebo, and Twinbase. *Machines* **2022**, *10*, 225. https://doi.org/10.3390/machines10040225.
25. van der Valk, H.; Haße, H.; Möller, F.; Arbter, M.; Henning, J.L.; Otto, B. A Taxonomy of Digital Twins. *AMCIS 2020 Proc.* **2020**, *4*.
26. Redelinghuys, A.J.H.; Kruger, K.; Basson, A. A Six-Layer Architecture for Digital Twins with Aggregation. In *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*; Borangiu, T., Trentesaux, D., Leitão, P., Giret Boggino, A., Botti, V., Eds.; Series Title: Studies in Computational Intelligence; Springer International Publishing: Cham, Switzerland, 2020; Volume 853, pp. 171–182. https://doi.org/10.1007/978-3-030-27477-1_13.
27. Resman, M.; Protner, J.; Simic, M.; Herakovic, N. A Five-Step Approach to Planning Data-Driven Digital Twins for Discrete Manufacturing Systems. *Appl. Sci.* **2021**, *11*, 3639. https://doi.org/10.3390/app11083639.
28. Somers, S.; Oltramari, A.; Lebiere, C. Cognitive twin: A cognitive approach to personalized assistants. In *Proceedings of the CEUR Workshop Proc*; Martin, A., Hinkelmann, K., Fill, H.-G., Gerber, A., Lenat, D., Stolle R., van Harmelen F., Eds.; CEUR-WS: Palo Alto, CA, USA, 2020; Volume 2600.
29. Zheng, X.; Lu, J.; Kiritsis, D. The emergence of cognitive digital twin: vision, challenges and opportunities. *Int. J. Prod. Res.* **2021**, *0*, 1–23. https://doi.org/10.1080/00207543.2021.2014591.
30. Abburu, S.; Berre, A.; Jacoby, M.; Roman, D.; Stojanovic, L.; Stojanovic, N. COGNITWIN - Hybrid and Cognitive Digital Twins for the Process Industry. In Proceedings of the 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 15–17 June 2020. https://doi.org/10.1109/ICE/ITMC49519.2020.9198403.
31. Lu, J., Zheng, X., Gharaei, A., Kalaboukas, K., Kiritsis, D. Cognitive Twins for Supporting Decision-Makings of Internet of Things Systems. In *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing*; Wang L., Majstorovic, V.D., Mourtzis D., Carpanzano E., Moroni G., Galantucci L.M.; Springer International Publishing: Cham, Switzerland, 2020;pp. 105–115. https://doi.org/10.1007/978-3-030-46212-3_7.
32. Rožanec, J.; Lu, J.; Rupnik, J.; Škrjanc, M.; Mladenić, D.; Fortuna, B.; Zheng, X.; Kiritsis, D. Actionable cognitive twins for decision making in manufacturing. *Int. J. Prod. Res.* **2022**, *60*, 452–478. https://doi.org/10.1080/00207543.2021.2002967.
33. Rožanec, J.; Jinzhi, L.; Košmerlj, A.; Kenda, K.; Dimitris, K.; Jovanoski, V.; Rupnik, J.; Karlovčec, M.; Fortuna, B. Towards actionable cognitive digital twins for manufacturing. In *Proceedings of the CEUR Workshop Proc*; Garcia-Castro, R., Davies, J., Antoniou, G., Fortuna, C., Eds.; CEUR-WS: Palo Alto, CA, USA, 2020; Volume 2615.
34. Li, H.; Wang, G.; Lu, J.; Kiritsis, D. Cognitive twin construction for system of systems operation based on semantic integration and high-level architecture. *Integr. -Comput. -Aided Eng.* **2022**, *29*, 277–295. https://doi.org/10.3233/ICA-220677.
35. digitaltwin.io. Digital Twin Knowledge Repository-Platforms. Available online: https://digitaltwin.io/platforms.html (accessed on 16 February 2021).
36. General Electric. Digital Twin | GE Digital. 2021. Available online: https://www.ge.com/digital/applications/digital-twin (accessed on 16 February 2021).

37. Colin Parris. What is a Digital Twin? Available online: https://www.ge.com/digital/blog/what-digital-twin (accessed on 29 August 2022).
38. IBM. Exchange | IBM Digital Twin Exchange. Available online: https://digitaltwinexchange.ibm.com/exchange/ (accessed on 16 February 2021).
39. PTC. What Is Digital Twin? | PTC. Available online: https://www.ptc.com/en/industry-insights/digital-twin (accessed on 16 February 2021).
40. McMahon, C.; Dertien, S. State of Digital Twin. Available online: https://www.ptc.com/-/media/Files/PDFs/Manufacturing/State-of-Digital-Twin-2022.pdf (accessed on 22 August 2022).
41. Microsoft. Azure Digital Twins. Available online: https://sourceforge.net/software/product/Azure-Digital-Twins/ (accessed on 16 February 2021).
42. baanders. Data history (with Azure Data Explorer)-Azure Digital Twins. Available online: https://docs.microsoft.com/en-us/azure/digital-twins/concepts-data-history (accessed on 22 August 2022).
43. Microsoft. Digital Twins–Modeling and Simulations | Microsoft Azure. Available online: https://azure.microsoft.com/en-us/services/digital-twins/ (accessed on 18 February 2021).
44. Ansys. Twin Builder: Digital Twin Predictive Maintenance Software | Ansys. Available online: https://www.ansys.com/products/systems/ansys-twin-builder (accessed on 16 February 2021).
45. Ansys Inc. White Paper on Digital Twins: Making the Vision Achievable. Available online: https://www.ansys.com/resource-center/white-paper/digital-twins-making-vision-achievable (accessed on 22 August 2022).
46. Ansys Inc. White Paper on How Simulation-Based Digital Twins and the Industrial Internet of Things Can Improve Product and Process Performance. Available online: https://www.ansys.com/resource-center/white-paper/how-sim-based-digital-twins-and-the-industrial-iot-can-improve-product-and-process-perf (accessed on 18 February 2021).
47. SAP. SAP Digital Twin Software&Technology. Available online: https://www.sap.com/finland/products/supply-chain-management/digital-twin.html (accessed on 16 February 2021).
48. SAP Leonardo. The Network of Digital Twins. Available online: https://community.sap.com/topics/digital-supplier-network (accessed on 22 August 2022).
49. Oracle Corporation. About the IoT Digital Twin Framework. Available online: https://docs.oracle.com/en/cloud/paas/iot-cloud/iotgs/iot-digital-twin-framework.html (accessed on 16 February 2021).
50. Bosch. When digitalization shows us the way. Available online: https://www.boschbuildingsolutions.com/xc/en/news-and-stories/digitalization/ (accessed on 16 February 2021).
51. Emerson. Digital Twin Solutions | Emerson US. Available online: https://www.emerson.com/en-us/automation/operations-business-management/dynamic-simulation/digital-twin-solutions (accessed on 16 February 2021).
52. Emerson. Emerson Digital Twin: A Key Technology for Digital Transformation. Available online: https://www.emerson.com/documents/automation/white-paper-emerson-digital-twin-a-key-technology-for-digital-transformation-mimic-en-5262472.pdf (accessed on 22 August 2022).
53. ABB. Digital twin applications-What is new | ABB. Available online: https://new.abb.com/control-systems/features/digital-twin-applications (accessed on 16 February 2021).
54. Mathworks. Digital Twins for Predictive Maintenance-04. Available online: https://explore.mathworks.com/digital-twins-for-predictive-maintenance/chapter-04-36US-761YW.html (accessed on 22 August 2022).
55. Mathworks. Digital Twins for Predictive Maintenance. Available online: https://explore.mathworks.com/digital-twins-for-predictive-maintenance/chapter-05-36US-763YW.html (accessed on 16 February 2021).
56. COMSOL Multiphysics® Software-Understand, Predict, and Optimize. Available online: https://www.comsol.com/comsol-multiphysics (accessed on 24 February 2021).
57. NVIDIA. NVIDIA Omniverse for Digital Twins. Available online: https://www.nvidia.com/en-us/omniverse/solutions/digital-twins/ (accessed on 18 August 2022).
58. NVIDIA. Develop on NVIDIA Omniverse Platform. Available online: https://developer.nvidia.com/nvidia-omniverse-platform (accessed on 22 August 2022).
59. Visual Components. Digital Twins and Virtual Commissioning in the Manufacturing Industry (Updated for 2022). Available online: https://www.visualcomponents.com/resources/blog/digital-twins-and-virtual-commissioning-in-industry-4-0/ (accessed on 6 September 2022).
60. Siemens. Tecnomatix | Siemens Software. Available online: https://www.plm.automation.siemens.com/global/en/products/tecnomatix/ (accessed on 6 September 2022).
61. iconics. Make Better Decisions About Smart Building Operations with Digital Twin Technology | ICONICS Software Solutions. Available online: https://iconics.com/Resources/ICONICS-Blog/2022/Make-Better-Decisions-About-Smart-Building-Operations-with-Digital-Twin-Technology (accessed on 22 August 2022).
62. Microsoft. Pushing the boundaries of renewable energy production efficiency with Azure Digital Twins. Available online: https://customers.microsoft.com/en-in/story/848311-doosan-manufacturing-azure-digital-twins (accessed on 22 August 2022).
63. Visual Components. Product Overview-Visual Components 4.5. Available online: https://www.visualcomponents.com/products/ (accessed on 6 September 2022).

64. Foundation, B. Blender is for Simulation. Available online: https://www.blender.org/features/simulation/ (accessed on 22 August 2022).
65. Maya Software | Get Prices and Buy Maya 2023 | Autodesk. Available online: https://www.autodesk.com/products/maya/overview (accessed on 22 August 2022).
66. Caulfield, B. NVIDIA, BMW Blend Reality, Virtual Worlds to Demonstrate Factory of the Future. Available online: https://blogs.nvidia.com/blog/2021/04/13/nvidia-bmw-factory-future/ (accessed 22 August 2022).
67. Wang, K.; Wang, Y.; Li, Y.; Fan, X.; Xiao, S.; Hu, L. A review of the technology standards for enabling digital twin. *Digit. Twin* **2022**, *2*, 4. https://doi.org/10.12688/digitaltwin.17549.1.
68. International Organization for Standardization (ISO). ISO 23247 Automation systems and integration—Digital twin framework for manufacturing. Available online: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/50/75066.html (accessed on 22 August 2022).
69. Shao, G. Use Case Scenarios for Digital Twin Implementation Based on ISO 23247. Available online: https://doi.org/10.6028/NIST.AMS.400-2 (accessed on 22 August 2022)
70. International Organization for Standardization (ISO). ISO/TR 24464:2020 Automation systems and integration—Industrial data—Visualization elements of digital twins. https://www.iso.org/standard/78836.html (accessed on 22 August 2022).
71. International Organization for Standardization (ISO). ISO/IEC AWI 30172 Digital Twin—Use cases. Available online: https://www.iso.org/standard/81578.html (accessed on 22 August 2022).
72. International Organization for Standardization (ISO). ISO/IEC AWI 30173 Digital twin — Concepts and terminology. Available online: https://www.iso.org/standard/81442.html (accessed on 22 August 2022).
73. Andika, L.; Dyck, A.; Lichter, H. Towards A Design for An Extendable Reporting Interface. *Procedia Comput. Sci.* **2017**, *116*, 318–325. https://doi.org/10.1016/j.procs.2017.10.082.
74. Khalil, M.E.; Ghani, K.; Khalil, W. Onion architecture: A new approach for XaaS (every-thing-as-a service) based virtual collaborations. In Proceedings of the 2016 13th Learning and Technology Conference (L&T), Jeddah, Saudi Arabia, 10–11 April 2016; pp. 1–7. doi:10.1109/LT.2016.7562859.
75. Jeffrey Palermo. The Onion Architecture: Part 1. Available online: https://jeffreypalermo.com/2008/07/the-onion-architecture-part-1/ (accessed on 22 November 2021).
76. Robert C. Martin. The Clean Architecture. Available online: https://blog.cleancoder.com/uncle-bob/2012/08/13/the-clean-architecture.html (accessed on 22 November 2021).
77. Martin, R.C. *Clean Architecture: A Craftsman'S Guide to Software Structure and Design*, 1st Ed.; Martin series; Prentice Hall: London, UK, 2018.
78. ISA (Society). *Enterprise-Control System Integration. Models and Terminology*; ISA: Research Triangle Park, NC, USA, 2000.
79. IOF. Technical Principles–IOF. Available online: https://www.industrialontologies.org/technical-principles/ (accessed on 6 September 2022).
80. J Neil Otte. IOF-BFO. Available online: https://github.com/NCOR-US/IOF-BFO (accessed on 6 Septemebr 2022).
81. Kulvatunyou, B.S.; Wallace, E.; Kiritsis, D.; Smith, B.; Will, C. The Industrial Ontologies Foundry Proof-of-Concept Project. In Proceedings of the APMS 2018 International Conference Advances in Production Management Systems (APMS 2018), Seoul, Korea, 26–30 August 2018.
82. Garetti, M.; Fumagalli, L.; Negri, E. Role of Ontologies for CPS Implementation in Manufacturing. *Manag. Prod. Eng. Rev.* **2015**, *6*, 26–32. https://doi.org/10.1515/mper-2015-0033.
83. Ontologies-EU Vocabularies-Publications Office of the EU. Available online: https://op.europa.eu/en/web/eu-vocabularies/ontologies (accessed on 14 June 2022).
84. Hyvönen, E.; Tuominen, J.; Alonen, M.; Mäkelä, E. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In *The Semantic Web: ESWC 2014 Satellite Events*; Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A., Eds.; Springer International Publishing: Cham, Switzerland, 2014; Volume 8798, pp. 226–230. https://doi.org/10.1007/978-3-319-11955-7_24.
85. Mohammed, W.M.; Ramis Ferrer, B.; Iarovyi, S.; Negri, E.; Fumagalli, L.; Lobov, A.; Martinez Lastra, J.L. Generic platform for manufacturing execution system functions in knowledge-driven manufacturing systems. *Int. J. Comput. Integr. Manuf.* **2018**, *31*, 262–274. https://doi.org/10.1080/0951192X.2017.1407874.
86. Iarovyi, S.; Mohammed, W.M.; Lobov, A.; Ferrer, B.R.; Lastra, J.L.M. Cyber–Physical Systems for Open-Knowledge-Driven Manufacturing Execution Systems. *Proc. IEEE* **2016**, *104*, 1142–1154. https://doi.org/10.1109/JPROC.2015.2509498.
87. Ramis Ferrer, B.; Mohammed, W.M.; Ahmad, M.; Iarovyi, S.; Zhang, J.; Harrison, R.; Martinez Lastra, J.L. Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing. *Knowl. Inf. Syst.* **2021**, *63*, 1271–1304. https://doi.org/10.1007/s10115-021-01558-4.
88. Ramis Ferrer, B.; Muhammad, U.; Mohammed, W.; Martínez Lastra, J. Implementing and Visualizing ISO 22400 Key Performance Indicators for Monitoring Discrete Manufacturing Systems. *Machines* **2018**, *6*, 39. https://doi.org/10.3390/machines6030039.
89. Ferrer, B.R.; Mohammed, W.M.; Chen, E.; Lastra, J.L.M. Connecting web-based IoT devices to a cloud-based manufacturing platform. In Proceedings of the IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 29 October–1 November 2017; pp. 8628–8633. https://doi.org/10.1109/IECON.2017.8217516.

90. Stanford. Protégé-A free, open-source ontology editor and framework for building intelligent systems. Available online: https://protege.stanford.edu/products.php#desktop-protege (accessed on 6 September 2022).
91. FAST-Lab, Tampere University. Olingvo—An application for editing RDF-Based ontologies. Available online: https://www.zdmp.eu/iprdocuments/subcall/Olingvo (accessed on 6 September 2022).

PUBLICATION V

FASTory digital twin data

Wael M Mohammed, Jose L Martinez Lastra

Data in brief, 2021, 35,106912
Doi: 10.1016/j.dib.2021.106912

Data in Brief 35 (2021) 106912

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

FASTory digital twin data

Wael M. Mohammed*, Jose L. Martinez Lastra

FAST-Lab, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 600, Tampere FI-33014, Finland

A R T I C L E I N F O

Article history:
Received 22 January 2021
Revised 17 February 2021
Accepted 25 February 2021
Available online 26 February 2021

Keywords:
Digital twin
Data engineering
Linked data
Discrete manufacturing process
Assembly process

A B S T R A C T

The vast adoption of machine learning techniques in developing smart solutions increases the need of training and testing data. This data can be either collected from physical systems or created using simulation tools. In this regard, this paper presents a set of data collected using a digital twin known as the FASTory Simulator. The data contains more than 100 K events which are collected during a simulated assembly process. The FASTory simulator is a replica of a real assembly line with web-based industrial controllers. The data have been collected using specific-developed orchestrator. During the simulated process, the orchestrator was able to record all the events that occurred in the system. The provided data contains raw JavaScript Object Notation (JSON) formatted data and filtered Comma Separated Values (CSV) formatted data. This data can be exploited in machine learning for modelling the behaviour of the production systems or as testing data for optimization solution for the production system. Finally, this data has been utilized in a research for comparing different data analysis approaches including Knowledge-based systems and data-based systems.

* Corresponding author.
E-mail address: wael.mohammed@tuni.fi (W.M. Mohammed).

https://doi.org/10.1016/j.dib.2021.106912

Specifications Table

Subject	Industrial Engineering, Manufacturing Engineering
Specific subject area	Discrete manufacturing data includes the events of simulated assembly process
Type of data	CSV, JSON
How data were acquired	The data has been collected using a digital twin known as the FASTory Simulator [1]. This simulator mimics an actual assembly system located in Tampere university. The simulator was controlled by an inhouse developed orchestrator for the purpose of data collection using web services.
Data format	Raw and filtered
Parameters for data collection	The collected data contains 106,154 events during a simulated assembly process [2]. The collection process lasted for 12 h distributed on 3 days. For the data collection, an orchestrator has been set for managing 3 pallets at the same time with continues utilization.
Description of data collection	The data was collected using an orchestrator which acted as event logger for this experiment. The orchestrator uses REST-formatted HTTP notifications for generating the events. The orchestrator needed to subscribe to all notifications in order to collect the data. Besides, the orchestrator managed the assembly of 1017 simulated product. [3]
Data source location	FASTory Simulator [3] Future Automation Systems and Technologies Laboratory (FAST-Lab) Tampere University Tampere, Finland
Data accessibility	Dataset: https://doi.org/10.23729/7e299722-c246-4695-990f-dbe5a7839ad2 [2] Software (simulator and orchestrator): https://doi.org/10.5281/zenodo.4541008 [3]
Related research article	R.F. Borja, M. Wael M., A. Mussawar, I. Sergii, Z. Jiayi, H. Robert, M.L. Jose L., Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing, Knowledge and Information Systems, In Press.

Value of the Data

- The collected data include more than 100 K events generated during a simulated assembly process of electronic devices using the FASTory Simulator. The data can be important for the purposes of studying different techniques and methods for balancing and optimizing discrete processes.
- The data can be utilized by researchers and developer of manufacturing systems that are specialized in optimizing manufacturing processes. The data can also be important for companies who develop Manufacturing Execution Systems (MES) and Enterprise Resources Planning (ERP) systems [4].
- The FASTory Digital Twin data can be exploited in modelling discrete manufacturing processes or testing and validating optimization solutions for balancing the utilization of the manufacturing resources. In addition, the data can be used for training machine learning models that targets replicating an industrial discrete manufacturing process.

1. Data Description

1.1. Data generation using the FASTory digital twin

The FASTory line is an education and research assembly line located at the FAST-Lab in Tampere University Tampere, Finland. The assembly line was utilized for assembling mobile phones during the early 2000s. Afterwards, the line was relocated to the university and retrofitted with web-based Remote Terminal Units (RTUs) to control the conveyors and the robots alongside handling other sensors and actuators as Fig. 1 shows on the right. In addition, the line is transformed to mimic the assembly process by drawing main components of mobile phones (frame, screen, and keyboard) with 3 different models and 3 possible colours for each part. These

Fig. 1. On the left, the FASTory line product and on the right, the FASTory line [1].

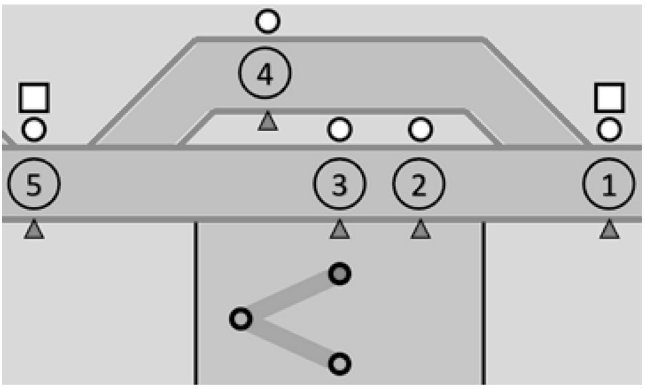


Fig. 2. The location of zones in a workstation numbered from 1 to 5.

variations generate a total of 729 different product. As show in Fig. 1 on the left, the pallets have been modified to hold the papers which represents the main component in the assembly. Once the paper is placed on a pallet, the drawing processes of the three parts are executed. Then the paper is extracted from the pallet. In this regard, the paper with the drawn phone represent a fully assembled product.

The FASTory line consists of 10 identical workstations (numbered from 2 to 6 and 8 to 12) for executing the drawing processes, one workstation (numbered as 1) for loading and unloading the papers and one workstation (numbered as 7) for loading and unloading pallets. Each workstation is equipped with RFID for recognizing the incoming pallets. In addition, each workstation (2–6 and 8–12) include a main conveyor and bypass conveyor. The main conveyor is used to locate the pallet in the position for drawing the mobile parts. Meanwhile, the bypass conveyor is used to skip the workstation without disturbing the ongoing process. For workstation 1 and 7, only main conveyor is used. Both main and bypass conveyors are divided into zones. These zones are numbered from 1 to 5 as Fig. 2 depicts. Each zone is equipped with a proximity sensor to detect the presence of the pallet and with a pneumatic actuator to hold the pallet in the zone. The functionality of these zones include:

- Zone 1 for identifying the incoming pallet;
- Zone 2 for holding 1 pallet in the workstation while the robot is executing a drawing process. This zone works as a buffer for queueing the pallet;
- Zone 3 is dedicated for executing the drawing process;

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "title": "event",
  "type": "object",
  "properties": {
    "timeStamp": {"type": "integer"},
    "event": {
      "type": "object",
      "properties": {
        "id": {"type": "string"},
        "senderId": {"type": "string"},
        "payload": {
          "type": "object",
          "properties": {
            "palletId": {"type": "string"},
            "recipe": {"type": "string"},
            "color": {"type": "string"},
          },
          "required": ["palletId"]
        }
      },
      "required": ["id","senderId","payload" ]
    }
  },
  "required": ["timeStamp","event"]
}
```


Fig. 3. JSON schema of the events.

- Zone 4 is used to bypass the workstation and allows holding the pallet to avoid any collision in the conjunction of the main and the bypass conveyors. This zone does not exist in workstation 1 and workstation 7;
- Zone 5 represents the exit of the workstation. This zone is overlapping with the next workstation's zone 1.

As mentioned above, the FASTory line is controlled by web based RTUs that use the HTTP protocol. In fact, each conveyor and robot is controlled by an RTU. In total, 23 RTUs (12 conveyors and 11 robots) are used for controlling purposes. These RTUs accept HTTP services to execute operations like drawing, paper handling and pallets handling. In addition, the RTUs employ publish/subscribe mechanism for notifying third party applications about the occurring events. The schema of these events is depicted in Fig. 3. As shown in the figure, each event contains a time stamp representing the issuing time, event id identifying the event type and sender id identifying the equipment that generated the event. Afterwards, each event includes a payload that consists of the event-specific information based on the event id. Based on the event's description, the following are the possible events that can be generated by the FASTory line:

- Z1_Changed: this event is triggered once a pallet arrives or leaves the zone. If the pallet arrives to the zone, the payload will contain the pallet id. If the pallet leaves the zone, the payload will contain "−1" as pallet id;
- Z2_Changed: similar to Z1_Changed but for zone 2, see Figs. 4 and 5;
- Z3_Changed: similar to Z1_Changed but for zone 3;
- Z4_Changed: similar to Z1_Changed but for zone 4;
- Z5_Changed: similar to Z1_Changed but for zone 5;
- DRAW_START_EXECUTION: this event is triggered once the process of drawing is started. The payload includes the pallet id where the drawing is conducted and the recipe of the drawing, see Fig. 7. the recipe represents the part and model of the phone;
- DRAW_END_EXECUTION: this event is triggered once the process of drawing is ended. The payload includes the pallet id where the drawing has been conducted, the recipe of the drawing and the colour of drawing, see Fig. 6;
- PAPER_LOADED: is triggered once the pallet receives a new paper. The payload includes the pallet id, see Fig. 9;

```
{
  "timeStamp":1501751782579,
  "event":{
    "id":"Z2_Changed",
    "senderId":"SIM_CNV11",
    "payload":{
      "palletId":"1501749696163"
    }
  }
}
```


Fig. 4. Example of zone changed event once the pallet leaves the zone.


```
{
  "timeStamp":1501751789604,
  "event":{
    "id":"Z2_Changed",
    "senderId":"SIM_CNV12",
    "payload":{
      "palletId":"-1"
    }
  }
}
```


Fig. 5. Example of zone changed event when the pallet arrives to the zone.


```
{
  "timeStamp":1501751790656,
  "event":{
    "id":"DRAW_END_EXECUTION",
    "senderId":"SIM_ROB4",
    "payload":{
      "palletId":"1501749696165",
      "recipe":"1",
      "color":"BLUE"
    }
  }
}
```


Fig. 6. Example of draw end execution event.

- PAPER_UNLOADED: is triggered once the paper is unloaded form the pallet. The payload includes the pallet id, see Fig. 8;
- PEN_CHANGED: this event is triggered once the robot changes the colour of the pen. This event is not included in the data set due to the fixed colour for each workstation.

Table 1
Description of the fields of the filtered and rearranged data.

Field name	Data type	Possible values
timeStamp	Integer	Integer represents the time according to universal time. Which is calculated since 1 January 1970 00:00:00 in milliseconds
eventId	String	Z1_Changed, Z2_Changed, Z3_Changed, Z4_Changed, Z5_Changed, PAPER_LOADED, PAPER_UNLOADED, DRAW_START_EXECUTION, DRAW_END_EXECUTION
senderId	String	SIM_CNV1, SIM_CNV2, SIM_CNV3, SIM_CNV4, SIM_CNV5, SIM_CNV6, SIM_CNV7, SIM_CNV8, SIM_CNV9, SIM_CNV10, SIM_CNV11, SIM_CNV12, SIM_ROB1, SIM_ROB2, SIM_ROB3, SIM_ROB4, SIM_ROB5, SIM_ROB6, SIM_ROB8, SIM_ROB9, SIM_ROB10, SIM_ROB11, SIM_ROB12
palletId	String	String of integer with 15 digits.
recipe	String	1,2,3,4,5,6,7,8,9 (1,2,3 for frame, 4,5,6 for screen, 7,8,9 for keyboard)
colour	String	red, green, blue


```
{
  "timeStamp":1501751786648,
  "event":{
    "id":"DRAW_START_EXECUTION",
    "senderId":"SIM_ROB4",
    "payload":{
      "palletId":"1501749696165",
      "recipe":"1"
    }
  }
}
```


Fig. 7. Example of draw start execution event.


```
{
  "timeStamp":1501751142646,
  "event":{
    "id":"PAPER_UNLOADED",
    "senderId":"SIM_ROB1",
    "payload":{
      "palletId":"1501749696165"
    }
  }
}
```


Fig. 8. Example of paper unloaded event.

1.2. Raw and filtered data

Besides sharing the raw data as JSON files, a filtering process has been conducted to rearrange the data as CSV. This data includes the similar content as described above but in computer-friendly format. Table 1 presents the description of the filtered data. It is important to mentioned in the CSV formatted data, the field recipe and colour are empty for the events that do not include a colour or a recipe.

```json
{
  "timeStamp":1501751149235,
  "event":{
    "id":"PAPER_LOADED",
    "senderId":"SIM_ROB1",
    "payload":{
      "palletId":"1501749696165"
    }
  }
}
```


Fig. 9. Example of paper loaded event.

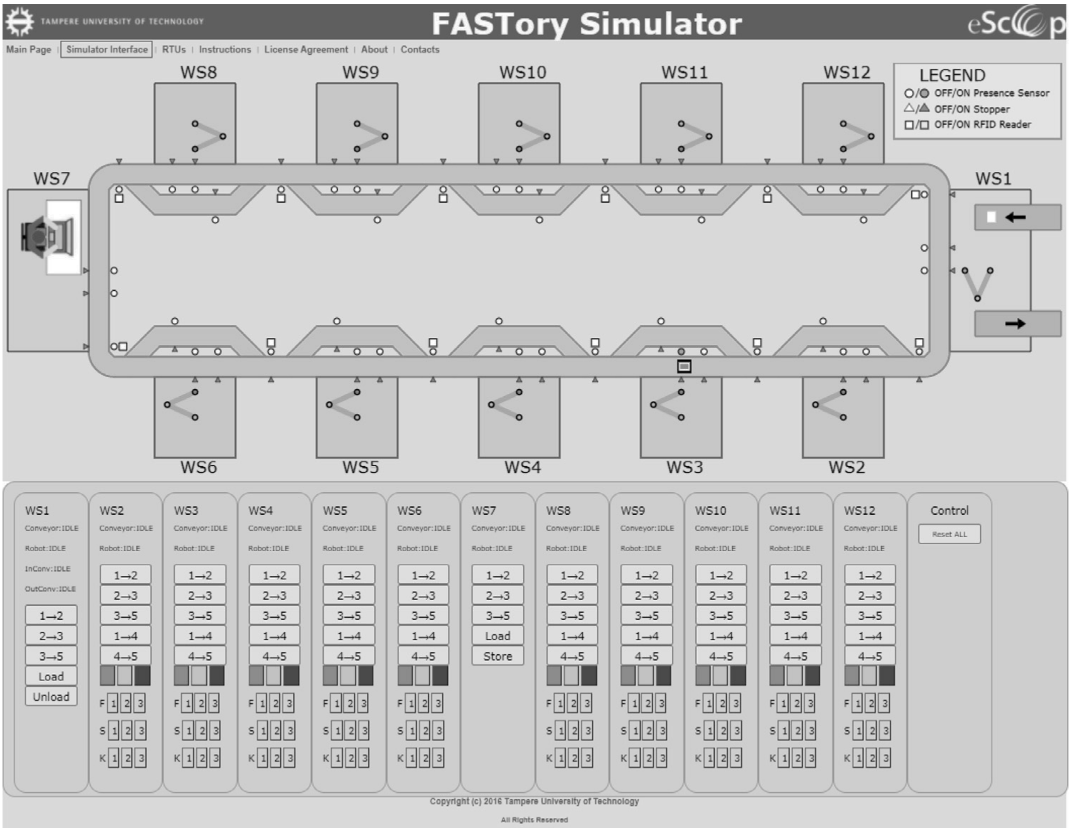


Fig. 10. FASTory digital twin [3,5].

2. Experimental Design, Materials and Methods

During the eScop research project [6], the FASTory Simulator[1] has been developed to help the project developers to test the various project applications remotely [7,8]. This allowed the developers to continuously develop and test the applications without producing any disruption or harm. For this paper, the collected data has been gathered using the FASTory simulator. This

[1] Accessible using: http://130.230.141.228:3000/.

simulator is a digital twin that mimics and simulates the services, data, timing and events. Fig. 10 depicts the interface of the simulator.

The process of gathering the data involved developing an orchestrator that subscribes to all events in the simulator and then start production process of large number of orders aiming at collecting as much data as possible. In this regard, the orchestrator starts by creating a product and then map it to the pallet. Afterwards, the orchestrator moves the pallet to load the paper, draw the 3 main part and then unload the paper. This process is repeated to finish all waiting products. For the purpose of collecting data, maximum 4 pallets where introduced at the same time during the long production process to avoid any misbehaviour or collisions.

Ethics Statement

The authors declare that the work does not involve the use of human subject or animal experiments which need for any informed consent.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgement

The research work by Wael M. Mohammed is supported by the Doctoral School of the President call 2018 of the Tampere University of Technology.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.106912.

References

[1] W.M. Mohammed, A. Lobov, B.R. Ferrer, S. Iarovyi, J.L.M. Lastra, A web-based simulator for a discrete manufacturing system, in: Proceedings of the 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON), 2016, pp. 6583–6589, doi:10.1109/IECON.2016.7793563.
[2] W.M. Mohammed, FASTory Digital Twin Dataset, (2021) 0. doi:10.23729/7E299722-C246-4695-990F-DBE5A7839AD2.
[3] W.M. Mohammed, Waeltut/Digitaltwin 1.0.1, Zenodo, 2021. doi:10.5281/ZENODO.4541008.
[4] W.M. Mohammed, B.R. Ferrer, S. Iarovyi, E. Negri, L. Fumagalli, A. Lobov, J.L.M. Lastra, Generic platform for manufacturing execution system functions in knowledge-driven manufacturing systems, Int. J. Comput. Integr. Manuf. 31 (2018) 262–274, doi:10.1080/0951192X.2017.1407874.
[5] R.F. Borja, M. Wael M., A. Mussawar, I. Sergii, Z. Jiayi, H. Robert, M.L. Jose L., Comparing ontologies and databases: a critical review of lifecycle engineering models in manufacturing, Knowl. Inf. Syst. (2021).
[6] Embedded Systems Service-based Control For Open manufacturing and Process automation | E-SCOP Project | FP7 | CORDIS | European Commission, (n.d.). https://cordis.europa.eu/project/id/332946. Accessed January 12, 2021.
[7] S.O. Afolaranmi, B.R. Ferrer, W.M. Mohammed, J.L.M. Lastra, M. Ahmad, R. Harrison, Providing an access control layer to web-based applications for the industrial domain, in: Proceedings of the IEEE 15th International Conference on Industrial Informatics (INDIN, 2017, pp. 1096–1102, doi:10.1109/INDIN.2017.8104926.
[8] S. Iarovyi, W.M. Mohammed, A. Lobov, B.R. Ferrer, J.L.M. Lastra, Cyber-physical systems for open-knowledge-driven manufacturing execution systems, in: Proceedings of the IEEE, 104, 2016, pp. 1142–1154, doi:10.1109/JPROC.2015.2509498.