CRANFIELD UNIVERSITY

ROSARIO DAVIDE D'AMICO

COGNITIVE DIGITAL TWINS: ENABLING HOLISTIC ASSET MANAGEMENT THROUGH SEMANTIC INTEROPERABILITY AND HUMAN EXPERTISE INTEGRATION

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING
Centre for Digital Engineering and Manufacturing

PhD
Academic Year: 2020–2024

Supervisors: Prof. John A. Eroyuncu, Dr. Pavan Addepalli
March 2024

CRANFIELD UNIVERSITY

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING
Centre for Digital Engineering and Manufacturing

PhD

Academic Year: 2020–2024

ROSARIO DAVIDE D'AMICO

Cognitive Digital Twins: enabling holistic Asset Management through Semantic Interoperability and Human Expertise Integration

Supervisors: Prof. John A. Eroyuncu, Dr. Pavan Addepalli
March 2024

© Cranfield University 2024. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

Academic integrity declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

The growing industrial interest in Digital Twins (DTs) has led to an expansive network of standalone DT solutions. Many of these twins operate in isolation, frequently lacking the features that enable seamless integration and communication with other systems. Based on these facts, this PhD project aims to propose a novel framework for creating DTs with advanced semantic capabilities, which are increasingly being referred to as Cognitive Digital Twins (CDTs). Beyond advancements of traditional DTs, these CDTs provide a means to formalise and transfer human expertise, which is crucial for enhancing DTs' decision support, operational efficiency, and adaptability in complex asset management. Moreover, CDTs also have the potential to achieve significantly improved semantic interoperability among twins. This enhanced interoperability facilitates a consistent and efficient data flow and augments asset management procedures. Furthermore, it boosts awareness and insights regarding the operations and performance metrics of the mirrored assets.

This comprehensive approach to DT development offers a reference frame that delineates how assets communicate, function, and evolve in a digitally twinned environment. This PhD thesis explores the potential of CDTs in improving the management of complex engineering assets by enhancing the transfer of human expertise and semantic interoperability within and among DTs.

The thesis makes several contributions, starting with a systematic literature review to identify gaps in existing academic research and an industrial review that assesses current industry practices, highlighting the relevance of the research. A novel five-step methodology for developing CDTs is introduced, emphasising the use of top-level ontologies (TLOs) and the integration of human expertise. This methodology ensures semantic consistency within the DT framework and has been validated in a laboratory setting. A CDT framework is also presented,

prioritising flexibility, interoperability, and reusability aspects provided by this novel approach.

The study primarily focuses on the manufacturing industry and underscores the role of ontologies in enhancing CDTs, bridging the gap between human understanding and automated systems. Future research directions are proposed, mainly aiming to optimise asset management, availability, and sustainability by integrating human knowledge into digital systems and promoting collaborative learning among CDTs.

Keywords

Asset Management; Basic Formal Ontology (BFO); Cognitive Digital Twins (CDTs); Federation of Twins; Human Expertise Formalisation; Ontology; Semantic Interoperability; System of System; Top-Level Ontologies (TLOs).

Contents

Academic integrity declaration **v**

Abstract **vii**

Contents **ix**

List of Figures **xi**

List of Tables **xiii**

List of Publications **xv**

Acknowledgements **xvii**

1 Introduction **1**
1.1 Background and context 1
1.2 Problem statement 5
1.3 Structure of the thesis 6

2 Research aim and objectives **9**
2.1 Research aim and objectives 9
2.2 Research process formation 10
2.3 Processes to achieve the stated objectives 12
2.4 Thesis plan and alignment with objectives 15
2.5 Support for the reader 17

3 Systematic literature review **19**
3.1 Introduction 20
3.2 Literature review 22
3.3 Methodology 24
3.4 Results 32
3.5 Characterisation of a DT for maintenance and health management 48
3.6 Discussion and conclusion 53

4 Industrial review **57**
4.1 Introduction 58
4.2 Background and literature review 59
4.3 Online survey 63
4.4 Interviews 69
4.5 Discussion 80

4.6 Conclusion and future work 86

5 Methodology **91**
5.1 Introduction 92
5.2 Literature review 95
5.3 Methodology 100
5.4 The use case: Festo sorting station 114
5.5 Discussion 132
5.6 Conclusion 136

6 The CDT concept and its role in interoperability **139**
6.1 The CDT framework 139
6.2 The connected DTs using the CDT concept 143
6.3 The TLO approach 144
6.4 Flexibility 144
6.5 Interoperability 146
6.6 Reusability 147
6.7 Interoperability index 148

7 Discussion **151**
7.1 Critical discussion of the results 151
7.2 Contribution to the knowledge 154
7.3 Limitations 156
7.4 The research question 158

8 Thesis conclusions and future work **161**
8.1 Conclusions 161
8.2 Future works 163

A Appendix A Ethical approval letter **167**

References **169**

List of Figures

1.1 Structure of the central part of the thesis with relative publications 7

2.1 The five steps of the methodology - presented in Chapter 5 13
2.2 Thesis breakdown between motivation, chapters, and objectives 16

3.1 Number of publications containing ”digital twin” in title/abstract/keywords on the Scopus database from 2015 to 2021 21
3.2 Distribution of literature review publications about DTs (Tool: sankeymatic.com) 23
3.3 SALSA methodology steps description. Inspired by [Booth et al., 2012]. 24
3.4 Classification of the category's division for the review of DTs 29
3.5 Distribution of the reviewed publications per year 32
3.6 Distribution of the reviewed publications between journals and conferences . . 33
3.7 Sankey diagram with the distribution of published papers divided per year and type of publication 33
3.8 List of authors having two or more articles within the final selection 34
3.9 List of keywords used at least 3 times in the reviewed publications 34
3.10 Distribution per IC of the reviewed articles 36
3.11 Distribution of the reviewed papers per field of application 37
3.12 Stages of the life cycle where reviewed papers had more focus 40
3.13 Example of the machine to machine (M2M) communication with the OSI model. The left side shows how a machine receives information from another machine (receiver) and, on the right side, how a machine sends information to another machine (sender) 41
3.14 Comparison between OSI 7-layer model and TCP/IP 4-layer model 42
3.15 Distribution of DT types through the life cycle and information flow [from [Grieves, 2012]] 49
3.16 Physical asset life cycle and information life cycle 50
3.17 DT conceptual model characterisation 52

4.1 Distribution of job roles of respondents 64
4.2 Challenges to address with DT adoption, where (MA) represents the category Managing Assets, (A) represents Assurance, and (S&P) represents Strategy and Planning 64
4.3 Sectoral representation of respondents. 66
4.4 Interest in DT implementation across the life cycle. 66
4.5 The size of the DT projects. 66
4.6 The focus area of DT. 67
4.7 Purpose area of DT. 67
4.8 Status of the DT project. 68

4.9 Level of maturity of DT projects. 68
4.10 BEP (in years) of DT projects, divided between completed projects (left), large projects (centre), and small projects (right). In the abscissa the number of the project, while in the ordinate the time length in years. 69
4.11 Barriers and how to overcome them. 83
4.12 Roadmap for wider adoption of DTs in the manufacturing industry. 85

5.1 Number of publications regarding CDTs. 96
5.2 Methodology to develop CDT. 101
5.3 Mereotopology diagram that shows the relations between artefacts, their functions, and their capabilities. 104
5.4 Levels of abstraction in ontology theory. 109
5.5 Database approaches based on data complexity. 110
5.6 Relation between the components of the CDT. 113
5.7 Festo Sorting Station, schema, components, actions. 114
5.8 Mereotopological representation of the system along with functions. 116
5.9 Modulo junction diagram from AD theory. 117
5.10 Ontology classification based on the level of abstractions. 123
5.11 Screenshot of the Protégé tool focusing on the class Sensing Station Festo. . . . 123
5.12 Diagram that represents the process ActOfCarrying using BFO properties. . . . 124
5.13 Diagram that represents the process ActOfMeasuringColor using BFO and CCO properties. 124
5.14 Model representation of the ontology. 126
5.15 Data flow in the CDT. 128
5.16 Snapshot of the KG populated. Part of the ontology (left), and data population containing data related to workpiece ID001 in space $s_in(right)$. 128
5.17 Example of SPARQL query to send data point to the KG. 129
5.18 Malfunctioning of the red sorting unit. 131
5.19 Schema for the reification. 131
5.20 State description of workpiece ID 004. 132
5.21 Malfunctioning for the Sensing Unit. 132
5.22 Malfunctioning for the Sensing Unit. 133

6.1 Cognitive Digital Twin framework based on a Top-Level Ontology. 140
6.2 The 5D DT model. 141
6.3 The graphical difference between model, shadow, and twin according to the basic model presented in this paper. 142
6.4 Example of representation of Connected DTs. 143
6.5 Representation of ontology dependence with two levels of abstractions. 149

List of Tables

2.1 Thesis plan, alignment with objectives, and status of paper submissions 15

3.1 Research scope definition - PICOC framework 25
3.2 Inclusion and exclusion criteria 26
3.3 Screening criteria 27
3.4 Records for the synthesis analysis 28
3.5 List of journals 35
3.6 List of conference names 38
3.7 List of definitions of DT found in the reviewed articles in chronological order . 39
3.8 Communication protocols in the network architecture for the DT 44
3.9 Descriptions and references for various data/file formats. 46
3.10 List of tools employed for modelling and simulation, visualisation, and virtualisation. 47

4.1 Categorisation of challenges derived from the answer to the online survey. Where (MA) represents the category Managing Assets, (A) represents Assurance, and (S&P) represents Strategy and Planning. 65
4.2 Steps of the thematic analysis. 70
4.3 Profiles of interview meetings (online and 60 mins duration) and participants' details. 71
4.4 Themes resulted from the thematic analysis. 72
4.5 Table of DT drivers. 74
4.6 Table of inhibitors. 74
4.7 Table of DT attributes. 76
4.8 Table of DT design. 76
4.9 Table of DT visualisation. 77
4.10 Table of DT semantics. 77
4.11 Results from the return-on-investment questions. 78
4.12 Next steps and needs. 80

5.1 Definitions of Cognitive Digital Twins in the literature. 97
5.2 Colour detection from sensors' output combination. (DS = Diffuse sensor, DCS = Diffuse colour sensor, IPS = Inductive proximity sensor) 127

List of Publications

Journal papers

- **Rosario Davide D'Amico**, John Ahmet Erkoyuncu, Sri Addepalli, Steve Penver; *Cognitive digital twin: An approach to improve the maintenance management*; CIRP Journal of Manufacturing Science and Technology; Volume 38, 2022; Pages 613-630; https://doi.org/10.1016/j.cirpj.2022.06.004

- **Rosario Davide D'Amico**, Sri Addepalli, and John Ahmet Erkoyuncu. 2023. "*Industrial Insights on Digital Twins in Manufacturing: Application Landscape, Current Practices, and Future Needs*" Big Data and Cognitive Computing 7(3): https://doi.org/10.3390/bdcc7030126

- **Rosario Davide D'Amico**, Arkopaul Sarkar, Mohamed Hedi Karray, Sri Addepalli, John Ahmet Erkoyuncu; *Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins*; CIRP Journal of Manufacturing Science and Technology; Volume 52; 2024; Pages 366-385; https://doi.org/10.1016/j.cirpj.2024.06.007

Conference papers

- **Davide D'Amico**, John Ekoyuncu, Sri Addepalli, Christopher Smith, Ed Keedwell, Jim Sibson, Steven Penver; *Conceptual framework of a digital twin to evaluate the degradation status of complex engineering systems*; Procedia CIRP; Volume 86; 2019; Pages 61-67; https://doi.org/10.1016/j.procir.2020.01.043

- **Rosario Davide D'Amico**, Pavan Addepalli, and John Ahmet Erkoyuncu, *Is a Top Level Ontology Based Digital Twin the Solution to Human-Machine Interoperability?* (October 18, 2021). Proceedings of the The 10th International Conference on Through-Life Engineering Services 2021 (TESConf 2021), http://dx.doi.org/10.2139/ssrn.3945058

- **Rosario Davide D'Amico**, Arkopaul Sarkar, Hedi Karray, Sri Addepalli, John Ahmet Erkoyuncu; *Detecting failure of a material handling system through a cognitive twin*; IFAC-PapersOnLine; Volume 55, Issue 10; 2022; Pages 2725-2730; https://doi.org/10.1016/j.ifacol.2022.10.128

Book chapters

- Farsi, Maryam and Namoano, Bernadin and Latsou, Christina and Subhadu, Vaishnav Venkata and Deng, Haoxuan and Sun, Zhen and Zheng, Bohao and **D'Amico, Davide** and Erkoyuncu, John Ahmet"; *Digital Twin Architecture for a Sustainable Control System in Aircraft Engines*; Sustainable Materials and Manufacturing Techniques in Aviation; 2024; Springer Nature Switzerland; https://doi.org/10.1007/978-3-031-62987-7$_5$

Non-peer-reviewed articles

- **Davide D'Amico**, Usama Shazlee, Damien Mazeas, Sri Addepalli, Samuel Court, and John Ahmet Erkoyuncu; *Emerging technologies to support asset management*; June 16, 2021; Maintenance and Engineering, Available at: https://www.maintenanceandengineering.com/2021/06/16/emerging-technologies-to-support-asset-management/

Acknowledgements

Embarking on this doctoral journey, I was surrounded by an incredible group of mentors, supporters, and friends, each playing a pivotal role in shaping this journey and completing my thesis.

Foremost, I express my profound gratitude to Prof. John Ahmet Erkoyuncu, whose exceptional mentorship and boundless encouragement propelled me to push my limits academically, professionally, and personally. His guidance opened doors to invaluable opportunities, enriching my PhD experience beyond measure. Alongside, my sincere thanks go to Dr. Pavan Addepalli, whose diligent support and unwavering dedication significantly elevated the quality of my research.

I extend my sincere appreciation to Prof. Hedi Karray and Dr. Arkopaul Sarkar of ENIT, France, for their warm hospitality and invaluable guidance. Their deep expertise in Ontologies has not only enriched my understanding but also connected me with a network of esteemed professionals, for which I am deeply grateful.

My research has thrived thanks to the generous support of Babcock International. Special acknowledgement to Andy Cooper, Jim Sibson, and Steve Penver, whose relentless motivation and support have inspired my drive for innovation and significantly broadened the impact of my work. The supportive environment of the Centre for Digital Engineering and Manufacturing at Cranfield University, UK, and the Laboratoire Génie de Production at ENIT, France, provided a foundation of support and academic team spirit. The students and researchers I met during my tenure have indelibly marked my experience, leaving lasting memories.

Beyond academics, my deepest gratitude extends to my family and friends, the bedrock of my resilience and success. To my parents, my sister, and her family, your unwavering love and faith have been my guiding lights. To my girlfriend, your love and patience have made my journey worthwhile. To my 'road to happiness' friends and my bubble in Bedford, you have infused my journey with strength and joy.

Chapter 1

Introduction

1.1 Background and context

1.1.1 Emergence of Digital Twins

The Digital Twin (DT) is a virtual, digital equivalent of a physical product. The main idea of a DT is to shift the workload from the physical world to the digital one [Grieves, 2015]. The rise of DTs in engineering asset management can be attributed to their ability to create dynamic digital models that learn and update the status of physical assets [Lu et al., 2011]. Moreover, A DT is made possible by data and simulators to facilitate better decision-making, monitoring, controlling, and real-time prediction, optimisation, and monitoring [Rasheed et al., 2020]. These DTs are the basis of complex software ecosystems, integrating information and co-creating value [Camposano et al., 2021]. However, their widespread use is hindered by the lack of standardisation [Macchi et al., 2018]. Despite this, DTs are increasingly recognised for their potential in the life cycle and prognostic decision-making for engineered assets [Krishnamenon et al., 2021].

The rise of DTs in engineering asset management can be attributed to their potential to reform decision-making processes and improve asset performance. DTs can integrate real-time data from sensors, simulations, and historical information to provide a comprehensive understanding of the asset's behaviour and condition [Krishnamenon et al., 2021].

One of the primary drivers of adopting DTs in engineering asset management is the need for an improved decision-making process. DTs enable asset managers to make informed decisions throughout the asset's life cycle, from design and construction to operation and maintenance [Krishnamenon et al., 2021]. By simulating different scenarios and analysing real-time data, DTs provide valuable insights into asset performance, maintenance needs, and potential risks, allowing for proactive decision-making and optimisation of asset management strategies [Krishnamenon et al., 2021, Madni et al., 2019].

Furthermore, DTs offer the potential for predictive maintenance, which can significantly

reduce downtime and maintenance costs. By continuously monitoring asset performance and analysing data, DTs can detect early signs of asset degradation or failure, enabling timely maintenance interventions [Krishnamenon et al., 2021, Madni et al., 2019]. This shift from schedule-based to condition-based maintenance can optimise asset performance and extend asset lifespan [Madni et al., 2019].

DTs also facilitate the integration of physical and digital systems, enabling seamless communication and collaboration between stakeholders in asset management. By providing a shared platform for data exchange and decision-making, DTs enhance coordination and cooperation among asset owners, operators, maintenance teams, and other stakeholders [Camposano et al., 2021, Madni et al., 2019].

Technological advancements, such as the Internet of Things (IoT), big data analytics, and cloud computing, have played a crucial role in developing and implementing DTs. These technologies provide the necessary infrastructure for collecting, storing, and analysing vast amounts of data from assets, making it possible to create accurate and dynamic digital replicas [Krishnamenon et al., 2021, Madni et al., 2019].

1.1.2 The semantic gap

Data management presents intricate and multifaceted challenges across technological, conceptual, and managerial dimensions [Varga, 2010]. These challenges evolve and become more complex in response to the dynamic nature of data itself. The exponential increase in the volume of data, coupled with its growing diversity and inherent unpredictability, adds layers of complexity to data management practices. Given the interconnectedness of digital ecosystems, the need for systems and platforms to effectively communicate and exchange data has never been more critical. This requires the development of robust standards and protocols that can accommodate the diversity of data formats and structures [Gleim and Decker, 2020].

Furthermore, a significant semantic gap exists between the raw data collected and stored within these technological systems and the human expertise required to interpret and utilise this data. This gap represents a fundamental challenge in aligning vast quantities of data with meaningful human insight, thereby impacting the efficiency of information retrieval, the fluidity of knowledge sharing, and the precision of decision-making processes. Bridging this gap is crucial for ensuring that data serves its intended purpose in enhancing understanding and guiding informed actions. In this context, the application of ontologies emerges as a promising approach. Employing ontologies makes it possible to formally and explicitly represent the intricate web of concepts, relationships, and constraints that define a particular domain, thereby facilitating a more intuitive and effective interface between human cognitive capabilities and the digital data landscape. This methodology not only aids in narrowing the semantic gap but also enhances data systems' overall coherence and usability, making them more aligned with human reasoning and knowledge structures.

An ontology, in the context of computer science, refers to an explicit formal model representing a shared conceptualisation of a domain [Gruber, 1993]. It specifies the concepts in a domain and their relationships, providing a standardised vocabulary for knowledge representation and facilitating interoperability between different systems and applications [Guarino and Welty, 2009].

Interoperability in digital systems refers to the ability of diverse systems and organisations to work together [Gottschalk, 2009]. It involves the exchange of data and information across systems, applications, or components, and is crucial for digital government and digital innovation [Hodapp and Hanelt, 2022]. Interoperability can lead to benefits such as innovation, competition, and choice, but it also poses risks such as security and privacy concerns [Gasser, 2015].

Therefore, using an ontology can reduce the semantic gap by providing a common understanding and a shared vocabulary for humans and machines. Ontologies enable integrating and organising diverse data sources, allowing for more effective information retrieval and knowledge discovery. They also facilitate the development of intelligent systems that can reason and make inferences based on the formalised knowledge represented in the ontology [Guarino and Welty, 2009]. Moreover, ontologies can serve as a basis for capturing and representing human expertise and domain knowledge. They provide a structured framework for explicitly specifying the concepts, relationships, and constraints relevant to a particular domain. This explicit representation enables the encoding of expert knowledge and facilitates the sharing and reusing knowledge within and across different domains [Sicilia, 2005].

Human expertise can be mapped into an ontology through knowledge elicitation. This process involves capturing and representing the knowledge and expertise of individuals in a structured and formalised manner. Here are a few approaches and citations that discuss mapping human expertise into a semantic model or ontology:

1. Ontology-Driven Human Language Technology for Semantic-Based Business Intelligence [Declerck et al., 2008]: This article describes the development of textual analysis and ontology-based information extraction in the European R&D project ”MUSING”, focused on Business Intelligence. It discusses how semantic-based knowledge and content systems can be integrated into Business Intelligence tools and modules.

2. Ontology-Based Competency Management [Sicilia, 2005]: This article discusses using ontologies as the enabling semantic infrastructure of competency management in knowledge-intensive organisations. It explores how ontologies can be used to represent and manage human expertise and competencies within an organisation.

3. An Ontology-Based Modeling Approach for Developing a Competencies-Oriented Collective Intelligence [Brut et al., 2008]: This article presents a modelling approach that uses ontology to develop well-managed and semantically enriched user profiles and doc-

ument annotations. It discusses how ontologies can be used to represent and capture user competencies, facilitating personalised and collaborative facilities in an e-learning platform.

4. Interactive Ontology-Based User Knowledge Acquisition [Aroyo et al., 2006]: This research paper presents an approach to eliciting a user's conceptualisation through an ontology-driven dialogue. It discusses how ontologies can be used to capture and represent user knowledge, facilitating personalised knowledge acquisition and recommendation systems.

Please note that while these articles discuss using ontologies and semantic models for mapping human expertise, they may not provide specific details on the mapping process itself. There is limited availability for research regarding mapping human expertise into a semantic model or ontology, and it can vary depending on the application's specific domain, context, and requirements.

1.1.3 Importance of bridging the gap

Ontologies and semantic technologies are employed in asset management throughout the life cycle to enhance decision-making, optimise asset performance, and facilitate system interoperability. Several works discuss the utilisation of ontologies and semantic technologies in asset management.

Kiritsis aimed to propose a semantic data model for engineering asset management, specifically focusing on the operation and maintenance phase of its life cycle. The study suggested using ontologies due to their ability to capture data semantics, establish a common vocabulary for sharing knowledge within the domain, and provide reasoning capabilities [Kiritsis, 2013].

Another research effort noted that while semantic technologies and ontologies are increasingly popular in engineering applications, their adoption in industry remains limited, and applications tend to be of small scale. The research sought to broaden the understanding of the advantages of applying such technologies in the complex context of asset life cycle management and to provide a platform for implementing ontology models in industrial settings [Koukias et al., 2013].

Additionally, it was observed that ontologies define a shared vocabulary for researchers and practitioners involved in information exchange within a specific domain. Given the growing volume of information published in the asset management domain, the importance of knowledge-based systems and maintaining consistent representation and terminology for such information has grown. This underscores the rationale for developing an asset management ontology [Frolov et al., 2009].

Furthermore, the developed ontology in one of the studies has multiple potential applications, including sharing and annotating asset management information, identifying gaps

in prevailing asset management paradigms, visualising the holistic nature of asset management, classifying asset management knowledge, and constructing a relational asset management knowledge-based system [Frolov et al., 2009].

1.2 Problem statement

1.2.1 Challenges in current approaches

The DT concept is becoming crucial in various sectors, from improving manufacturing processes to enhancing the functionality of cyber-physical systems. This innovative approach allows for real-time analysis and optimisation of complex engineering systems [Bécue et al., 2020, Yun et al., 2017]. Despite its potential, the adoption of DTs faces obstacles. A key issue is the lack of a clear, universally accepted definition of a DT, which leads to inconsistencies in its development and use. Moreover, there is a pressing need for standardised methods in creating and managing DTs to avoid inefficiencies and fragmentation in their implementation [Dahmen and Rossmann, 2021].

Furthermore, a federated twin model has been proposed in the literature. This model involves a network of DTs that can operate both independently and in collaboration, offering a unified and more accurate virtual representation of their physical counterparts. Such a system promises to enhance DTs' cooperation and data sharing [Haag and Anderl, 2018, Stark et al., 2019]. However, interoperability remains a significant hurdle, with varied systems and platforms often employing incompatible formats that obstruct data exchange between DT systems. This challenge is compounded by the prevalent siloed approach to DT development, where twins are created for specific, isolated tasks. This isolation limits the scope of each DT's functionality and prevents the holistic benefits that could be achieved through integrated, cross-domain applications. Breaking down these silos is crucial for leveraging the full spectrum of advantages offered by DTs, fostering an environment where insights, knowledge, and resources can be shared and utilised more effectively across different domains [Fuller et al., 2020, Khan et al., 2022].

The federated twin model suggests a way forward, advocating for standardised communication protocols to enable DTs to share data and work together more effectively. This approach could transform how organisations use DTs, allowing for more informed decision-making by leveraging collective insights from multiple DTs [D'Amico et al., 2023].

Nevertheless, the transition to a federated model is not without its challenges. The development of universal standards for communication and data exchange is a complex task, and ensuring the security and privacy of interconnected systems is paramount. Furthermore, the governance and coordination of such a federated network demands meticulous planning and collaboration [Khan et al., 2022].

1.2.2 Research focus

In light of the challenges posed by interoperability issues and the siloed nature of existing DT implementations, there is a growing recognition of the need for innovative approaches to enhance DTs' capabilities and collaboration potential. Within this context, the concept of Cognitive DTs emerges as a promising solution. The Cognitive DT (CDT) concept is a promising evolution of the current DT concept towards a more intelligent, comprehensive, and full life cycle representation of complex systems [Zheng et al., 2022a]. The CDT represents a significant advancement in the realm of DTs. It builds upon the foundations of traditional DTs. It introduces augmented semantic capabilities, making it capable of not only replicating physical assets but also incorporating and understanding human expertise within its framework [Lu et al., 2020a].

The concept of the CDT is further analysed throughout this thesis. The difference between the classical DT approach and the reasons why the CDT satisfies the needs highlighted by industries are part of the outcomes of this PhD work. The primary focus of this study deals with the capabilities of the CDTs in the context of complex engineering asset management to address the challenges highlighted above. In particular, the aim is to create a CDT framework that turns low-semantics data into semantically enriched data, supporting the management of complex engineering assets and facilitating the transferability of human expertise. By addressing this, the aim is to bridge the gap between human expertise and DT systems, ultimately advancing the field of asset management and DTs.

The subsequent chapters will delve into the complexity of a CDT and its potential to support asset management by enhancing semantic interoperability and incorporating human expertise.

1.3 Structure of the thesis

This thesis is presented in a "paper-format" style, wherein the core chapters are crafted in the guise of journal articles. This approach highlights the intellectual contributions of the work through distinct chapters, each offering a comprehensive package of work that includes an introduction, methodology, results, and discussion, thus ensuring each chapter is a self-contained account of the work undertaken. A systematic literature review and a paper delineating industry best practices and future directions, based on surveys and interviews, have been published in peer-reviewed journals. Furthermore, a methodology paper has been also published in a peer-reviewed journal. To ensure uniformity, these papers have been adapted to the thesis format. This format offers a multitude of advantages, particularly in fostering the ability to produce succinct, self-contained reports that effectively communicate research findings despite potential repetitions. Repetitions can be found, especially where the results of one paper inform another or where background information overlaps across papers. Adopting this format style facilitates a structured and coherent presentation of the research and distinctly highlights the

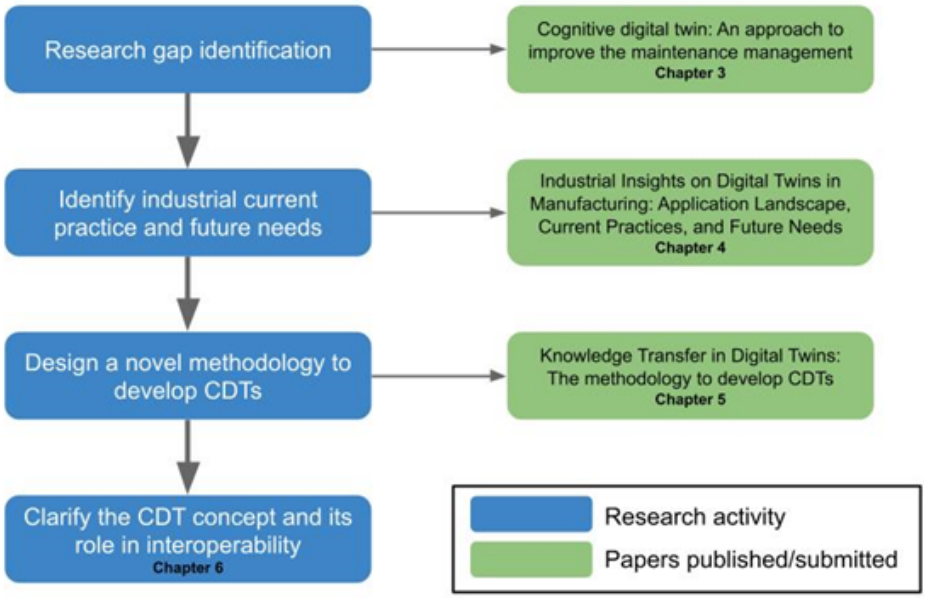


Figure 1.1: Structure of the central part of the thesis with relative publications

various contributions to knowledge made throughout the PhD. This format effectively showcases the incremental advancements and the breadth of investigation undertaken, providing a clear and comprehensive narrative of the scholarly contributions within the field.

The structure of this PhD thesis provides a comprehensive exploration of the research topic. It begins with Chapter 2, where the research aim and objectives are defined, setting the stage for the study. This section includes the processes used in this PhD research, including its formation, context, and data collection techniques, all of which serve as the methodological underpinning for the subsequent chapters.

Chapters 3 to 5 constitute the core of the thesis, delving into the intricacies of the research and their relations with the research activity, which are reported in Figure 1.1. Chapter 3 introduces the concept of CDTs and conducts a thorough systematic literature review, laying the groundwork for the methodology detailed in Chapter 5. It presents results from the literature review, offering valuable insights into publication elements, study elements, characteristics, and conclusion elements. Chapter 4 expands the scope of the research by investigating industrial insights through online surveys and interviews, shedding light on current practices and future needs in the manufacturing sector.

Chapter 5 presents the methodology for developing CDTs, representing a pivotal contribution to the thesis. Chapter 6 builds upon this foundation, elucidating the CDT concept, framework, and its significance in achieving interoperability among DTs. The TLO approach, with its emphasis on flexibility, interoperability, and reusability, is discussed.

Chapter 7 engages in a critical discussion, contextualising the findings within established theories and emphasising their relevance. The contributions to DTs and maintenance management are highlighted, along with an acknowledgement of the study's limitations.

Finally, in Chapter 8, the thesis summarises the key findings and a perspective on future research directions. Overall, this PhD thesis significantly contributes to understanding CDTs and their application in complex engineering asset management.

Chapter 2

Research aim and objectives

This chapter outlines the aim, objectives, and research question of this study. It also elucidates the rationale behind the selection of a suitable research methodology. Section 2.1 introduces the research aim and the corresponding objectives. Section 2.2 delves into the current research techniques, available research strategies and data collection methods. Additionally, this section discusses the approaches employed to ensure the generalisability and validity of the research findings. Section 2.3 provides a high-level overview of the methodology adopted in this study.

2.1 Research aim and objectives

This research aims *to develop a cognitive digital twin (CDT) framework to facilitate the transferability of human expertise and support the management of complex engineering assets, converting data low on semantics to semantically enriched data. This paradigm breaks up the siloed approach, facilitating interoperability and access to data throughout the life cycle.*

The literature review throughout the research set the basis for this study by capturing the current trends and research gaps. Furthermore, the findings and challenges observed by the industry also supported the determination of the scope of this research. Five objectives have been defined in order to accomplish the aim. The research objectives are:

1. Define the context and elements of the CDT framework. It involves understanding the current limitations and challenges of the traditional DT framework. Conceptualise the fundamental aspects of a CDT that can effectively integrate human expertise.

2. Develop methods for enhancing the metadata - the data about data - ensuring the CDT can effectively understand and process the data it receives.

3. Design interoperability protocols. Break down the siloed approach by developing standardised methodologies that facilitate seamless data exchange and communication between different systems. Investigate how those standardised methodologies can be incorporated into the CDT framework.

4. Outline expertise transfer mechanism. Understand the different modalities of human expertise in complex engineering assets management. Incorporate that expertise within the CDT framework.

5. Verify and validate the CDT framework. Implement a prototype based on the designed CDT framework to verify and validate it in a laboratory setup.

The main research question this PhD aims to answer is the following:

Can a CDT approach facilitate the transferability of human expertise and the associated semantic interoperability between digital twins, improving the management of complex engineering assets?

In section 2.2, the author reviews the research process that has guided the development of this study's research methodology.

2.2 Research process formation

This section presents the factors considered in the design of a comprehensive research approach for this study. The research context is described, providing a foundation upon which the research strategy has been formulated. Following this, a discussion on the motivation behind the research and an overview of the data collection techniques employed are presented.

2.2.1 Research context

This study delves into the concept of transferring human expertise within the realm of DTs. Insights from a systematic literature review and recognised industrial challenges steered the research towards developing a redesigned DT framework. This framework aims to formalise and transfer human expertise from the tangible world to its digital counterpart while enhancing the interoperability of DTs, especially in scenarios where multiple DTs are integrated or federated. This approach can be conceptualised in the literature as the CDT. This is one interpretation, but CDTs can also be understood as entities with different past 'experiences' that they can record and share with other CDTs undergoing similar current or ongoing conditions. These other CDTs can then learn from past scenarios encountered by neighbouring CDTs. Given the nascent state of research on CDTs within the manufacturing domain, this study sought to bridge various research disciplines, including "digital twins", "ontology", "top-level ontology", "maintenance", and "complex engineering assets". As this domain holds significant industrial relevance, insights from industry experts regarding prevailing practices, challenges, and future aspirations were important in painting a holistic picture and addressing core challenges.

2.2.2 Data collection techniques

The selection of data collection techniques depends on the type of information required, the source of that information, and the circumstances under which it is collected. Driven by the nature of this study, the following data collection techniques have been utilised.

Literature review

The literature review offers insights into existing research that spans current practices and potential future directions. In some instances, analysing literature can provide as much intellectual and practical value as collecting first-hand data. There are several reasons to conduct a literature review, including [Hart, 2018]:

- Identifying completed work
- Preventing duplication of efforts
- Avoiding pitfalls and mistakes from prior research
- Assisting in designing the research
- Identifying gaps in existing research

Prominent types of literature encompass books, articles, reports, conference papers, official documents, and reviews.

Survey

Surveys are widely used across research domains because they can rapidly, inexpensively, and accurately gather information on a subject [Zikmund, 2013]. The merits of the survey method inspired the author to adopt it as a primary data collection technique. It allows for multiple questions about a topic, providing flexibility in analysis. Furthermore, there is versatility in how questions can be presented, be it orally, in writing, or electronically. Standardising questions also ensures more precise result comparisons.

Interviews

Interviewing typically involves direct interaction, either in person or through videoconferencing tools like Zoom or Teams, where the researcher poses questions and records responses. Interviews are classified into three types: structured, semi-structured, and unstructured [Fellows and Liu, 2021]. Semi-structured interviews, used in this research, mix standardised questions with those tailored to the individual. This method's strength lies in its flexibility in discussing a broad range of topics.

Documents

Document analysis encompasses written materials, such as newspapers, and letters, and non-written content, like videos or seminars. During this research, the author received several documents detailing current processes.

2.3 Processes to achieve the stated objectives

The ontological approach is pivotal in developing the CDT framework. At a basic level, an ontology is a structured method to represent and categorise knowledge, drawing upon clearly defined terms, relational expressions, and natural language definitions. Further solidifying these definitions is a formal theory ensuring they align with their intended interpretations, as stated in ISO-IEC 21838-1, 2021. This formal semantics approach also enhances the quality of the system's metadata, ensuring that the CDT can effectively understand and process the data it receives by establishing a formalised lingua franca.

The concept of TLO - defined as an ontology created to represent the entities shared across a maximally broad range of domains (ISO-IEC 21838-1, 2021) — holds immense promise for this research. TLO provides a domain-neutral framework that can be employed across various domains. In the context of the CDT, the TLO approach is crucial for translating human expertise (such as that of a maintenance operator) into the DT. Moreover, it enhances semantic interoperability in scenarios where multiple systems are interacting with each other, commonly referred to as a system of systems environment. This ensures that the CDT can accurately interpret and integrate data from diverse sources.

Expanding on this, using a domain-neutral TLO as a foundation allows for the development of domain-specific ontologies that align with the upper-level ontology. This alignment ensures a high level of consistency and compatibility. For this study, ISO/IEC 21838-2, also known as the Basic Formal Ontology (BFO), was identified as the TLO to anchor the creation of the CDT. By employing this formal semantics approach, the quality of the metadata is improved, enabling the CDT to better understand, process, and leverage the data it receives for more accurate and efficient operations. Additionally, this formalisation ensures that the data is Findable, Accessible, Interoperable, and Reusable (FAIR), thereby maximising its utility and value across various applications [Wilkinson et al., 2016].

The methodology to implement this approach is defined in Chapter 5 and comprises five steps, reported in Figure 2.1. These steps converge into a systematic roadmap. This roadmap is a path that not only leads to the creation of a CDT but also ensures its efficiency, placing a pronounced emphasis on integrating knowledge, sharpening each ontology, and optimising data management and analytics. To further refine and validate the methodology, a systematic literature review and an industrial survey were carried out. These investigations provided critical insights into the current state of the art and industrial needs, directly informing the development

Figure 2.1: The five steps of the methodology - presented in Chapter 5

of the CDT framework and ensuring that the approach aligns with both academic and practical perspectives.

The five-step methodology to develop CDTs can be summarised as follows:

1. **Step 1: Requirements Analysis**

   - **Objective:** To comprehensively understand the system of interest by dissecting its structural and functional components, defining its scope and purpose, and modelling part-whole and spatial relationships.
   - **Approach:** This step employs mereotopology theory to formalise part-whole relationships (mereology) and spatial connections (topology) within the system. Additionally, Axiomatic Design (AD) theory is used to break down the system into functional requirements, design parameters, and process variables, ensuring that the design meets its intended functions.
   - **Output:** A detailed system framework, including its components, functions, capabilities, and their interrelationships, serves as the foundation for subsequent steps, guiding knowledge gathering and ontology development.

2. **Step 2: Knowledge Gathering**

   - **Objective:** To systematically collect essential information related to the system, including facts, assumptions, rules, behavioural patterns, periodic observations, and human expertise.
   - **Approach:** The knowledge-gathering process builds upon the framework established in the requirements analysis. It involves identifying processes within the system and formalising them using First Order Logic (FOL) axioms. Thematic roles such as ”Instrument” and ”Patient” are defined to clarify the roles of different entities in these processes.
   - **Output:** A comprehensive knowledge base that captures the essential elements and processes of the system, informing the ontology development step by providing the raw data and expertise needed to formalise the system’s knowledge.

3. **Step 3: Ontology Development**

- **Objective:** To formalise the knowledge gathered in the previous steps into a structured ontology, which serves as the foundation for data management and analysis within the CDT.
- **Approach:** This step begins with the formulation of competency questions (CQs) to define the ontology's purpose. The ontology is developed using established TLOs like Basic Formal Ontology (BFO) and mid-level ontologies (MLOs) like Common Core Ontology (CCO). The terminology and structure of the ontology are influenced by the mereotopology and AD theories to ensure consistency and logical coherence.
- **Output:** A formal ontology that represents the system's knowledge in a structured and computable format, providing the schema and vocabulary used to organise and query the data within the CDT.

4. **Step 4: Data Management**

   - **Objective:** To manage the data related to the CDT, including selecting the appropriate database technology (e.g., graph databases), modelling the data using the ontology vocabulary, and integrating the data into a knowledge graph.
   - **Approach:** The data management strategy leverages the ontology developed in the previous step to structure the knowledge graph. The data is transformed and imported into the graph database, ensuring that it aligns with the formalised processes and relationships defined by the mereotopology and AD theories.
   - **Output:** A populated knowledge graph that is structured according to the ontology, enabling advanced data analytics and reasoning, directly supporting the analysis and validation phase.

5. **Step 5: Analysis and Validation**

   - **Objective:** To validate the CDT's functionality by performing analytical and simulation activities, including testing its ability to detect system malfunctions and support decision-making.
   - **Approach:** This phase involves querying the ontology with the CQs, applying rules derived from the formal ontology, and using the knowledge graph to detect malfunctions in the system. The AD theory is used to validate the system design and ensure that all functions are realised effectively. The mereotopology framework supports the analysis by providing a logical structure for understanding the relationships and interactions within the system.
   - **Output:** Insights and feedback that validate the CDT's ability to meet its intended goals. The results from this analysis inform potential refinements in the ontology

Table 2.1: Thesis plan, alignment with objectives, and status of paper submissions

Chapter	Objectives	Title	Journal	Status
3	1, 2, 4	Cognitive digital twin: An approach to improve the maintenance management	CIRP Journal of Manufacturing Science and Technology	Published 08/2022
4	1, 2, 4	Industrial Insights on Digital Twins in Manufacturing: Application Landscape, Current Practices, and Future Needs	Big Data and Cognitive Computing	Published 06/2023
5	1, 2, 3, 4, 5	Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins	CIRP Journal of Manufacturing Science and Technology	Published 06/2024
6	1, 3, 5	The CDT concept and its role in interoperability	-	In preparation

and data management strategies, ensuring that the CDT remains accurate and effective in representing and managing the system.

2.4 Thesis plan and alignment with objectives

Table 2.1 provides an overview of the following thesis chapters, their alignment with research objectives, the titles of corresponding papers, the journals in which they are published, and the current status of these papers. Figure 2.2 reports the chapters along with the main motivations and how they align with the objectives. In the next chapter, a systematic literature review on CDT is presented.

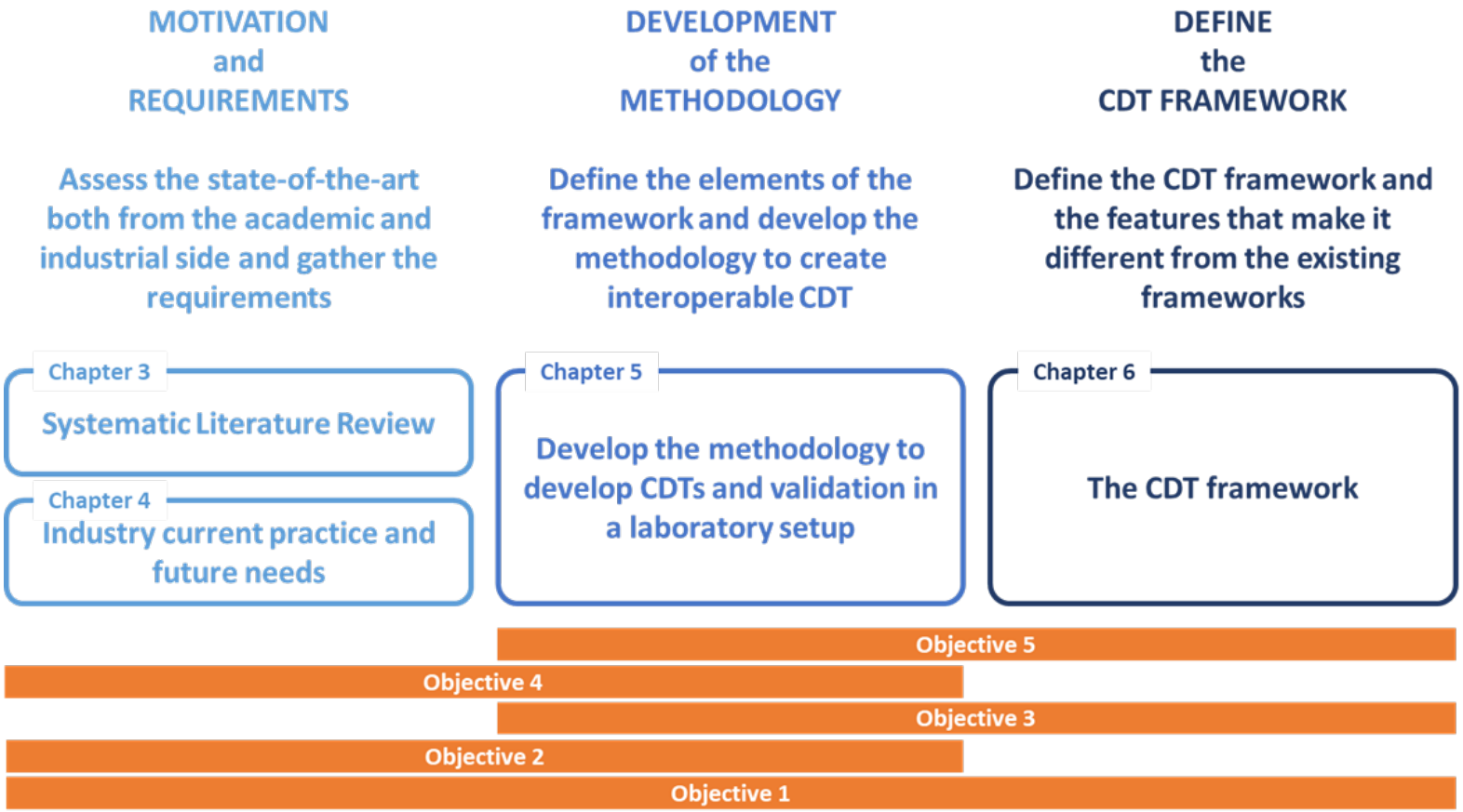


Figure 2.2: Thesis breakdown between motivation, chapters, and objectives

2.5 Support for the reader

This paragraph aims to further clarify the structure of the thesis as outlined in Section 1.3 and depicted in Figure 1.1. The first two chapters lay the groundwork by providing essential background information, including the problem statement, aims, objectives, and research questions. These foundational chapters are crucial for setting the stage for the more detailed and focused sections that follow.

The subsequent three chapters are presented as standalone papers to enhance readability, as they correspond to published articles (see Table 2.1). Each of these chapters contains its own literature review, methodology, results, discussion, and conclusion. This format ensures that each chapter can be read independently, providing a comprehensive understanding of each specific study. Additionally, this structure allows the reader to observe the chronological evolution and refinement of the CDT concept over the course of the PhD research.

Chapter 3 presents a systematic literature review conducted in April 2022. Although some numerical comparisons are made with data collected in 2023, these are not part of the original systematic methodology but are included to provide additional context and insights.

Chapters 4 and 5 delve into more recent and specialised reviews related to the thesis topic. Chapter 4 explores the industrial state of the art, identifying current requirements and future needs. This chapter underscores the motivation and relevance of the thesis, establishing a clear connection between academic research and industrial application. It is important to note that since this paper focused on DTs in general, it does not have a special emphasis on semantic technologies. Instead, it highlights other technologies identified by industry experts during the interviews, reflecting a broader perspective on the state of DTs in various sectors. Chapter 5, which forms the core of this thesis, introduces the innovative methodology for developing CDTs. This chapter is particularly important as it outlines the practical steps and novel contributions involved in CDT development, showcasing the unique approach of this research.

However, it is important to note that due to the structure of the thesis, the different chapters may not always seem to link seamlessly with each other in every aspect. This is partly due to the timeliness of the publications. For instance, the first literature review was published at the end of the first year of the PhD, during a period when the research topic was still being defined and refined. At that time, the focus on CDT was emerging but not fully developed. Consequently, the initial review covers a broader and more general domain.

As the research progressed, particularly in the survey and interview stages covered in Chapter 4, the application of CDT in the industry was not extensively explored, as the technology was still in its early stages. However, this chapter did highlight the significant need for semantics in industry, among other findings. By the time the methodology chapter was developed, the focus had narrowed significantly, focusing on Cognitive Digital Twins.

It is also worth noting that the field of CDT is not yet mature, which presents its own challenges. As a result, the literature remains somewhat high-level, and there is insufficient

material to thoroughly analyse every methodology employed, such as the Axiomatic Design Theory or the integration of knowledge transfer using logical reasoning through an ontological approach. Furthermore, the methodologies implemented, which are not extensively reviewed in existing literature, are part of the novelty of this PhD work, contributing to the advancement of knowledge in this emerging field.

For these reasons, readers may notice that the initial chapters cover a broad and general domain, gradually narrowing down to the more focused and specialised topic of CDTs in the methodology chapter. This progression reflects the natural evolution of the research as the topic became more defined and the scope was progressively refined.

Chapter 3

Systematic literature review

Cognitive digital twin: an approach to improve maintenance management

Published as Rosario Davide D'Amico, John Ahmet Erkoyuncu, Sri Addepalli, Steve Penver; Cognitive digital twin: an approach to improve the maintenance management; CIRP Journal of Manufacturing Science and Technology; 38 (2022), pp. 613-630; 10.1016/j.cirpj.2022.06.004

Abstract

Digital twin (DT) technology allows the user to monitor the asset, specifically over the operation and service phase of the life cycle, which is the longest-lasting phase for complex engineering assets. This paper aims to present a thematic review of DTs in terms of the technology used, applications, and limitations specifically in the context of maintenance. This review includes a systematic literature review of 59 articles on semantic digital twins in the maintenance context. Key performance indicators and explanations of the main concepts constituting the DT have been presented. This article describes the evolution of DTs and their characterisation for maintenance purposes. It provides an ontological approach to develop a DT and improve maintenance management leading to the creation of a structured DT or a Cognitive Digital Twin (CDT). Moreover, it points out that using a top-level ontology approach should be the starting point for the creation of CT. Enabling the creation of the digital framework that will break down silos, ensuring a perfect integration in a network of twins' scenario.

Keywords

Basic Formal Ontology (BFO), Cognitive Twin, Digital Twin, Ontology, Predictive maintenance, Top-Level Ontology (TLO).

3.1 Introduction

The DT definition has evolved in the last decade and still changes depending on applications and purpose. Generally, we can briefly define a DT as a dynamic digital representation of a real entity; it gains value if applied across the whole life cycle of the asset of interest.

DT cooperates well with other related paradigms, such as cyber-physical systems and Industry 4.0, and it is predicted that half of the large industrial companies will implement DTs by 2021, resulting in those organisations gaining a 10% improvement in effectiveness [Velosa and Kutnick, 2017]. The National Infrastructure Commission (NIC) in 2017 stated that *"the UK needs a digital framework for data on infrastructure to harness the benefits from sharing better quality information about its infrastructure; how it is used, maintained, and planned. A digital framework will enable a better understanding of interdependencies between infrastructure sectors and help to break down silos."* (National Infrastructure Commission, 2018).

Degradation is a continuous process pertaining to the deterioration of the structure of the asset. Complex engineering assets are characterised by their maintenance phase being the longest-lasting and resource-consuming phase of the life cycle and more efficient maintenance management could lead to efficient use of resources and optimisation of costs [Price and Bodington, 2004]. Degradation in complex engineering assets is a demanding and challenging area and understanding and establishing a cause-effect model will help maintain the asset. Such damage reduces performance and shortens the life of the product. Collecting data across the life cycle together with accurate degradation models leads the DT to predict better the remaining useful life (RUL) of components. This way, we can move from reactive to proactive maintenance, avoiding downtime and optimising maintenance planning and costs (predictive maintenance).

A DT stores information about its physical counterpart throughout the life cycle and can use those pieces of information to perform predictive analytics reducing unscheduled downtimes. It also helps in making more accurate decisions based on reliable information. Most organisations are siloed internally, resulting in an expensive lack of collaboration between those siloed departments. Managing assets with data and specifically with DTs can reduce business risks and improve collaboration [Oliveira, 2020]. Today, early in a system's acquisition, stakeholders invest in models that aid communication across the whole supply chain [Bickford et al., 2020].

Successful use of DTs faces significant data management challenges and needs well-organised guidance [Lu et al., 2020b]. Another challenge in deploying DTs is that there are limited standards for implementing them [Bickford et al., 2020], and different companies may use different definitions and relations within their data architectures, increasing the incompatibility of systems. A new framework for connected DTs should be developed and based on an ontological approach. Ontology is a branch of philosophy that studies the kind of things that exist and the relations between them. This framework facilitates the effective integration of the models, data, and systems (such as software, comms, etc). This paper's authors pursue em-

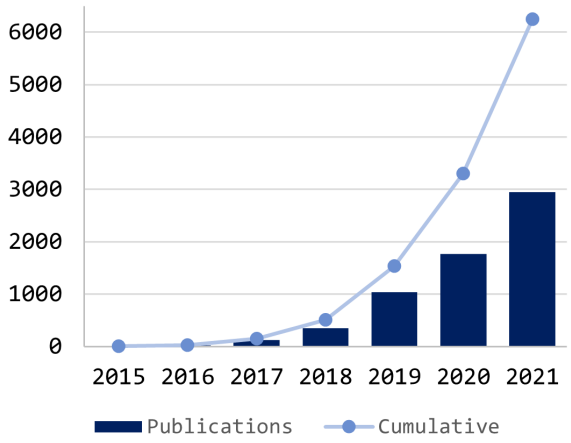


Figure 3.1: Number of publications containing ”digital twin” in title/abstract/keywords on the Scopus database from 2015 to 2021

ploying a top-level ontology-based DT to address the interconnectivity issue mentioned above. In fact, DTs based on the same domainless upper ontology facilitate the integration of twins, enabling the creation of a network of twins. The latest executive report from Accenture published in March 2021 highlighted that 87% of executives acknowledge DTs as an increasingly fundamental tool to enable a strategic collaboration with business partners [Musiolik, 2021].

Interest in DTs is growing fast across industry and academia. As seen in Figure 3.1, the total number of publications published in 2021 is about half of those published over the last seven years. Therefore, not all literature reviews published in 2020 take into account the majority of publications released so far.

There are already several reviews about DTs across the whole life cycle that also depict their main characteristics. Instead, this review focuses on analysing the maintenance and health monitoring phase of the life cycle, as one of the more complicated phases for complex engineering assets [Bickford et al., 2020]. For this reason, this paper aims to present a review of DTs in terms of technologies used, applications, and limitations, specifically for the maintenance phase of the asset's life cycle. Moreover, this paper aims to introduce the association between the concepts of DT and ontology to improve the interoperability of DTs. Previous research reported by D'Amico et al. provides an initial approach linking the use of an ontology to develop a DT [D'Amico et al., 2021].

This association of an ontology-based DT has led to the recent introduction of the concept of cognitive twin (CT) or cognitive DT. Lu et al. define the CT as a DT 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 [Lu et al., 2020a]. One of the aims of this paper is also to guide the reader from the evolution and review of the standard concept of DT used in maintenance to the opportunity of implementing the CT to improve maintenance management.

A Systematic Literature Review (SLR) has been carried out to address the above-mentioned aim. Booth et al. composed a comprehensive methodology to conduct an SLR [Booth et al., 2012] using a 4-stage analytical framework: Search, AppraisaL, Synthesis and Analysis (SALSA). This generic approach is well-validated and can be applied explicitly across different fields [Fernández del Amo et al., 2018], [Palmarini et al., 2018]. Therefore, it has been adopted in this review to provide a clear and valid analysis of the literature published in the field of DTs for maintenance and health monitoring, identifying relevant research gaps.

The rest of this paper is organised into four sections. Following this section, the literature review section reports complimentary review papers about DTs. The methodology section reports the procedure that has been carried out in this paper. The results section reports the results and the characterisation of the DT for maintenance, highlighting the concept of DT, its evolution over time, and the difference with existing models like BIM (Building Information Modelling). Finally, the conclusion section reports findings, limitations, and future research steps.

3.2 Literature review

Several literature reviews about DTs have been published so far. Searching on the Scopus database for (“digital twin” AND “review”), we find that 436 papers have been published from 2015 to 2021 – the same search string until 2023 (the time of this thesis) gave 1543 publications. Figure 3.2 shows the distribution of the literature review publications about DTs. Specifically, at the time of this review (2020), searching for (“digital twin” AND “systematic” AND “literature review”), 15 papers have been published between 2019 and 2020, and just 4 of them are journal articles. The same search performed until 2023 resulted in 210 papers with 49 journal articles.

Since this article is a systematic review, a brief analysis of the four systematic reviews mentioned above has been carried out.

After a systematic literature review, Bányai et al. introduced the production structure as a cyber-physical system focusing on logistics. In that research, DTs are solutions to support prediction and optimisation performance. The search string used there was (“manufacturing” AND “logistics” AND “optimisation”) and the database used was Web of Science. Bányai et al.’s study considered DTs only as part of their production system and exploited associated digital technologies to optimise both the manufacturing as well as the logistics associated with it [Ágota Bányai et al., 2019].

Wanasinghe et al. focused their systematic review on the application of DTs in the oil and gas (O&G) sector. The authors of that review considered journals, books and conference papers published between 2003 and 2020 in Elsevier, IEEE Xplore, OnePedro, Scopus, and Springer databases. The search string used in that review was (“digital twin” OR “digital twins” OR

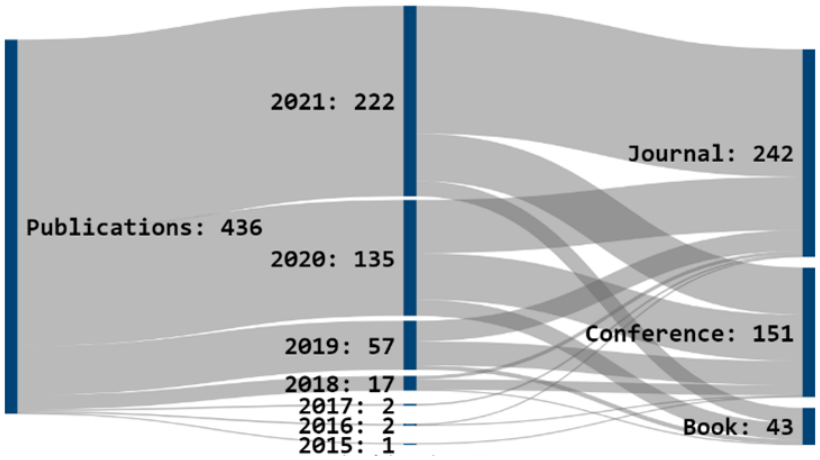


Figure 3.2: Distribution of literature review publications about DTs (Tool: sankeymatic.com)

"digital models" OR "digital environment" OR "digital environments" OR "virtual twin" OR "virtual twins" OR "virtual environment" OR "virtual environments") AND ("oil" AND "gas"). In their review, Wanasinghe et al. looked at what a DT is, the status of DT deployment in the O&G industry, and key opportunities and challenges in DT deployment in the O&G industry. They highlighted the top ten enabling technologies for DTs as well as the top ten O&G-related application areas. Asset integrity monitoring has been highlighted as the top DT application in the latter classification. The review also pointed out that existing data integration is one of the major challenges in today's implementation of DTs, particularly in the O&G industries [Wanasinghe et al., 2020].

Götz et al. performed a systematic literature review to map use cases of DT, Internet of Things (IoT), blockchain and smart contract technologies. The review aimed to investigate the feasibility of a blockchain-based DT for asset life cycle management. Mainly looking for functionalities, interoperability with current applications and integrability enablers for blockchain-based DT. Götz et al. looked at Scopus and Science Direct databases and limited their search to publications between 2019 and 2020 with the following search strings: ("digital twin" OR "Internet of Things" OR "Blockchain" OR "smart contract") AND ("asset" OR "facility" OR "building" OR "real estate" OR "life cycle" OR "management"). The authors in the conclusions highlighted the significant potential of the DT applications, specifically focusing on the real estate industry. They also pointed out the importance of creating a common ontology ecosystem to implement the blockchain-based DT for pilot testing better [Götz et al., 2020].

Jones et al. provided a systematic literature review and a characterisation of the DT, in addition to identifying gaps and required areas of future research. The research was performed at the end of 2018, the search string was ("digital twin") in the Google Scholar database [Jones et al., 2020]. They also added all the papers that cited the three seminal papers of DTs written by Grieves [Grieves, 2015, Grieves and Vickers, 2016] and Tao and Zhang, resulting in a total of 92 publications [Tao and Zhang, 2017]. The authors identified 13 characteristics of

SALSA framework	
Step	Description
SEARCH	Find the scope (PICOC framework)
	Preliminary literature search
	Full literature search and reference management
	Selection of articles
APPRAISAL	Quality assessment and selective data extraction
SYNTHESIS	Data synthesis
ANALYSIS	Data analysis
	Reporting

Figure 3.3: SALSA methodology steps description. Inspired by [Booth et al., 2012].

the DT and gaps in research that include DT across the product life cycle. Jones et al. pointed out that the DT would benefit from a more detailed comparison and review, mentioning health monitoring and prognostics as an example.

Results from the articles mentioned above strengthen our motivation to publish this systematic literature review. As mentioned before, the scope of this paper is to investigate applications of DTs that focus on the maintenance and health monitoring phase and not on the whole life cycle.

3.3 Methodology

The literature review is defined by [Booth et al., 2012] as "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesising the existing body of completed and recorded work made by researchers, scholars, and practitioners". Therefore, to make the literature systematic, the methodology applied in this article follows the frameworks given in the book "systematic approaches to a successful literature review" by Booth et al. [Booth et al., 2012]. The analytical framework proposed by Booth et al. is well-validated and consists of four steps: search, appraisal, synthesis, and analysis (SALSA framework). Figure 3.3 presents a description of the methodology's steps. The four steps are described in the following subsections.

Table 3.1: Research scope definition - PICOC framework

Concept	Description
Population	DTs and CPPS use semantic capabilities that are used to track through-life degradation in the maintenance phase.
Intervention	Investigation of existing techniques – DTs for machine health monitoring, maintenance, and asset management purposes.
Comparison	Current industrial practices – How does the new proposal compare to the existing methodology?
Outcomes	Characterisation of an open framework of DT, with flexible and reliable interfaces that enable interconnections between agents.
Context	Maintenance engineering.

3.3.1 Step 1 - Search

Defining the research scope is necessary to mould clear, answerable questions. This review adopted the PICOC (Population, Intervention, Comparison, Outcome, Context) framework (Table 3.1) to frame the aim, objectives, and research questions.

Aim, objectives, and research questions

Through the PICOC framework (Table 3.1), the aim has been derived, considering also the motivation highlighted in previous sections. Therefore, this review aims to investigate the possible standards, principles, and methodologies in using DTs for the maintenance of complex engineering assets, as well as to characterise a DT model that can best suit this application.

Actions taken to achieve the aim are the following objectives:

- Identify current practices in using through-life DTs to track the health status of assets.
- Investigate DTs that can predict the remaining useful life of an asset.
- Perform a thematic analysis to understand key aspects of a DT for maintenance management.
- Characterise a DT model for maintenance planning.

The *research questions* that this paper aims to answer are the following:

- How are DTs for the maintenance phase of the life cycle developed?
- How can a DT for maintenance be characterised?

Search strategy for identification of articles

A defined literature search strategy is required for this review to be systematic. These will then be filtered according to the inclusion and exclusion criteria, which define the filters that

Table 3.2: Inclusion and exclusion criteria

Inclusion criteria (IC)	IC1. Publications related to health and degradation assessment IC2. Publications related to maintenance IC3. Publications related to asset management
Exclusion criteria (EC)	EC1. Inclusion criteria not met EC2. Not written in English EC3. Full paper not available EC4. Published before 2011

have been considered for the selection of the studies for this review (see Table 3.2). Search databases, search strings, and article types have also been identified for this review.

After a preliminary literature search, described in the previous section, the authors became aware of synonyms used in research for DTs and research gaps highlighted in other reviews. Eventually, the search string chosen for this review must contain the word "digital twin" and its synonyms, maintenance-related words, and semantic/ontology words. Cyber-physical production systems (CPPS) have been considered a synonym for DTs. The reviewers suggested CPPS publications, which, together with the semantic approach, are aligned with the purpose of this review. Later in the methodology, the authors will refer just to DTs, considering both DTs and CPPS, because DTs have been chosen as the main focus of the research project. The search string adopted can be seen in Table 3.3. The search string is applied for searching in title, abstract, and keywords among the papers in three scientific databases: Scopus, IEEE eXplore, and Web of Science.

Mendeley (https://mendeley.com) has been used to manage all the references due to its strong community and support, as well as the integrated PDF viewer and the add-in for Microsoft Word used to write this review. All the available pdf files have been downloaded and uploaded to Mendeley.

3.3.2 Step 2 - Appraisal

The appraisal phase consists of evaluating search results to refine the number of publications to those relevant and in accordance with the scope of this review. In order to make this process systematic, a set of specific inclusion and exclusion criteria based on the PICOC framework has been used (Table 3.1). Table 3.2 lists the inclusion and exclusion criteria adopted within this systematic review.

The literature search was conducted on the 6th of April 2022, resulting in a total of 59 publications. Of these, seven papers were not available for download. Based on their abstracts, these seven papers were deemed not relevant to the research according to the inclusion criteria (see Table 3.2). Therefore, although the 'Publication elements' of these papers were initially recorded (see Table 3.4), they were subsequently excluded from further analysis due to their

Table 3.3: Screening criteria

Database	Scopus, IEEE eXplore, and Web of Science
Search string	(”digital twin” OR ”digital thread” OR ”virtual twin” OR ”cyber twin” OR ”digital replica” OR ”cognitive twin” OR ”cognitive digital twin” OR ”digital data hub” OR ”cyber physical production system” OR ”CPPS” OR ”digital shadow” OR ”digital mirror”) AND (”degradation” OR ”deterioration” OR ”maintenance” OR ”failure” OR ”damage” OR ”defect” OR ”downtime” OR ”maintenance repair and operation” OR ”MRO” OR ”health management” OR ”condition monitoring” OR ”CBM” OR ”condition based monitoring” OR ”remaining useful life” OR ”RUL”) AND (”ontology” OR ”semantic”)
Document type	Journal and conference paper

unavailability and irrelevance (see exclusion criteria in Table 3.2). The remaining publications were thoroughly analysed, and all relevant information was extracted and classified in the following steps.

3.3.3 Step 3 - Synthesis

The synthesis phase consisted of extracting and classifying relevant information from the selected articles. This review step will identify what the literature says, and in the next step - the analysis - what the literature means. The data extraction phase consisted of filling a Microsoft Excel spreadsheet with data from articles, to manage and allow different studies to be appraised consistently [Booth et al., 2012]. This included records that can be summarised in four main categories: publication elements, study elements, characteristics, and conclusion elements. Publication elements were recorded for the 59 sources found. All the records selected for the synthesis phase are listed in Table 3.4. The eligibility was established in three main stages: i) title - abstract - keywords screening; ii) introduction - conclusions reading; iii) full-text reading. This allowed papers to be summarised into thematic categories, identifying relationships for synthesis.

While the subdivisions of Publication elements, Study elements, and Conclusion elements categories are typical of almost every SLR, the Characteristics subdivision identifies attributes related only to DTs. Taking advantage of other reviews and characterisations already published, a comprehensive subdivision that best represents the typical characterisation of a DT has been implemented. An example of a technical and comprehensive review of techniques used to create DTs was published by Lim et al. [Lim et al., 2020].

That review contains specific technical aspects that everyone who wants to develop a DT needs to deal with. Other broader characterisations linked with the DTs' world can also be

Table 3.4: Records for the synthesis analysis

Publication elements	DOI Title Authors Journal/Conference Name Year Keywords Inclusion criteria
Study elements	Field/Sector Purpose of the paper Research question(s)
Characteristics	Definition of DT DT or CPPS Life cycle stage(s) Ontology information Reasoning engine Taxonomy Knowledge graph information Architecture of DT Data exchange protocols for DT Data mapping dictionary and format Middleware platforms Databases and data management Data format, data model, and representation Modelling and simulation software Virtualisation microservices Integration in existing standards Interface Visualisation Cloud Edge computing Data gathering rate
Conclusion elements	Benefits Findings Limitations Future works

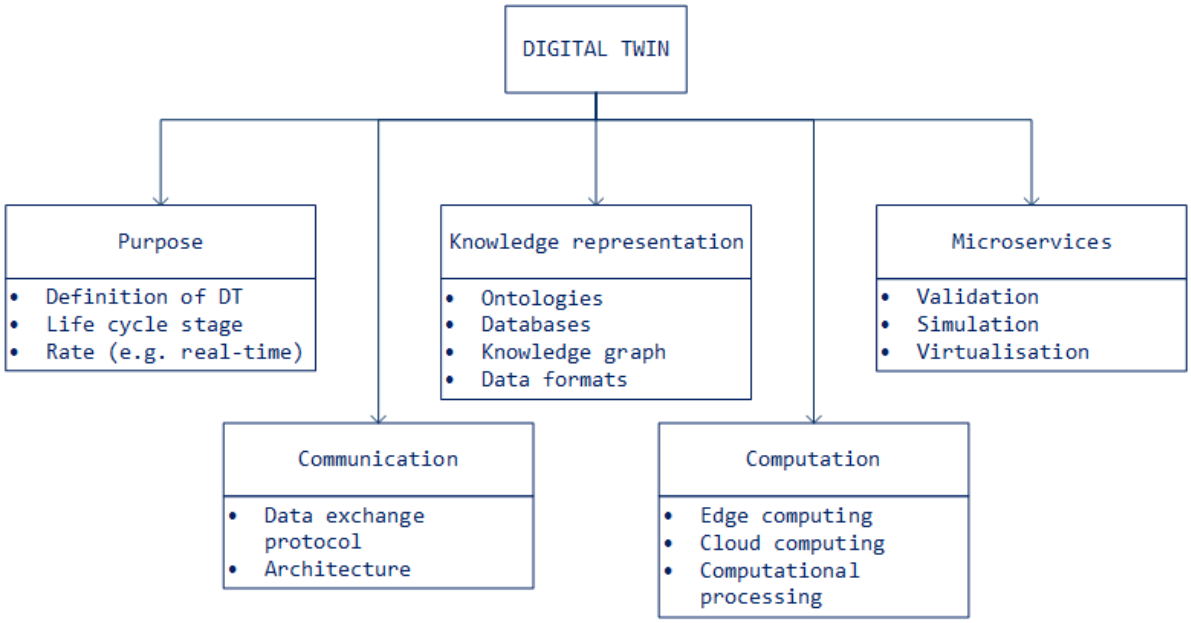


Figure 3.4: Classification of the category's division for the review of DTs

found in other systematic reviews, such as the ones mentioned in the literature review section of this article. Figure 3.4 shows the results of the thematic analysis dedicated to the characterisation of DTs. This helps the understanding of the specific category of Characteristics in Table 3.4.

3.3.4 Step 4 - Analysis

The analysis phase consisted of the extraction of data from the reviewed articles into the categories previously introduced in the previous section (Step 2 - Appraisal). This phase is comprised of three steps [Booth et al., 2012]: i) thematic analysis, which covers each theme and correlations over the reviewed articles, assessing either qualitative data or quantitative results. This step is described in this section; ii) discussion of results, which evaluates the outcomes to answer the research questions of this review and will be discussed in the results section; iii) conclusions are drawn, which include findings, limitations of the study and future research opportunities, discussed in the conclusion section.

The following sub-section will describe the four characteristics found in the previous subsection: publication elements, study elements, characteristics, and conclusion elements.

Publication elements

This part of the analysis includes general basic elements about the 59 publications found before reading the full-text:

- DOI (Digital Object Identifier)
- Title of the publication

- Authors
- Journal or conference name
- Year of publication
- Keywords
- Inclusion criteria

The points (a to e) are standard data, and the last one (g) is the inclusion criteria, a record added to understand the inclusion criteria the examined article has met.

Study elements

In this part of the analysis, further elements of the articles were collected, such as:

- Sector
- Purpose of the paper
- Research question(s)

In this phase, there was further contextualisation about the purpose of each reviewed article, for instance, the type of sector and industry to which the DT was applied, or the purpose that led to the DT deployment. Also, the fundamental research questions in the publications have been identified.

Characteristics

This was the core part of the analysis, where specific characteristics of each DT were gathered. As can be seen in Figure 3.4, each record is classified within a subcategory. The subcategories are:

- Purpose
- Communication
- Knowledge representation
- Computation
- Microservices

The following is a brief description of what is included in each subcategory.

Purpose Since there is limited standardisation about DTs, today, everyone shapes the definition of a DT according to the purpose for which they want to use this approach. Therefore, in the purpose subcategory, specific definitions of the DT applied to each case have been categorised, the pace, or rather the rate of data collection (if real-time or other), and the life cycle stage in which the DT is applied. All the selected articles handle DTs for maintenance, specifically, looking at the Operation and Service (O&S) phase of the life cycle. Nevertheless, in this analysis, we wanted to check if any DT considered for that specific purpose, was also designed to assist the asset through the whole life cycle (from the concept to the retirement and disposal stage) or in particular from the service to the design stage (feedback loop).

Communication The communication subcategory includes information about how a DT acquires and transmits data. The focus will be on the DT architecture, the structure that allows the DT to receive and transmit data. Data exchange protocols and middleware platforms to ease information exchange are also considered in this analysis. The specific characterisation of the technologies used in this part of the DT will be discussed in the results section of this chapter.

Knowledge representation The knowledge representation subcategory includes information about how the knowledge of the DT is represented, and how the semantics of the DT is represented. In case there is an ontology, is it dedicated or does it refer to other top-level ontologies? Details about the reasoning engine running over the ontology have been considered. Knowledge graphs, tools, and query language have been reviewed. The data format, representation tools, and relational/non-relational databases have been evaluated.

Computation The computation subcategory includes information about storage allocation, computational models for data processing, and machine learning methodologies employed in the specific application of the DT, including statistical optimisation approaches for data analysis.

Microservices The microservices subcategory includes virtualisation tools that allow monitoring of the physical process. Furthermore, the tools allow the simulation to improve the decision-making process, and validation tools enable task verification to confirm accuracy and integrity.

Conclusion elements

This subsection includes the fourth and last category analysed for this review (see Table 3.4). This is to collect and list the outcomes of the reviewed articles, including the benefits of using DTs, findings, limitations, and future work suggestions.

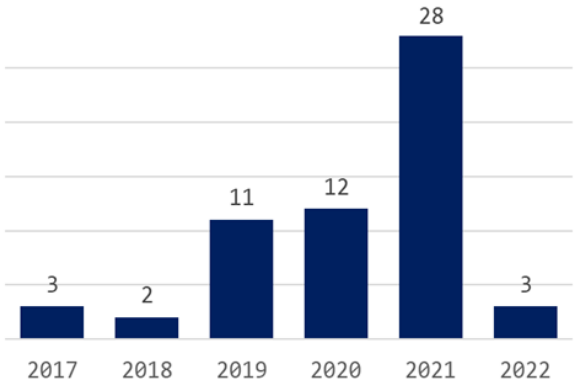


Figure 3.5: Distribution of the reviewed publications per year

3.4 Results

This section presents and discusses results for each of the four subcategories (publication elements, study elements, characteristics, and conclusion elements). Moreover, the characterisation of a DT for maintenance has been proposed below.

3.4.1 Publication elements

Figure 3.5 shows the distribution of the selected 59 publications according to the publication year. As can be seen, in 2021 alone, the same number of papers have been published as were published from 2017 to 2020 (28 papers published in 2021 and 28 papers from 2017 to 2020). This trend clearly shows the increased growth in this area of research. The year 2021 by far has seen the largest number of publications in the area of DTs with semantic capabilities for maintenance, highlighting the importance and timeliness of this article.

Figure 3.6 shows the distribution of the reviewed publications between journal and conference papers. Conference papers are 34 out of 59 (58%), while journal articles are 25 out of 59 (42%). The Sankey diagram in Figure 3.7 shows the distribution of the 59 papers per year and per publication type.

Table 3.5 and Table 3.6 show respectively, the list of journals and conferences where the reviewed papers have been published. All the 59 selected articles are divided into these two tables.

Authors of the selected articles have been gathered and ranked. Figure 3.8 shows just the list of authors that published two or more articles in this field (not only as of the first author). From this analysis, Ansari F., Kiritsis D., and Erkoyuncu J. are the top 3 authors in this specific field.

Figure 3.9 shows the keywords used at least three times in the reviewed articles. DT and synonyms have been used 31 times as keywords. The extended version and the acronym have been grouped (i.e., Asset Administration Shell and AAS).

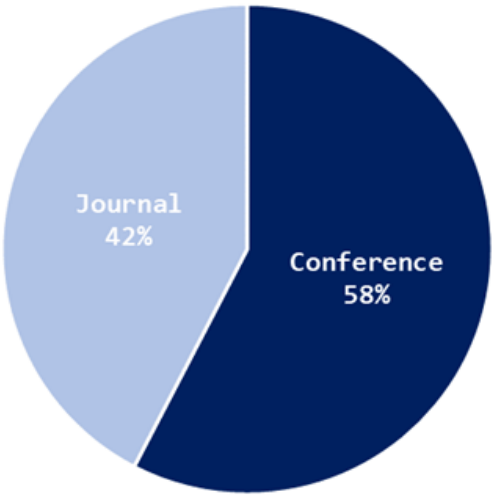


Figure 3.6: Distribution of the reviewed publications between journals and conferences

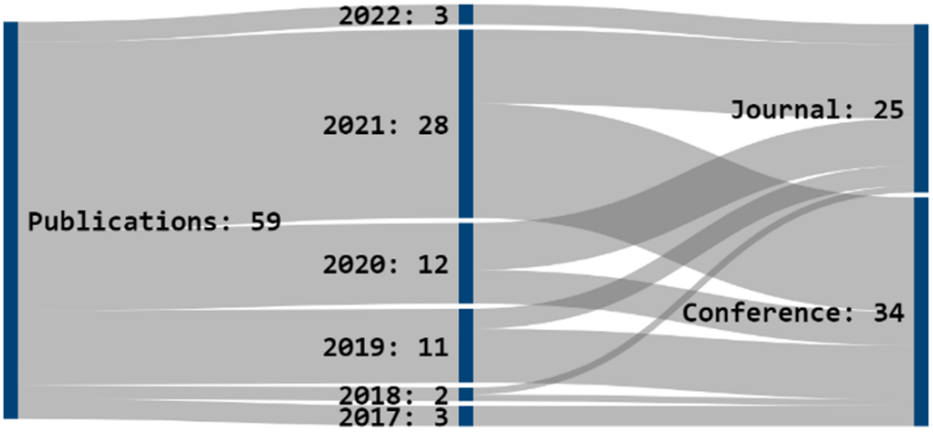


Figure 3.7: Sankey diagram with the distribution of published papers divided per year and type of publication

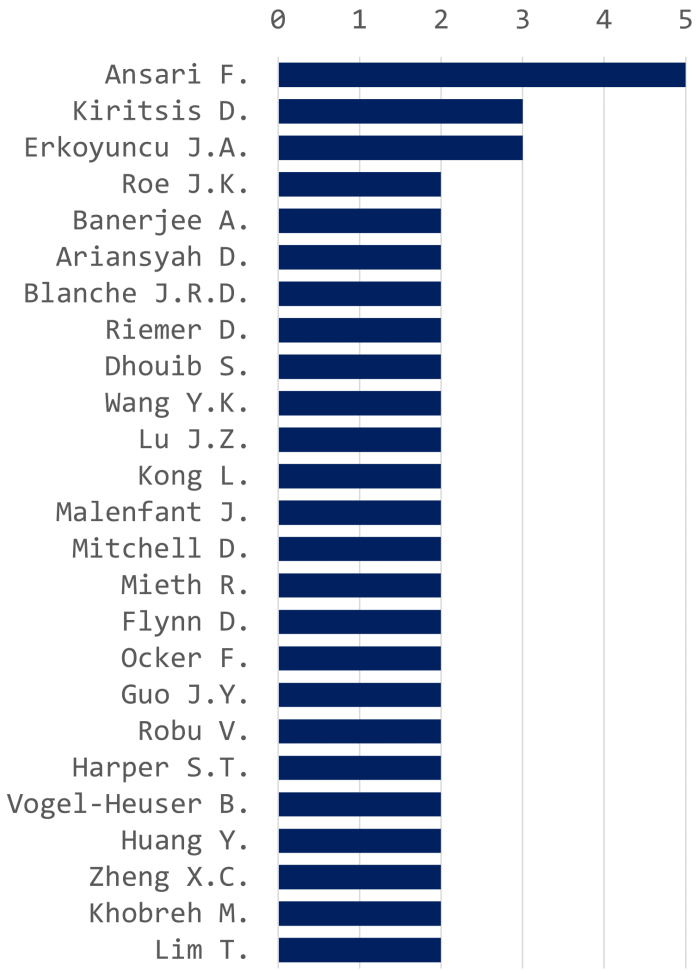


Figure 3.8: List of authors having two or more articles within the final selection

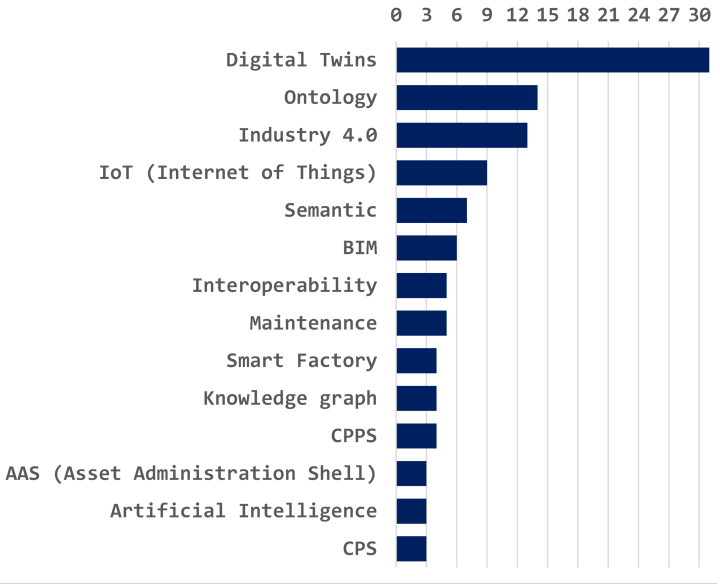


Figure 3.9: List of keywords used at least 3 times in the reviewed publications

Table 3.5: List of journals

Journal name	**Reference**
Annual Reviews in Control	[Longo et al., 2019] [May et al., 2022]
Applied Sciences	[Moretti et al., 2020] [Sahal et al., 2021]
CIRP Annals - Manufacturing Technology	[Erkoyuncu et al., 2020]
CIRP Journal of Manufacturing Science and Technology	[Ansari et al., 2018] [Ansari et al., 2020]
Energy and Buildings	[Hosamo et al., 2022]
GENES	[Santos et al., 2019]
IEEE Access	[Mitchell et al., 2021]
IEEE Systems Journal	[Lu et al., 2021]
International Journal of Automation Technology	[Montavon et al., 2020]
International Journal of Computer Integrated Manufacturing	[Ansari, 2019]
International Journal of Production Research	[Lu et al., 2022]
Journal of Ambient Intelligence and Humanized Computing	[Jiang et al., 2021]
Journal of Construction Engineering and Management	[Li et al., 2022]
Journal of Nondestructive Evaluation	[Vrana, 2021]
Knowledge-Based Systems	[Zhang et al., 2020]
Magazine of Civil Engineering	[Badenko et al., 2021]
Naval Engineers Journal	[Kumar et al., 2020]
PFG-Journal of Photogrammetry Remote Sensing and Geoinformation Science	[Lehner and Dorffner, 2020]
Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences	[Vogel-Heuser et al., 2021]
Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture	[Singh et al., 2021]
Sensors	[Bamunuarachchi et al., 2021]
Tunnelling and Underground Space Technology	[Yu et al., 2021]

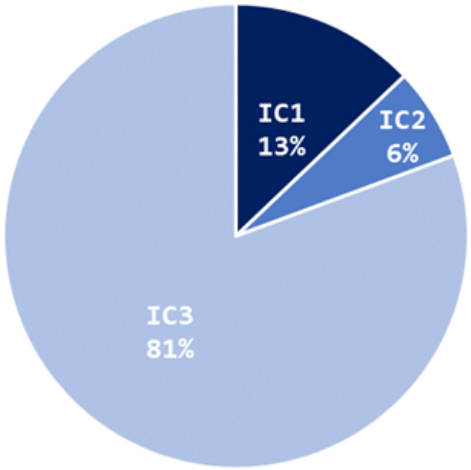


Figure 3.10: Distribution per IC of the reviewed articles

Figure 3.10 shows the distribution of the reviewed articles per IC (see Table 3.2). The pie chart shows that 50 (81%) papers are aligned with IC3 (asset management), whereas eight papers (13%) are aligned with IC2 (maintenance). The remaining 6% (four papers) are aligned with IC1 (health and degradation assessment).

3.4.2 Study elements

Figure 3.11 shows the distribution of 59 publications per field of application. Following, a description of each field of application adopted for this review now follows.

- 'Manufacturing' concerns DTs using semantics used for health monitoring, maintenance, or asset management in manufacturing processes, such as CNC machine tools, gearboxes, etc. This also includes publications about smart manufacturing, CPPS, robotics, and Industry 4.0.
- The 'Civil' category focuses on monitoring structures, mostly bridges, tunnels, and buildings.
- 'Generic' is a category where DTs were studied with a domainless approach.
- 'Infrastructures' refer to electrical systems components in power plants or buildings (e.g., batteries or fuel cells), or systems like renewable energy systems.
- 'Transports' refer to DTs designed to monitor mobility systems, not only the car or the train but also the integration of intelligent vehicles in a smart city environment.
- In 'Food', there is an article about semantic digital technologies for cost evaluation in the dairy sector.

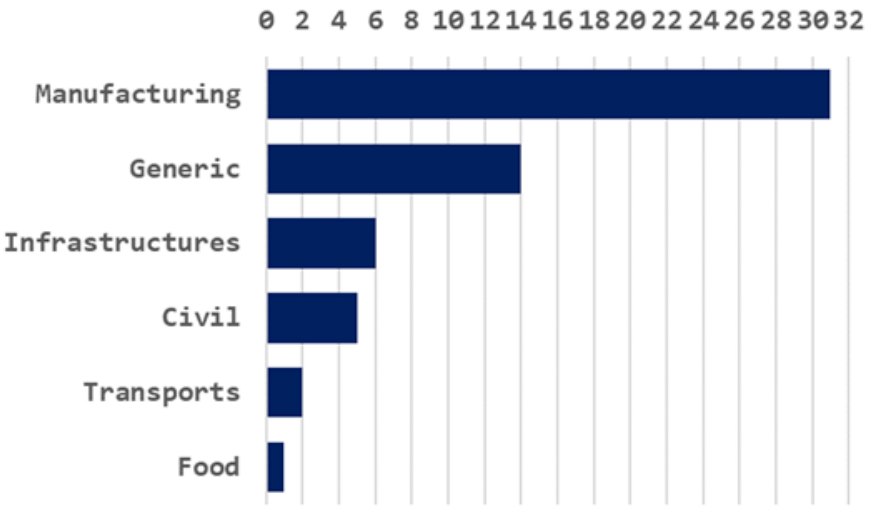


Figure 3.11: Distribution of the reviewed papers per field of application

3.4.3 Characteristics of the DT

Results concerning the characteristics category can be divided into five subcategories (purpose, communication, representation, computation and microservices) as described in the previous methodology section. The description of each subcategory results and the classification are reported in Figure 3.4.

Purpose of the DT

The three aspects highlighted in this review concerning the purpose are the definition of the DT, the life cycle stage, and the rate at which the DT works.

Concerning the definition of the DT, most publications adopted the one used in the white paper of Michael Grieves, which is defined simply as a virtual, digital equivalent to a physical product [Grieves, 2015]. All the other definitions found in the publications were collected in Table 3.7.

Table 3.6: List of conference names

Conference name	**Reference**
ACM WEBSCI	[Banerjee and Mittal, 2017]
ASME	[Zhiqiang et al., 2019]
CITI	[Gómez-Berbís and de Amescua-Seco, 2019]
DAI-SNAC	[Salama et al., 2021]
DASC/PiCom/CBDCom /CyberSciTech	[Xiao and Li, 2021]
DCOSS	[Cho et al., 2019]
ESWC	[Kainzner et al., 2021, Vakaj and Cheung, 2021]
IABSE	[Törmä et al., 2019]
IECON	[Huang et al., 2021b]
IEEE BigData	[Riemer, 2019] [Zehnder and Riemer, 2019]
IEEE CBI	[Zinnikus et al., 2017]
IEEE ETFA	[Deng et al., 2021, Huang et al., 2021a] [Tuli et al., 2021, Rahm et al., 2017]
IEEE ICCC	[Liu et al., 2019]
IEEE ICPS	[Redeker et al., 2021] [Sahlab et al., 2021]
IEEE IEEM	[Ocker et al., 2022]
IEEE INDIN	[Prenzel and Provost, 2019]
IEEE ISIE	[Najdataei et al., 2019]
IFAC PapersOnLine	[Ansari, 2019]
IFIP AICT	[Defèr et al., 2020]
IPIN	[Lee et al., 2021]
ISWC	[Martín-Recuerda et al., 2020]
MLSD	[Grebenyuk et al., 2021]
OTC	[Zaki et al., 2020]
Procedia CIRP	[Farsi et al., 2021]
SAM IoT	[Tao et al., 2020]
SeDiT	[Beetz, 2021, Ihsan et al., 2021]
WIT Transactions on the Built Environment	[Agostinelli, 2021]

Table 3.7: List of definitions of DT found in the reviewed articles in chronological order

Definition	Reference
"Digital Twin models are computerized clones of physical assets that can be used for in-depth analysis."	[Banerjee and Mittal, 2017]
"The digital twin (DT) is a digital representation of a physical asset that can be used to describe its properties, condition, and behaviour through modelling, analysis, and simulation."	[Erkoyuncu et al., 2020]
"Digital representation of an observable manufacturing element with a means to enable convergence between the element and its digital representation at an appropriate rate of synchronization. (ISO 23247)"	[Farsi et al., 2021]
"DT for buildings can be seen as BIM models extended to capture real-world data and feed it back into the model, thus neatly closing the information loop of asset life cycle management."	[Lee et al., 2021]
"An integrated information flow that connects all the phases of the product life cycle using accepted authoritative data sources (e.g., requirements, system architecture, technical data package (TDP), three-dimension (3D) CAD models)."	[Deng et al., 2021]
"The DT is an exact digital replica of something in the physical world."	[Sahal et al., 2021]
"An integrated plant model 'maintained and kept consistent throughout the entire life of the plant' is a DT."	[Vogel-Heuser et al., 2021]
"DT is a digital representation of a real physical asset that represents the relevant features of the real asset within a model."	[Ihsan et al., 2021]
"DT is a digital copy of a physical asset that accurately represents and alters the asset state and behavior within a certain context."	[Grebenyuk et al., 2021]
"An (ideal) digital twin is a virtual representation of an asset and like a monozygotic twin it shows the same behavior and development as the asset."	[Vrana, 2021]
"Digital Twin (DT) is the combination of logically integrated models of a physical asset to give useful insights using data associated with those models."	[Singh et al., 2021]

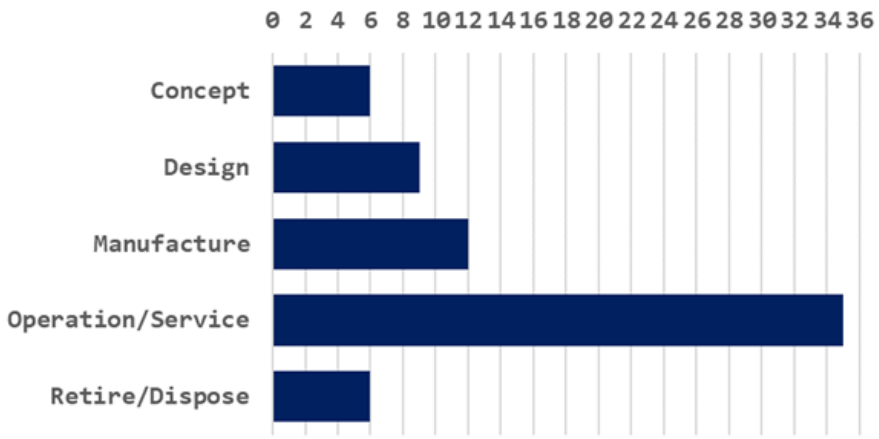


Figure 3.12: Stages of the life cycle where reviewed papers had more focus

In the life cycle stages, the main part of the life cycle where the DT has been implemented has been reviewed. Although this review focuses on maintenance, the DT should consider the whole life cycle, from the conceptualisation to the end of life. Figure 3.12 shows the focus of the reviewed papers per each phase of the life cycle. The consideration of a DT that covers the whole life cycle, which allows feedback from the O&S to the conceptualisation/design stage of the life cycle, ensures the overall system optimisation throughout the life cycle. Concerning the operational rate at which DTs operate in the reviewed articles, 22 papers out of 59 explicitly mention that the data collection rate has been performed in real-time. Although real-time operation rate is widely used and considered, I do not consider the real-time feature mandatory for a DT, especially for assets that cannot always be connected to the network, such as planes, ships, helicopters, etc.

The last factor checked in this subsection was whether the reviewed papers dealt with DT or CPPS. It has been found that 84% of the papers dealt with DT, and the remaining 16% dealt with CPPS.

Communication

In this subsection, communication protocols and DT architectures will be reviewed. The OSI Open Source Interconnection (see EN ISO/IEC 7498-1:1995 ”Information Technology - Open Systems Interconnection - Basic Reference Model: The Basic Model”) is a well-established network communication reference model proposed by Zimmermann in 1980 [Zimmermann, 1980]. A brief description of how this model works will help the reader to better understand the following classification based on communication protocols. The OSI is a seven-layer model that defines any communication between two or more machines (broadcast or unicast). The seven layers are physical, data link, network, transport, session, presentation, and application. Figure 3.13 reports the structure of the OSI model. The same figure also reports the information path either as a receiver or sender.

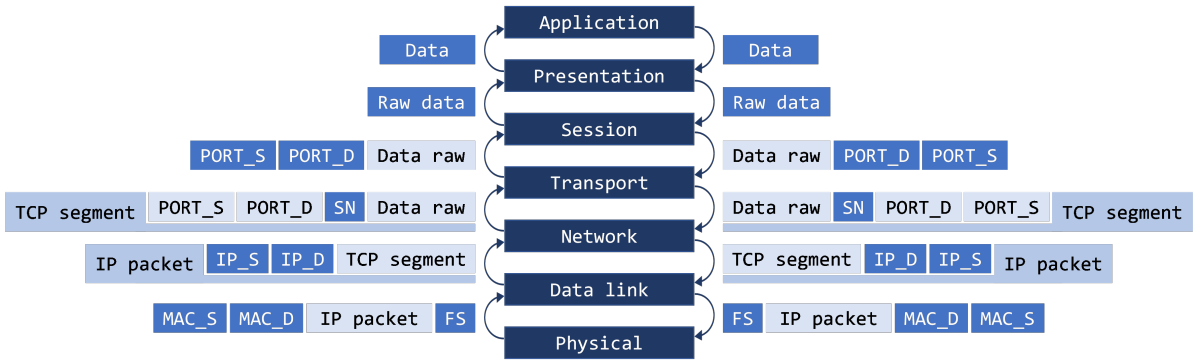


Figure 3.13: Example of the machine to machine (M2M) communication with the OSI model. The left side shows how a machine receives information from another machine (receiver) and, on the right side, how a machine sends information to another machine (sender)

For example, going from the top to the bottom of Figure 3.13is where a machine wants to send information to another machine in the network (e.g., a HyperText Transfer Protocol - HTTP request).

The first step happens in the application layer, which is the link between the application and the communication model. This is especially useful for those applications that are not network-aware.

In the next step, inside the presentation layer, the information is converted into raw data, which is readable from a machine (e.g., a text file converted into an SSL or ASCII file).

The session layer checks the ports, both destination and source port, and adds them to the information for the next layer. The source port ($PORT_S$) comes from the application and the destination port ($PORT_D$) depends on the machine (e.g., in computers, the first 1024 ports are known and fixed, and port 80 is for the HTTP port, the web port). This layer connects the server to a client on a remote system.

The transport layer puts the SN (sequence number), in the information, because sometimes a big piece of information needs to be split into several small chunks numbered with an SN for connection bandwidth issues (e.g., SN = 1/20 means that this data is the first part of 20). The transport layer then has the function of disassembling the information, in the case of a sender, and assembling the information, in the case of the receiver. At this stage, the TCP (Transmission Control Protocol) segment is composed and consists of the string that contains ($PORT_S$) (source port), ($PORT_D$) (destination port), SN (Sequence Number), and “raw data”.

The network layer can append the IP_D (destination IP address) and IP_S (source IP address) to the information. The string is now called the IP (Internet Protocol) packet.

The data link layer checks the integrity of the information that is given from the FS (frame check sequence), which will then be appended to the information.

The physical layer appends the MAC (Media Access Control) addresses (MAC_D and MAC_S) to the information. After appending the MAC addresses, the whole information packet will then be sent to the network through the physical network board (e.g., the NIC - Network Interface

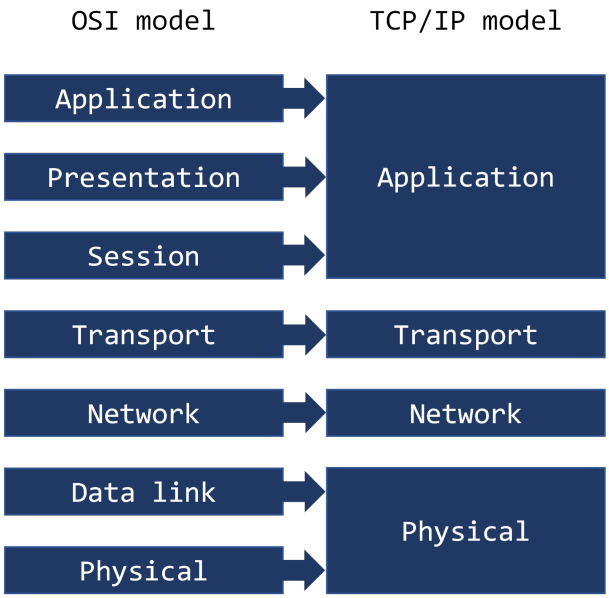


Figure 3.14: Comparison between OSI 7-layer model and TCP/IP 4-layer model

Controller).

The path is mirrored in case the information is received, as illustrated on the left side of Figure 3.13. In the above-described process, the information string has been simplified describing only the main parts of typical information to make it understandable. Modern computers do not follow the OSI model but follow the TCP/IP model. The latter is a four-layer model which can perfectly replace the former. The analogy between the two models can be seen in Figure 3.14. Hypothetically, we can replace the OSI model with the TCP/IP model in Figure 3.13, and nothing will change in the communication path. We must consider that in the TCP/IP model, the physical layer works as the OSI model's physical and data link layers, and the application layer works as the session, presentation, and application layers of the OSI model.

Having explained that the classification in terms of communication protocols used becomes easier to understand, and the results can be seen in Table 3.8. The table shows the five protocols used within the reviewed papers.

Regarding the architecture of the DTs in the reviewed articles, most of the proposed architecture models were similar to the original model proposed by Michael Grieves [Grieves, 2015]. He initially defined a DT with three layers: the physical layer, the digital counterpart, and the data exchange layer between the two. Three papers present a 4-layer architecture [Ansari, 2019, Jiang et al., 2021, Zhiqiang et al., 2019], and one presents a 6-layer architecture [Grebenyuk et al., 2021].

Knowledge representation

This subsection will review semantic capabilities, the database and knowledge graph used, and data formats.

All the 59 reviewed articles implemented a semantic aspect in their digital platform. In particular, 28 of them explicitly mentioned an ontology implementation. Only 5 out of these 28 referred to an application ontology referred to a top-level ontology (TLO), an explanation of which will follow in the conclusion section. 9 of these 28 papers have explicitly indicated the tool used to create and manage the ontology, which is Protégé, one has also specifically mentioned the use of HermIT reasoner [Zinnikus et al., 2017]. HermIT is a reasoner for ontologies written using the Web Ontology Language (OWL).

Interoperability was also one of the key elements in some of the reviewed papers. To improve interoperability, some of the papers implemented a standard terminology or a shared vocabulary. In order to do that, standard ontologies or international standards have to be applied. The Semantic Sensor Network (SSN) Ontology has been employed in three papers [Bamunuarachchi et al., 2021, Hosamo et al., 2022, Vakaj and Cheung, 2021] used a taxonomy from ISO 2041, ISO 13372, and ISO 17359:20. In [Grebenyuk et al., 2021] taxonomies from IEC61968, IEC61970, and IEC 62325 CIM. An alternative solution to improve interoperability was the use of a standard architecture. In this respect, five papers [Huang et al., 2021a, Huang et al., 2021b, Redeker et al., 2021, Vrana, 2021, Zheng et al., 2022a] exploited RAMI 4.0, the well-known Reference Architecture Model Industrie 4.0 introduced by the German Electrical and Electronic Manufacturers' Association (ZVEI) to support Industry 4.0 initiatives [Koschnick et al., 2015]. The exchange of information is enabled through the concept of an Asset Administration Shell (AAS) used to describe an asset in a standardised manner [Koschnick et al., 2015].

Three different kinds of databases have been found: relational, non-relational, and graph databases (also known as knowledge graphs). Respectively, six papers employed relational databases, three papers used non-relational databases, and 15 papers used graph databases. The main difference between a relational and a non-relational database is the structure; the non-relational database does not rely on a predefined structure. MS SQL, MySQL, MS Access, and SQLite 3 have been used as relational databases. MongoDB and Elasticsearch are used for non-relational, or NoSQL, solutions. Regarding the graph databases, GraphDB, Neo4J, and Apache Jena were the solutions found in the reviewed articles. These solutions use different query languages, SPARQL for GraphDB and Apache Jena, and Cypher for Neo4J. The query language is generally the query language used to query the databases. Using one kind of database does not exclude the use of the others. Indeed, it would be possible to use more than one kind of database within the same DT.

Table 3.9 shows a classification of the data formats found in the reviewed articles. A brief description of each data format can also be found in the table. The analysis showed that OWL and RDF are the primary data formats exploited due to the employment of the semantic approach and knowledge graphs.

Table 3.8: Communication protocols in the network architecture for the DT

Data exchange protocol	**Description**	**Reference**
TCP/IP	TCP/IP (Transmission Control Protocol / Internet Protocol) is one of the most common communication models used among network devices.	[Bamunuarachchi et al., 2021] [Prenzel and Provost, 2019] [Zhiqiang et al., 2019]
TCP/UDP	Both communication protocols are based on the transport layer. The difference between TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) is that the latter is not connection-oriented (connectionless) and prioritises time over reliability.	[Tao et al., 2020]
OPC UA	In OPC UA (Open Platform Communications United Architecture) the communication occurs at the TCP level. It is a machine-to-machine (M2M) communication protocol used in industrial automation environments. It can use either client/server or publish/ subscribe pattern.	[Huang et al., 2021a] [Huang et al., 2021b] [May et al., 2022] [Vrana, 2021]
MQTT	MQTT (Message Queuing Telemetry Transport) is an open publish-subscribe network protocol BS ISO/IEC 20922 (BS ISO/IEC 20922:2016 Message Queuing Telemetry Transport - MQTT, 2016). It allows communication between devices usually over a TCP/IP model in the application layer.	[Bamunuarachchi et al., 2021] [Redeker et al., 2021] [Tao et al., 2020] [Zhang et al., 2020]
HTTP	HTTP (HyperText Transfer Protocol) is an application layer protocol (the higher layer for both OSI and TCP/IP model). It is a request-response protocol in a client-server pattern.	[Longo et al., 2019] [Tao et al., 2020]

3.4.4 Computation

Computation starts with the selection of the storage engine. This can be local, as in the case of edge computing, or remote, as in the case of cloud computing.

Cloud computing allows data storage and/or computing power on demand and remotely. Edge computing, instead, brings storage and computation close to the location where it is needed. Both cloud computing and edge computing can coexist within the same system. Cloud computing is not always the best choice if we consider isolated systems such as ships or aeroplanes, which cannot be easily connected to a network. Edge computing can probably be a better solution for these kinds of isolated systems.

From the analysis of the articles, five out of 59 specifically mentioned the storage location. three out of 59 articles adopted a hybrid approach (both cloud and edge computing) [Redeker et al., 2021, Vrana, 2021, Zheng et al., 2022a], while [Mitchell et al., 2021] explained the use of cloud computing and [Sahal et al., 2021] explained the implementation of edge computing. In seven publications, the exploitation of the REST-API interface implies the presence of the middleware [Bamunuarachchi et al., 2021, Hosamo et al., 2022, Longo et al., 2019, May et al., 2022, Redeker et al., 2021, Tao et al., 2020, Zinnikus et al., 2017]. In four publications [Mitchell et al., 2021, Tao et al., 2020, Tuli et al., 2021, Zaki et al., 2020], ROS (Robot Operating System) has been used as middleware, while Erlang middleware has been employed in [Prenzel and Provost, 2019] and a Java solution in [Zinnikus et al., 2017].

Microservices

This section contains the results of collecting the major modelling and simulation tools, visualisation and virtualisation tools that have been reviewed. Table 3.10 lists the tools used in the reviewed publications.

3.4.5 Conclusion elements

This section gathered the benefits of using DTs, along with major findings, limitations, and future works from the reviewed articles. All the conclusions were summarised and grouped into broader concepts. The benefits of using DTs found in the reviewed articles can be summarised in the following points:

- Minimise the spread of damage
- Enable fault prediction
- Enable proactive maintenance

Table 3.9: Descriptions and references for various data/file formats.

Data/file format	**Description**	**Ref**
CSV	CSV (Comma-Separated Value) is a text-based file format used for the import and export of data in tables (e.g., databases). Each value is separated through a predefined separator (e.g., comma).	[Zaki et al., 2020]
IFC	IFC (Industry Foundation Classes) is a data model used to describe construction industry data. IFC is a platform-neutral, open format, and object-based file format developed to ease interoperability specifically in the construction industry. ISO 16739-1:2018	[Moretti et al., 2020]
JSON	JSON (JavaScript Object Notation) is a lightweight open standard file format for storing and transporting data. It is intuitive, easy to read and write, works across platforms and is used as a data-interchange language.	[Longo et al., 2019, Montavon et al., 2020, Redeker et al., 2021, Salama et al., 2021, Tao et al., 2020]
OWL	OWL (Web Ontology Language) is a markup language for explicitly representing the meaning and semantics of terms and relationships between them. OWL was developed by the W3C consortium based on XML and is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs.	[Deng et al., 2021, Hosamo et al., 2022, Huang et al., 2021b, Huang et al., 2021a, Kainzner et al., 2021, May et al., 2022, Singh et al., 2021, Vakaj and Cheung, 2021, Vrana, 2021, Yu et al., 2021, Zinnikus et al., 2017]
RDF	The RDF (Resource Description Framework) is the basic tool proposed by the W3C consortium for encoding, exchanging and reusing structured metadata and enables semantic interoperability between applications sharing information on a network.	[Cho et al., 2019, Hosamo et al., 2022, May et al., 2022, Ocker et al., 2022, Rahm et al., 2017, Sahal et al., 2021, Vogel-Heuser et al., 2021, Yu et al., 2021, Zehnder and Riemer, 2019, Zinnikus et al., 2017]
XML	XML (Extensible Markup Language) is a markup language that defines a set of rules for encoding data. It is a software/hardware-independent tool for storing and transporting data.	[Jiang et al., 2021, Longo et al., 2019, Ocker et al., 2022, Prenzel and Provost, 2019, Sahal et al., 2021]

Table 3.10: List of tools employed for modelling and simulation, visualisation, and virtualisation.

Tool	Reference
Autodesk Revit	[Hosamo et al., 2022]
Eclipse Papyrus	[Huang et al., 2021b, Huang et al., 2021a]
Halve	[Zinnikus et al., 2017]
Inversiv	[Zinnikus et al., 2017]
JavaScript	[Zinnikus et al., 2017]
Python	[May et al., 2022, Tuli et al., 2021]
StreamPipes	[Riemer, 2019]
Unity	[Tuli et al., 2021]
VoluSoft	[Montavon et al., 2020]

- Enhance the capability of interconnections of different systems and the integration in a wider environment
- Increase the information available to help analytics
- Enable a long-term vision to apply a cost-effective maintenance strategy
- Reduce downtimes, increase operation
- Increase overall efficiency and reliability

The limitations gathered can be summarised in the following points:

- High initial costs
- Management of large datasets
- Lack of automatic feedback correction in previous phases of the life cycle to improve production
- The gap between as-built and as-is condition
- Lack of guidance in the implementation
- The need for specific and dedicated solutions
- Unknown green performance evaluation
- Lack of stakeholders' awareness within the same supply chain.

Finally, future works proposed are summarised as follows:

- Application of the proposed methodology in a real case
- Expand the range of damage types to monitor

- Consider cloud and/or edge computing
- Find and fight the source of degradation
- Enable real-time monitoring
- Try the same method in different fields, check the flexibility
- Provide feedback to improve the design and production phase
- Analyse the value of having a higher-level DT that contains sub-level DTs
- Perform a sustainability evaluation
- Improve accuracy, refine DT for specific applications
- Balance costs and benefits
- Create structured knowledge for the O&S phase
- Consider a DT for the whole life cycle

3.5 Characterisation of a DT for maintenance and health management

In this section, a general characterisation of a DT for maintenance and health monitoring has been proposed. The following subsection presents the DT concept and its evolution and the conceptual model of the DT for maintenance.

3.5.1 DT concept evolution

The main reason why the DT concept rose was to move the work activity into the virtual space, perform all tests on a high-fidelity digital counterpart, see errors, correct them, and once ready, transfer the information back to the physical item.

To have a higher value from it, the DT must be supported and fed over the whole life cycle. To cover this need, Grieves divided the DT into three types that operate together in a DT Environment (DTE): the DT Prototype (DTP), the DT Instance (DTI), and the DT Aggregate (DTA) [Grieves, 2019].

DTP describes prototypical physical assets containing information about the early stage of creating the physical artefact. DTP covers all products that can be made.

DTI describes a specific corresponding physical asset to which the DT will be linked throughout life. DTI covers all products that are made.

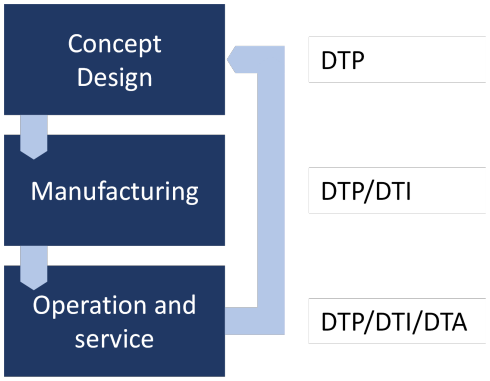


Figure 3.15: Distribution of DT types through the life cycle and information flow [from [Grieves, 2012]]

DTAs are the aggregation of all the DTIs. The first valuable aspect is to correlate previous state changes with subsequent behavioural outcomes. This enables, for example, component failure prediction when specific sensor data predefined thresholds are met. The second value can occur via a learning process when a small group of DTIs learn from actions. That learning can be conveyed to the rest of the DTIs. DTA covers all products that have been made.

The three DT types can be applied in several asset life cycle phases. DTP can be applied in the concept and design phase. Consecutively, DTI can be added in the manufacturing phase, and eventually DTA in the O&S phase until the retire/dispose phase. Figure 3.16 would highlight the two different routes of the physical and information life cycle. While the physical life cycle is usually an open cycle, the information life cycle is closed. Valuable information, mainly from O&S and disposal phases, can help improve the initial phases of successive assets. As in Figure 3.15, in the Grieves model, the feature of the DT to bring back information from the later stages of the life cycle to the early stages of the other asset's life cycle is one of the most important added values of using this technology. Product Life cycle Management (PLM) feedback loops were introduced by Grieves [Grieves, 2012], and their basic principle is using information from each phase of the life cycle to improve the design and creation phase of the assets.

All the above considerations let us think about a world of interconnected twins, where the DT of a system can be composed of a system of smaller connected DTs. For a DT to be able to learn from the experience of similar DTs in other environments, the interoperability and integrability of such platforms are necessary, thus creating a network of twins.

3.5.2 Difference between DT and model

In many cases, the difference between a DT and the model is not clear, mainly in the early stages of the life cycle. Wright and Davidson pointed out that the DT concept gathers several existing technologies and approaches and explained that a model is part of the DT. A model is primarily a static representation of an asset, but the model would not be valid for those

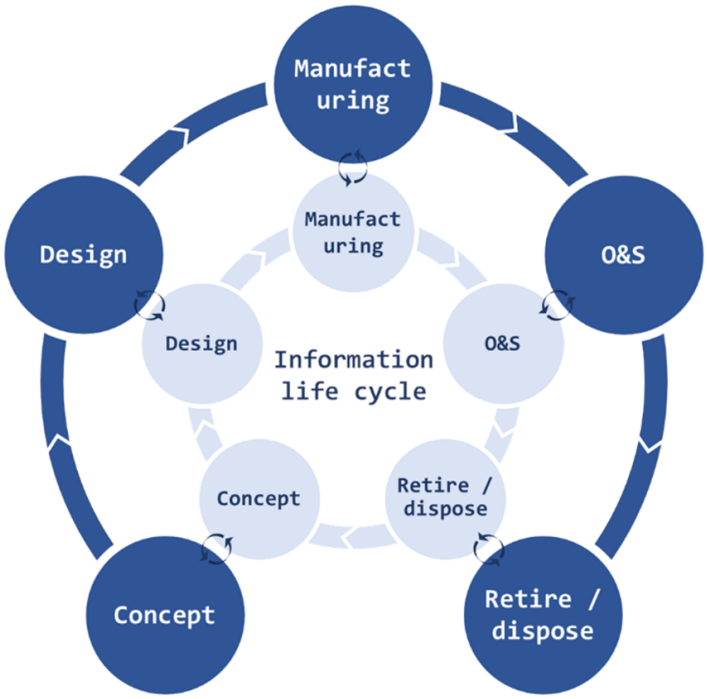


Figure 3.16: Physical asset life cycle and information life cycle

assets that change over time. For those assets, it is worth using a DT paradigm that follows the assets throughout their life [Wright and Davidson, 2020]. A DT thus adds context and time dimensions to the overall asset representation, which is merely a static model at the outset.

Kapteyn et al. pointed out that differences in material properties, manufacturing processes, and operational histories are just some of the many factors that ensure that no two engineering systems are identical, even if they share the same design parameters [Kapteyn et al., 2020].

Therefore, modelling many similar assets using a single model neglects the differences mentioned above. The DT paradigm aims to overcome those differences by providing an adaptive, comprehensive, and authoritative digital model tailored to each unique physical asset [Kapteyn et al., 2020].

3.5.3 Difference between DT and BIM

Building Information Modelling (BIM) has been revolutionary for the construction sector as it brings in a single digital platform that incorporates all the physical and functional characteristics of the building project, which were previously siloed. However, BIM does not provide a complete solution for whole-life asset management since it is a model. As said before, a model is not the perfect representation of an asset that changes over time, as a building does. From the perspective of information richness and analytical/decision-making capability, the concept of Digital Twins (DT) is broader than BIM [Xie et al., 2020]. DT is a more comprehensive solution to monitor the as-is condition of a building. Moreover, a DT integrates data analytics, control and simulation functions that allow maintainers to improve service plans

[Li et al., 2017].

3.5.4 DT characterisation

Figure 3.17 represents the conceptual model for a DT for degradation assessment. On the bottom end, in light blue, is the physical space (the real space), which represents the through-life degradation of an asset or process. Above this, in dark blue, is the virtual space, which represents the main components of a DT. Physical and virtual spaces are connected via data and information. Data acquisition goes from the physical world to the real world; information, instead, acts as dynamic adaptiveness from the digital world to the physical world.

DT ontology and taxonomy of degradation is the 'interface' block, where the dictionary of the DT will be developed. This is to give the DT context awareness. An ontology is fundamental because, if chosen correctly, it enables effective DT interconnection and the creation of the network of twins. The ontology must start from a domainless top-level-ontology. This review found that several upper ontologies exist, presented in the work from the Centre for Digital Built Britain (CDBB) [Partridge et al., 2020]. This review identified a variety of fundamental ontologies, such as BFO (Basic Formal Ontology), BORO (Business Objects Reference Ontology), and DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering). Regarding BFO, a draft version of an ISO standard (ISO/IEC 21838) was published in 2021. Choosing a standard would be helpful because it allows the DT developer to start with the same top-level ontology. Using the same top-level ontology enables the integration of the DT in a complex system composed of other DTs, even belonging to different field domains. This is because all the DTs use the same dictionary and can easily talk with each other, allowing the creation of a DT for an entire company or a city containing lower-level DTs. All of this has already been implemented successfully, specifically in biology with the implementation of the gene ontology (GO) [Smith et al., 2007]; and on the internet, with the Semantic Web.

The 'data architecture and data mining' component relies on the ontology chosen in the previous block to create the data structure of the DT and the optimal mining strategy.

'Data modelling and simulation' is essentially the DT's brain. Here, the DT can use the data to predict the behaviours of the physical asset or simulate several scenarios to choose the best configuration for the physical asset. Artificial intelligence algorithms can better help the DT predict failure based on previous patterns. This block could also be connected with other blocks, creating the twins' network, or rather the network of twins.

'Visualisation' enables the visualisation of the results or provides a graphical interface for the user to interact with the DT. This can be not only an application or a web app but also the graphical environment used in virtual reality (VR) or augmented reality (AR).

'The interface' also represents the DT's information exchange cornerstone. It allows the DT to talk with different modules and manage all the inlet and outlet information to enable the DT to work regardless of the system where it will be applied.

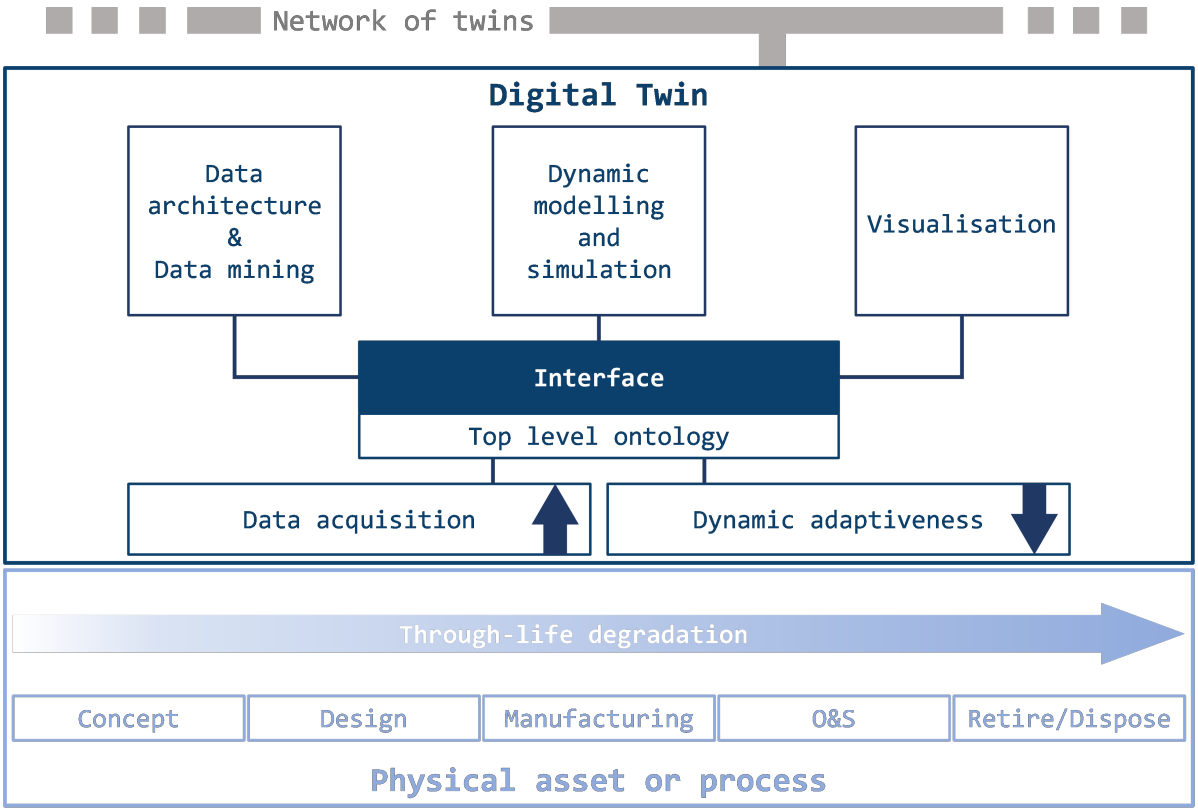


Figure 3.17: DT conceptual model characterisation

3.5.5 Maintenance and DT

A previous version of the conceptual model of DTs to monitor through-life degradation has been already published by the authors [D'Amico et al., 2020], which contains theoretical justification for using a DT for maintenance purposes. However, recent studies might empower the advantages of using a DT for maintenance. Bevilacqua et al. highlighted how a predictive maintenance DT prevents high-risk events for operators [Bevilacqua et al., 2020]. Aivaliotis et al. implemented degradation models in the DT that can be used to calculate the RUL of components [Aivaliotis et al., 2019]. This also improves decision-making and production planning [Aivaliotis et al., 2019].

In general, a DT used for maintenance can lower costs [Khajavi et al., 2019], improve the accuracy of predictions [Coraddu et al., 2019, Wang et al., 2019, Luo et al., 2019], and make the manufacturing system more sustainable, reliable, and efficient [Xu et al., 2019].

3.5.6 Cognitive Twins

Recent literature indicates a growing interest in combining ontologies with DTs to enhance their capabilities and interoperability. Boschert et al. coined the term "next generation Digital Twin" (nexDT) to describe a DT framework that enables the interconnectivity of systems through the use of semantic technologies such as ontologies [Rosen et al., 2019]. This concept highlights

the potential for DTs to go beyond their traditional boundaries by adopting advanced knowledge representation methods that facilitate system integration and communication.

Akroyd et al. further developed this idea by introducing the Universal Digital Twin concept, which employs a dynamic knowledge graph to enable cross-domain interoperability among DTs [Akroyd et al., 2021]. This approach allows for the seamless exchange of information between DTs across different domains, making it possible to integrate and leverage data from diverse sources, thus enhancing the overall functionality and adaptability of the DT framework.

Lu et al. presented the Cognitive Twin (CT) concept, which is characterised by augmented semantic capabilities that support the decision-making processes within manufacturing environments [Lu et al., 2020b]. The Cognitive Twin concept represents an evolution in DT technology, where ontologies are used to enrich the DT's ability to understand and interact with its environment, thereby improving its decision-making and predictive capabilities.

Li et al. implemented the Cognitive Twin concept in the context of co-simulation for complex engineering assets, combining various digital tools to achieve a more integrated and sophisticated simulation environment [Li et al., 2021]. Their work demonstrates how CTs can be applied to manage the complexity of modern engineering systems by providing a more comprehensive and interconnected approach to simulation.

Abburu et al. took this a step further by developing the COGNITWIN software toolbox for the process industry. They view the Cognitive Twin as an evolution of the Hybrid Digital Twin (HT), which itself is considered an advancement over the standard DT model [Abburu et al., 2020]. The COGNITWIN toolbox exemplifies how CTs can be tailored to specific industries, offering tools and frameworks that enhance the cognitive capabilities of DTs within the process industry.

Despite the relevance of the Cognitive Twin (CT) concept to this research, it was not included in the systematic analysis due to its relatively recent emergence in the academic literature. The term "Cognitive Twin" only began appearing in Scopus-indexed papers from 2020 onwards. At the time of this literature review, only five papers have been published in the broad field of engineering that specifically address CTs [Li et al., 2021, Lu et al., 2020b, Rožanec et al., 2020, Rožanec et al., 2021, Somers et al., 2020]. This limited body of work reflects the novelty of the concept, suggesting that while CTs hold significant potential, their practical applications and theoretical foundations are still in the early stages of development.

3.6 Discussion and conclusion

This paper presents a thematic review and characterisation of DTs in terms of technology used, applications, and limitations in the context of maintenance and degradation.

Regarding the first research question: *How are DTs for the maintenance phase of the life cycle developed?* Conducting the SLR enabled a systematic and reproducible methodology to

classify and review information about DTs related to the maintenance field. Randall divided maintenance into three phases of the Prognostics and Health Monitoring (PHM) process: detection; diagnostics, and prognostics. A DT should be able to satisfy each of these PHM steps [Randall, 2011].

Today's maintenance issues can be summarised in two main categories. The first issue is that maintenance today is mostly reactive, reaction once a failure occurs, or planned preventive when periodic maintenance is planned to prevent failures. This causes unexpected downtimes and unnecessary maintenance. The second issue is that departments within the same supply chain and even within the same company are siloed. They do not share data and information across the supply chain. This will cause a loss of important information and knowledge about the system. Early detection of faults, e.g., cracks or corrosion, can significantly reduce the through-life maintenance costs of complex systems and may help prevent unexpected breakdowns. This can be achieved using the DT paradigm. Using a DT to monitor the health of those systems will help the transition from time-based maintenance to condition-based maintenance, avoiding the waste of working components or waste of time. DTs have always existed, but they were just called by different names. The novelty of the DT concept is that it breaks through the walls of each department where information has been siloed so far.

The second research question: *How can a DT for maintenance be characterised?* The last subsection of the Results section presents a characterisation of a DT for degradation assessment. It should be possible to integrate a DT into a system-of-system scenario. This particular scenario identifies the federation of twins. This can be achieved by implementing a common semantic approach over the network, such as adopting a single top-level ontology. A DT must rely on a solid structure and contain modelling and simulation tools that use artificial intelligence (AI) algorithms to predict the remaining useful life (RUL) of the components within a system. A DT can collect data in real-time or when needed, depending on the operation of the asset. Assets that are not always connected to a network can discretely exchange information. A DT can have the computational power either on-site (edge computing), or off-site (e.g., cloud computing).

Of high importance is the capability to be interconnected in a network of twins, enabling, for instance, the DT integration across different sectors and stages within a supply chain. Emphasis has been given to this interconnection capability, which enables the real added value of using the DT paradigm. The same allows medicine to progress today through the worldwide adoption of gene ontology (GO) or the development of the internet through the adoption of the Semantic Web.

The limitation of this review is due to the non-standardisation of this technology. The search string can be improved to include other studies that do not use one of the words used for this article. Research projects might still approach similar topics, naming DT differently (e.g., virtual twin, digital thread, digital replica, surrogate, etc.). Additionally, the degradation topic

is broad; research projects may consider degradation topics without using the general words used in this review. However, several initial trials have been performed before reaching this search string, considering the string used, and its suitability for this review.

As previously anticipated, in the last two decades, the medical industry has made leaps and bounds, and part of the credit goes to adopting the ontological approach, particularly the top-level ontology (TLO). In fact, the creation of the Open Biomedical Ontology (OBO) foundry in 2002 has significantly impacted the industry as a whole. The OBO Foundry is a collection of all the biomedical ontologies, including GO, based on the same reference TLO, which is nothing but the BFO itself [Smith et al., 2007]. This approach improved the interconnection of different ontologies by providing a domain-neutral TLO as a starting point to develop domain-specific ontologies. This also allowed the reuse and expansion of existing ontologies instead of creating them from scratch without following any structured methodology.

In 2016, at the National Institute of Standard Technologies (NIST), a mixed group of researchers and industry experts created the Industrial Ontology Foundry (IOF) that reproduces the idea of the OBO foundry for the manufacturing sector [Montavon et al., 2020]. They also selected BFO as the reference TLO for the IOF. BFO is currently in the final phase to become an ISO standard, ISO/IEC 21838-2.

Future works will focus on investigating the TLO for the development of DTs, and evaluating the advantages and limitations of this approach. Exploring the opportunities that could give a structured DT framework, such as the one proposed in the previous section of this review, could potentially provide an opportunity for manufacturing industries to replicate the same successes that this approach brought to the biomedical sector. Developing structured DTs based on a standardised TLO, such as BFO, might improve the interconnections, enabling the creation of a network of twins or a federation of twins. The structured DT we presented in this paper can also be considered a cognitive twin, which we are going to develop using BFO as TLO. Other examples in research, such as Rozanec et al. presented the model of actionable CT based on BFO [Rožanec et al., 2021]. The network of twins might provide a more extensive training dataset for the artificial intelligence installed in the twins, which can improve the experience and, hence, the predictions of the twins.

Moreover, in a future human-machine interaction scenario, where both humans and machines will have a DT, it is envisaged that the human DT will likely be based on BFO, a TLO. The review also shows strong evidence that BFO is already the reference TLO for the biomedicine sector. Therefore, a machine CT based on BFO will easily exchange information with a human CT and vice versa. In general, the TLO-based CT could be the answer to the need, highlighted in the introduction, *"to have a unique digital framework able to break down silos"*.

Chapter 4

Industrial review

Industrial insights on digital twins in manufacturing: application landscape, current practices, and future needs

Published as Rosario Davide D'Amico, Sri Addepalli, and John Ahmet Erkoyuncu. 2023. "Industrial Insights on Digital Twins in Manufacturing: Application Landscape, Current Practices, and Future Needs" Big Data and Cognitive Computing 7, no. 3: 126. https://doi.org/10.3390/bdcc7030126

Abstract

The digital twin (DT) research field is experiencing rapid expansion, yet the research in industrial practices in this area remains poorly understood. This paper aims to address this knowledge gap by sharing feedback and future requirements from the manufacturing industry. The methodology employed in this study involves an examination of a survey that received 99 responses, and interviews with 14 experts from 10 prominent UK organisations, most of which operate in the defence sector. The survey and interviews explored DT design, return on investment, drivers, inhibitors, and future directions for DT development in manufacturing. The study findings indicate that DTs should possess adaptability, scalability, interoperability, and the ability to support assets throughout their life cycle. On average, completed DT projects reach the break-even point in less than two years. The primary motivators behind DT development were autonomy, customer satisfaction, safety, awareness, optimisation, and sustainability. Meanwhile, the main obstacles include a lack of expertise, funding, and interoperability. The study concludes that the federation of twins and a paradigm shift in industrial thinking are essential components for the future of DT development.

Keywords

Digital twins; federation of twins; industrial current practices; interviews; survey

4.1 Introduction

This study aims to examine the current state of DT usage in the industry through a mixed-methods exploration of DTs' applications, drivers, inhibitors, properties, and future needs. Data was gathered from UK industry experts through a combination of an online survey and interviews, with a focus on the application areas, development process, investment requirements, lessons learnt, and future needs for DT development. The three research questions guiding this study are:

1. What is the application landscape of DT in manufacturing?
2. What are the current practices in manufacturing?
3. What are the future needs?

The main contribution to knowledge is an insight into the current industry practices in adopting the DT paradigm and future steps and requirements. This paper provides experts' points of view on the potential impacts of DT utilisation. The article also provides a summary of challenges/barriers in adopting DT and how to overcome those, as well as a road map for adopting DTs.

The authors focused on a broad catchment area with the survey. At this point, gathering information from as many companies as possible and mapping out the present landscape of DT applications was important. The authors aimed to go deeper into understanding the DTs' features, drives, inhibitors, and future requirements when conducting the interviews. In the first phase, the survey was utilised to determine participants who were involved with DTs to capture their general understanding of the application landscape and the state of the project. As part of the survey, participants were invited to a second interview phase to capture specific knowledge on the DT projects they were associated with. Thus, for the second interview phase, only participants who agreed to disclose further information were involved in a semi-structured interview. This paper presents both the survey outcomes and the information captured during the interview phase. Care was taken to keep the information generic due to the commercial sensitivity associated with the information shared. It must also be noted that the analysis and opinions presented do not represent any specific organisation and are based on the authors' interpretation.

The methodology and the results have been divided into two sections, one dedicated to the online survey and one dedicated to the interviews, to separate and compare the results obtained from each data collection method. Therefore, the remainder of the paper is structured as

follows: Section 4.2 presents a literature review, Section 4.3 presents the online survey methodology and related results, Section 4.4 presents the methodology and results of the interviews, Section 4.5 presents the discussion, and Section 4.6 the conclusions, including limitations and future works.

4.2 Background and literature review

In this section, we report on the usage of DT in both industry and academia. In section 4.2.1, an analysis of the current industrial reports is presented, while in section 4.2.2, a review of the academic literature along with the current gaps is reported.

4.2.1 Industrial reports

With an increasing number of companies designing a digital workforce in manufacturing, Deloitte has provided some general guidelines outlining that four components are necessary for creating a framework for a digital workforce. These include collaboration, technology, governance, and value [Wright and Davidson, 2020]. Some important actions that are shaping the future of the digital workforce are:

- Reskilling the team
- Hiring individuals with the necessary skills
- Building an environment that attracts talent
- Retaining that talent

Although the study of DTs in infrastructure is still in its infancy, substantial investments and efforts are being made worldwide, including in the UK, to establish DT ecosystems. To implement the recommendations from the National Infrastructure Commission's 2017 "Data for the Public Good Report," for instance, the Centre for Digital Built Britain created the National Digital Twin Project (https://www.cdbb.cam.ac.uk/what-we-did/national-digital-twin-programme) [National Infrastructure Commission, 2018].

Accenture, in a report of 2022, states that manufacturing executives who successfully implement this next-generation manufacturing systems architecture will fundamentally outperform their competitors by increasing productivity, raising customer happiness, and lowering costs [Brosset et al., 2022]. Based on another analysis, between now and 2030, DTs can save CO2 emissions by 7.5 Gt and add up to $1.3 trillion in economic value [Verzelen et al., 2021].

ISO 23247:2021 is a set of standards for developing and using DTs in manufacturing. It provides a framework for defining, developing, and deploying DTs and some functional entities that DTs support. The standard is divided into four parts with the following scopes: i)

ISO 23247-1: General principles and requirements for developing DTs in manufacturing; ii) ISO 23247-2: Reference architecture with functional views; iii) ISO 23247-3: List of basic information attributes for the observable manufacturing elements; iv) ISO 23247-4: Technical requirements for information exchange between entities within the reference architecture (International Organisation for Standardisation [ISO], 2021).

ISO 23247 defines the DT as a *"(manufacturing) fit for purpose digital representation of an observable manufacturing element with synchronisation between the element and its digital representation"*. In this definition, digital representation is defined as a *"(manufacturing) data element representing a set of properties of an observable manufacturing element"*, and an observable manufacturing element (OME) is an *"item that has an observable physical presence or operation in manufacturing"*, such as personnel, equipment, material, process, facility, environment, product, and supporting documentation (International Organisation for Standardisation [ISO], 2021). The standard is intended for manufacturers to develop and deploy DTs to improve manufacturing processes. It is also intended for researchers and practitioners to develop new and innovative applications for DTs.

The AMRC report [Eyre et al., 2020] "Untangling the Requirements of a Digital Twin" provides a comprehensive overview of the requirements for developing and deploying DTs. The report highlights some quantitative benefits of the implementation of the DT in the industry, such as:

- 50% reduction in unplanned downtime (€40,000/minute by a German Automotive Industry).
- 40% reduction in maintenance costs (on a forging line proactively fixing saves $200,000).
- 1%-3% reduction of capital equipment costs.
- 5%-10% reduction in energy cost.

The AMRC report also highlights some gaps in the current state of DT development and deployment. These gaps include:

- Lack of a unified definition of DT. There is no single, agreed-upon definition of DT. This can make communicating about DTs and comparing different DT development and deployment approaches challenging.
- Lack of a systematic approach to DT development. There is no single, agreed-upon approach to developing DTs. This can make it challenging to develop DTs robustly and reliably.
- Lack of empirical evidence on the benefits of DTs. There is limited empirical evidence on the benefits of DTs. This makes it challenging to assess DTs' value and justify the investment in DT development and deployment.

- Lack of standards for DT development and deployment. There are no standards for DT development and deployment. This can make it challenging to develop and deploy DTs consistently.

- Lack of skilled personnel. There is a shortage of skilled personnel with the knowledge and experience to develop and deploy DTs. This can make it challenging to develop and deploy DTs in a timely and cost-effective way.

4.2.2 Academic literature review

DT definitions, features, applications, and design ramifications were thoroughly investigated by Barricelli et al. [Barricelli et al., 2019]. As of 2019, the authors looked at 75 articles from various disciplines, including manufacturing, precision medicine, hospital administration, and aviation. It must be noted that the study did not report on DTs for infrastructure. DTs were also the subject of literature reviews by Trauer et al. and Jones et al., which looked at a total of 57 and 92 publications, respectively [Trauer et al., 2020, Jones et al., 2020]. These studies covered DTs' definitions and traits but did not directly examine whether they were accepted in business, particularly the infrastructure sector. DT definition, functional components and differences between DT and cyber-physical systems are the subjects of the paper submitted by Boyes and Watson [Boyes and Watson, 2022]. In their study, information management, architecture and design, life cycle, and safety and security, are the areas that need further research. DT definition, application context, and research challenges are the subjects of the systematic review published by Semeraro et al. [Semeraro et al., 2021].

Despite the growing interest in DTs and their potential benefits, there is still a lack of understanding of the current industrial practices of DTs [Boyes and Watson, 2022]. This lack of understanding is also reflected in the difficulties that companies face when designing and implementing DTs. Many companies are struggling to find standardisation and guidance for designing DTs, which can lead to errors and inefficiencies in the process. This paper is an initial attempt to fill this gap, providing the current application landscape and the future needs from an industrial point of view.

Biesinger et al. argue that the benefits of a DT outweigh the costs [Biesinger et al., 2019]. The article argues that automotive manufacturers should consider implementing a DT to improve efficiency, reduce costs, improve quality, and increase flexibility. The article identifies several gaps in the literature on DTs. First, there is a lack of research on the specific benefits that DTs can provide for different types of businesses. Second, there is a lack of research on businesses' specific challenges when implementing DTs. Third, there is a lack of research on the specific strategies that businesses can use to overcome the challenges of implementing DTs. Meanwhile, Broo et al. also highlight the lack of a common set of standards for DTs [Broo et al., 2021]. There is currently no common set of standards for DTs, which makes it

difficult to share data and information between different DT implementations.

Other studies covered various aspects of the industry application of DT technologies. Ferko et al. conducted a literature review article including a survey with 33 respondents and four semi-structured interviews to investigate the current alignment between the ISO 23247 and DT software architectures in industries [Ferko et al., 2023]. The study highlighted how the adoption of standard architectures in the industry is still in the early stages and remains a gap in interoperability and maintainability. Liu et al. conducted semi-structured interviews with six companies for an industrial symbiosis network to analyse the needs of digital twins in supply chain collaboration, finding that large companies favour customised digital twins with an emphasis on environmental impact, while smaller companies prefer standardised digital twins focusing on cost and personnel reduction, leading to the development of an inter-company emission-monitoring framework and a general digital twin framework for supply chain collaboration [Liu et al., 2022].

Durão et al. analysed 19 articles and interviewed industry experts to identify the main criteria for developing DTs and industry understanding of the concept, highlighting that real-time data, integration, and fidelity are the key requirements in literature [Durão et al., 2018]. While the industry views DT as a coexistence of digital and physical worlds with a focus on simulation models, emerging research trends suggest a shift towards system independence and implementation over DT design requirements. Xie et al. presented a DT four-dimension (4D) fusion modelling method to solve issues in discrete manufacturing lines, such as low model accuracy, response delay, and insufficient production line control accuracy [Xie and Wan, 2023].

According to Timperi et al., a DT can substantially impact the business models of manufacturing organisations [Timperi et al., 2023]. DTs may improve operations, provide cost savings and business growth, and enable stakeholders to concentrate on core strengths while growing their organisations. Several barriers to exploiting DTs have been highlighted in the study, including data ownership, resource allocation, internal bureaucracy, and the difficulty of showing the actual value of data-based services to prospective consumers.

Both in industry and academia, studies need to outline insights about DT applications from numerous perspectives, such as areas of use, benefits achieved, challenges experienced, and development methods. This paper represents a first attempt to address these gaps by analysing the results of a detailed industrial survey and interviews with experienced stakeholders. Over 45 organisations have been involved in this research to share their point of view on DTs. Capturing and collating industrialist insights presents valuable guidance on current and future research needs.

4.3 Online survey

This section presents the methodology and the results of the online survey. The online survey was conducted in a series of workshops by industry experts from several companies.

4.3.1 Online survey methodology

The online survey mapped the current application landscape of DTs in manufacturing, primarily in the UK. The survey was conducted online with 16 questions, 12 requiring short answers or multiple-choice responses. The authors used the Forms tool within the Office 365 suite to collect the data. The results were then analysed using MS Excel and elaborated in a .xlsx spreadsheet. The questions (see Appendix) were grouped into themes: the first six pertained to the DT project itself, the following five (questions 7-11) focused on the size of the project, status, and investment, and the final five (questions 12-16) were about the company, including contact information and industry sector. Overall, we found significant interest in DTs among participants in the manufacturing industry.

The survey has been promoted through a series of workshops delivered by the Team Defence Information (TD-Info) group (`https://www.teamdefence.info/`), within the TD-Info - Digital Twin Community of Practice (DefSp DTCOP). A working group on DT mapping was hosted by Prof. John Ahmet Erkoyuncu from Cranfield University (UK), and an online survey was launched on the 23rd of June 2021. 99 responses were collated from 49 organisations. The outcomes have been divided into four macro groups: i) challenges, ii) area of implementation, iii) maturity, and iv) impact. This approach allowed the authors to present the content logically so that the reader could interpret the study's key findings and how they relate to the research question.

4.3.2 Results from survey

The individuals responding to the survey typically occupied senior positions such as 'Director', 'C' level, 'Head of Unit', 'Manager', or 'VP'. The respondents' occupation level distribution is illustrated in Figure 4.1.

Among the interviewees, only 13% of the respondents were in junior-level positions. Interestingly, senior respondents were more aware of DTs and expressed interest in facilitating faster development and implementation of DTs.

Challenges

The challenges to be addressed where DTs have been initially considered have been divided into categories and subcategories, as reported in Table 4.1.

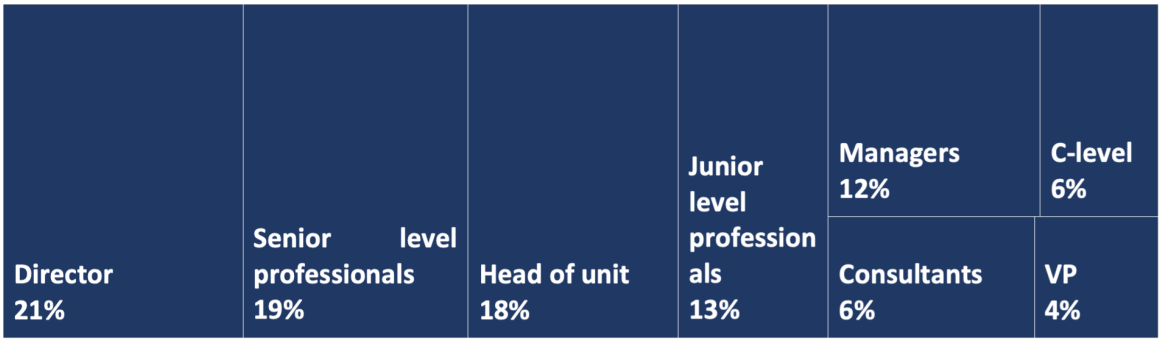


Figure 4.1: Distribution of job roles of respondents

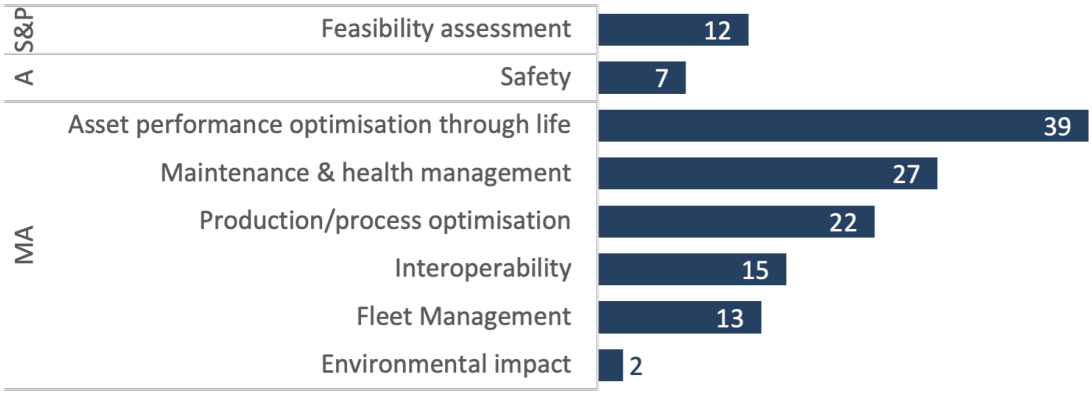


Figure 4.2: Challenges to address with DT adoption, where (MA) represents the category Managing Assets, (A) represents Assurance, and (S&P) represents Strategy and Planning

The result, reported in Figure 4.2, was derived from responses to the open-ended questions in the online survey. According to the survey respondents, the most significant challenge DT is being utilised to tackle is optimising asset performance through life, followed by maintenance and health management, and production/process optimisation. Additionally, the topic of interoperability was brought up during this phase, indicating that this matter is gradually garnering business attention.

Area of implementation

From the question *'What sector is your digital twin applied to?'*, the authors mapped the area of implementation of DTs in terms of industrial sectors. The question had multiple-choice answers and the users just needed to select the industrial sector(s). The defence, aviation, and maritime sectors are the primary sources for DT deployment, according to the companies involved in the survey, as illustrated in Figure 4.3.

From the question *'What areas are applicable to your digital twin?'*, the authors wanted to investigate the phases of the life cycle the DT was applied to.

The life cycle phases during operation, including in-service optimisation, maintenance, re-

Table 4.1: Categorisation of challenges derived from the answer to the online survey. Where (MA) represents the category Managing Assets, (A) represents Assurance, and (S&P) represents Strategy and Planning.

Categories	Subcategories	Definition
S&P	Feasibility assessment	In this field, the challenges regarding helping enterprises understand the digital maturity level and what tools they require include IT support.
A	Safety	In this field, the challenges regarding the safety of the system are included.
MA	Asset performance optimisation through life	In this field, the challenges regarding the optimisation of the asset performance throughout its life cycle including modelling and simulation to find optimal working operation; historic, current, or predicted conditions; performance validation are included.
MA	Maintenance & health management	In this field, the challenges regarding health monitoring; increasing availability; monitoring of the remaining useful life; management of resources to perform maintenance; degradation assessment; increasing the endurance of the asset are included.
MA	Production/process optimisation	In this field, the challenges regarding optimisation of the production processes, such as the design, development, and manufacturing processes, are included.
MA	Interoperability	In this field, the challenges regarding the interoperability of different systems, such as cross-department communication are included.
MA	Fleet Management	In this field, the challenges regarding the management of a fleet of assets are included.
MA	Environmental impact	In this field, the challenges regarding the monitoring of the environmental impact are included.

pair, and overhaul (MRO), and mid-life upgrade and modification, play a significant role in the implementation of DTs, as reported in Figure 4.4.

From the question *'What is the scale of the project?'*, the authors investigated the size of the DT projects. The distribution is almost evenly distributed between small, medium, and large projects, as reported in Figure 4.5.

From the question *'What is the complex system scope or focus?'* the authors investigated the focus area of the DT project.

The study highlights that across the proposed DT examples, there has been an emphasis on having a better understanding of the 'end-to-end capability' and 'processes' linked to the asset, as reported in Figure 4.6.

This emphasis on understanding the entire life cycle of the asset suggests that the goal of

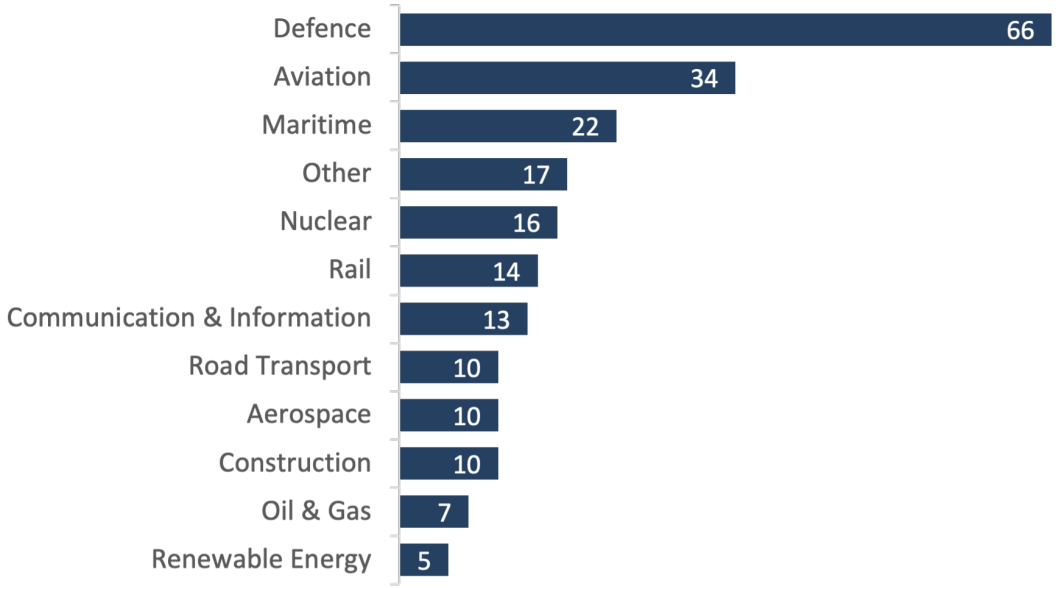


Figure 4.3: Sectoral representation of respondents.

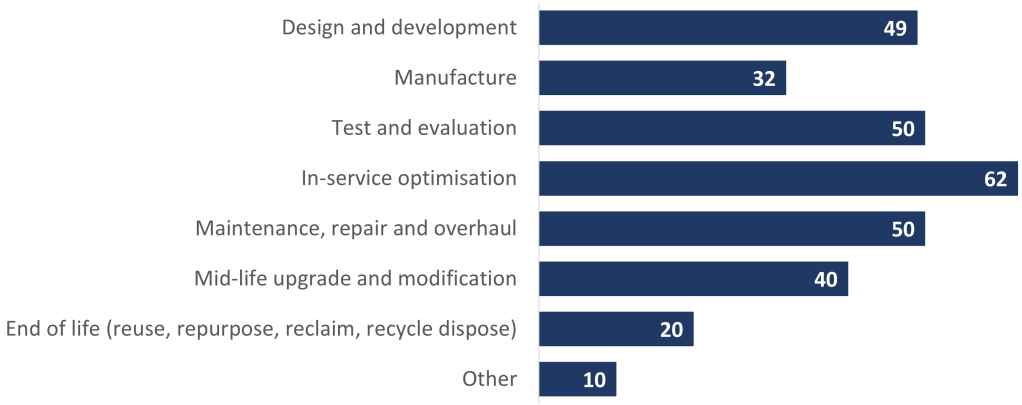


Figure 4.4: Interest in DT implementation across the life cycle.

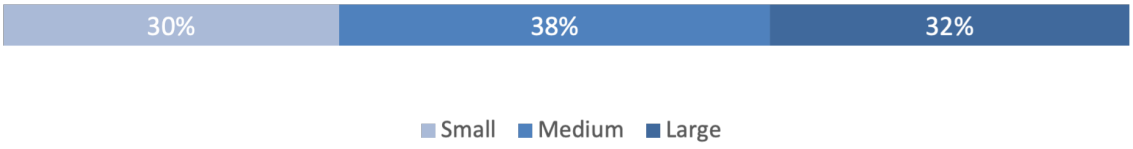


Figure 4.5: The size of the DT projects.

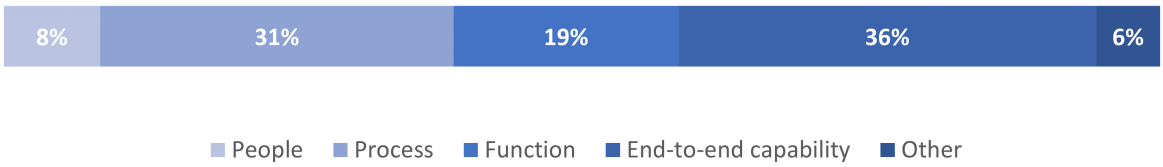


Figure 4.6: The focus area of DT.

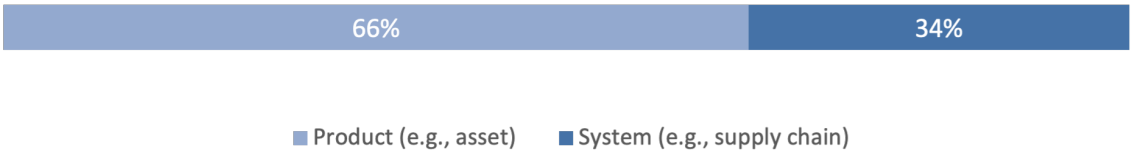


Figure 4.7: Purpose area of DT.

using DTs is to improve the asset and the outcomes that can be achieved from the asset.

With the question *'What is the theme of the digital twin application?'* The authors investigated whether DT is applied to products or systems. Results are reported in Figure 4.7.

The purpose of DT has primarily been centred on the product (e.g., asset) as declared by 66% of the survey respondents. Only 34% of the responses focused on the system (e.g., processes) that consisted of, among other things, the supply chain. This important message may indicate that DT has commonly been considered for products that generate commercial and performance gains for organisations. This could be because organisations can connect the commercial/performance gains accumulated from the asset and the added value that a DT can bring. This may have been more difficult to justify the value gathered from DTs for improving systems/processes. Moving forward, there will likely be a growing interest in enhancing systems/processes to improve outcomes that can be achieved through the product.

Maturity

From the question *'What is the status of your project?'*, the authors looked into the status of the DT projects of the companies taking part in the online survey. Results are reported in Figure 4.8.

The distinction here is between completed DT projects (with available outcomes); in-progress DT projects; scheduled DT projects (planned and funded); and aspirational DT projects (planned and unfunded). Interestingly, the combination of the shares of completed DT projects and in-progress projects together is more than 50% (about 68%), with the most significant portion (41%) dedicated to projects currently in progress. The remaining two portions are divided among planned, and scheduled projects, each making up approximately 16% of the chart, as reported in Figure 4.8.

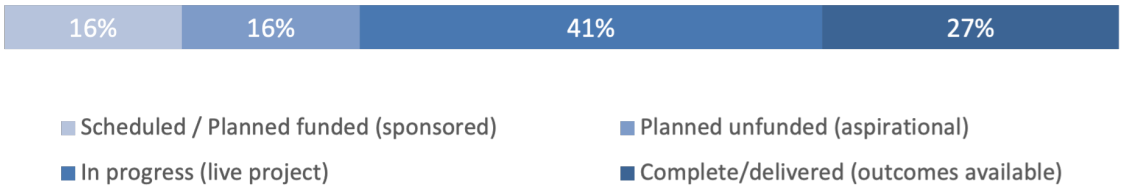


Figure 4.8: Status of the DT project.

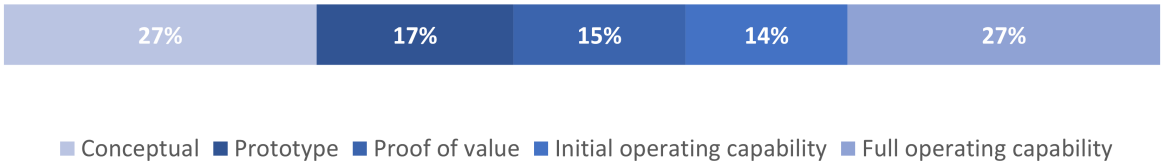


Figure 4.9: Level of maturity of DT projects.

Besides, with the question *'What level of maturity is your project?'*, the authors investigated the stage of the DT project. Specifically, the DT project maturity is considered across conceptual, prototype, proof of value, initial or full operating capability.

Almost one-third of the DT projects investigated have full operating capabilities (27%). With the same share, there are conceptual projects, which could enforce the premise that interest in this technology is growing and new projects have been considered by the companies involved, as reported in Figure 4.9.

Financial impact

Return On Investment (ROI) is a financial metric used to evaluate the efficiency of an investment. It is calculated by dividing the net benefit of an investment by the cost of the investment and expressing the result as a percentage. A positive ROI indicates a profit from the investment, while a negative ROI indicates a loss. It is commonly used to measure the performance of an investment over time and is expressed as a percentage or a ratio. It is used in many industries to assess the profitability and effectiveness of specific investments [Phillips et al., 2011].

What has been reported in Figure 4.10 is called the break-even point (BEP) in the literature. The break-even point is when the total revenue from a project or investment equals the total costs associated with that project or investment. In other words, it is the point at which the revenue generated by an investment or project is sufficient to cover all its associated costs, resulting in a zero net profit [Dayananda et al., 2002].

Results from the combination of question 10 *'What is the project start date?'* and question 11 *'What is the forecast return on investment (when will benefits in the business case be realised)?'* are reported in Figure 4.10. The forecasted return on investment is expressed in years

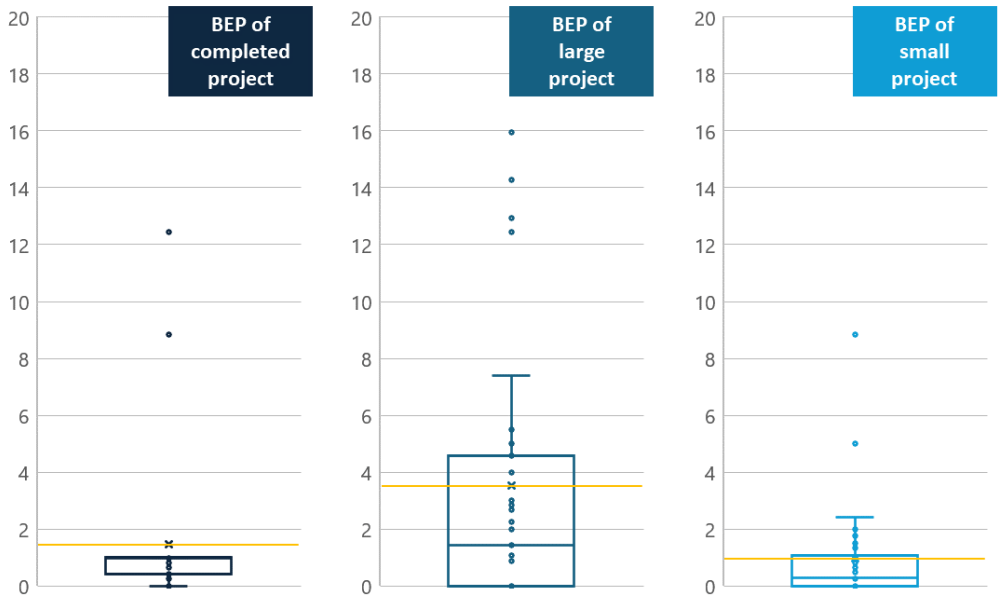


Figure 4.10: BEP (in years) of DT projects, divided between completed projects (left), large projects (centre), and small projects (right). In the abscissa the number of the project, while in the ordinate the time length in years.

and indicates the break-even point (BEP) of the DT projects. The three charts in the figure indicate the BEP in years for completed DT projects (on the left), for large DT projects (in the centre), and for small DT projects (on the right). The BEP was less than two years across the declared completed projects. Furthermore, the BEP average for large projects was just under four years, and around one year for small projects.

4.4 Interviews

This section presents the methodology and the results deduced from the interviews. The interviews were conducted based on the availability and preference selected from the online survey.

4.4.1 Interviews' methodology

Using the list of volunteers from the last survey question, the researchers contacted 14 experts from 10 UK-based companies. These experts agreed to participate in interviews, providing valuable insights into their DT projects, including requirements, lessons learned, and future scenarios. Initially, the questions were tested with lab members. Throughout the interview process, the questions were continually refined based on the information gathered to ensure a consistent set of answers for analysis.

The profile of the 14 experts from the 10 companies that participated in the interview is reported in Table 4.3. The questions evolved during the interview timeframe based on the

experience gained over the other interviews to have a homogeneous set of answers to analyse.

The interview strategy in this study was to use open-ended questions to gather as much detailed information as possible from the interviewees. The interviews aimed to understand the companies' perspectives on DTs and their approach. The questions were based on a semi-structured format, meaning they were open-ended but still followed a general structure. To avoid bias, the participants did not know the questions beforehand and were unaware of the answers provided by other participants, allowing for more objective data collection.

Thematic analysis is a method of analysing qualitative data, such as interview transcripts, that involves identifying and coding recurrent themes in the data. As described by Braun and Clarke [Braun and Clarke, 2006], the method is systematic, meaning that it is a step-by-step process that is applied consistently to all of the data. The steps of the thematic analysis, inspired by a paper by Braun and Clarke [Braun and Clarke, 2021], include initial familiarisation with the data, development of a coding scheme, identification of themes, and interpretation of the themes, as reported in Table 4.2. The specific steps may vary depending on the study and the research question being addressed. Thematic analysis can be useful in identifying patterns and themes in data, which can help to guide further research and analysis.

Table 4.2: Steps of the thematic analysis.

Step	**Description of the process**	**Application to this work**
Familiarising with data	Transcribing the interviews, reading and re-reading the transcripts, and noting down initial ideas.	Manual transcription of the recorded interviews.
Generating initial codes	Coding interesting features of the interviews in a systematic fashion across the entire data set, collating data relevant to each code.	Find common codes among all the interviews. Creation of codes in the NVivo tool.
Searching for themes	Collating codes into potential themes, gathering all data relevant to each potential theme.	Grouping codes based on themes and creation of themes.
Reviewing themes	Checking if the themes work in relation to the coded extracts and the entire data set.	Validate the themes by checking the codes for each interview.
Defining and naming themes	Generating clear definitions and names for each theme.	See Table 4.4.
Producing the report	Reporting the results obtained in a paper.	Writing the final paper.

Table 4.3: Profiles of interview meetings (online and 60 mins duration) and participants' details.

Agent	Company	Location	Date	Experience	Role	Industry
A00	Co_01	UK	14/09/2021	42 Years	IT manager	European multinational aerospace corporation
A01	Co_02	UK	21/09/2021	21 Years	Head of Digital	European multinational company for the aerospace, defence, transportation and security sectors
A02	Co_03	UK	24/09/2021	32 Years	Experimental and MRO Functional Lead	Aerospace design and manufacturing
A03	Co_03	UK	24/09/2021	13 Years	Head of Innovation and New Product Design	Aerospace design and manufacturing
A04	Co_04	UK	24/09/2021	14 Years	Commanding officer	British government department
A05	Co_05	UK	30/09/2021	10 Years	Propositions Manager	Multinational logistics, supply chain, manufacturing and consultancy company
A06	Co_05	UK	30/09/2021	26 Years	Chief of Staff and Enterprise Architect	Multinational logistics, supply chain, manufacturing and consultancy company
A07	Co_05	UK	30/09/2021	23 Years	Head of Solutions	Multinational logistics, supply chain, manufacturing and consultancy company
A08	Co_06	UK	12/10/2021	37 Years	Partner	Engineering and technology consultancy company
A09	Co_07	UK	13/10/2021	14 Years	Systems Engineering Information Technologist	Multinational Aerospace, Arms industry, Information security company
A10	Co_07	UK	13/10/2021	40 Years	Engineering Manager	Multinational Aerospace, Arms industry, Information security company
A11	Co_07	UK	13/10/2021	12 Years	Service Integration Manager	Multinational Aerospace, Arms industry, Information security company
A12	Co_08	UK	19/10/2021	40 Years	Head Of BI	Multinational Aerospace, Defence
A13	Co_09	UK	19/10/2021	22 Years	Head of Data & Analytics	IT consulting and services
A14	Co_10	UK	21/10/2021	25 Years	Head of Digital Integration	Multinational Aerospace and Defence company

Table 4.4: Themes resulted from the thematic analysis.

Theme	Description
Drivers & inhibitors	This theme gathers the factors that drive or interfere with DT development. These factors include the benefits, the challenges that the DT undertake, the barriers, and the problems.
Digital twin properties	This theme gathers the factors that characterise the DT. These factors include the attributes, the elements that constitute the DT, the semantic properties, and the visualisation properties.
Finance	This theme gathers the factors that characterise the economic aspect of the DT. These factors include the return on investment and the funding used to start a DT project.
Next steps & needs	This theme gathers the factors that provide a future vision in the DT context. These factors include the future scenario, the suggestions, the needs, and the next steps.

4.4.2 Results from the interviews

The ten interviews were recorded, and their transcripts were manually created in ten separate Microsoft Word documents. In three of these interviews, we hosted more than one expert, resulting in a total of fourteen interviewees across the ten interviews. The authors opted not to use any speech-to-text tools for transcription to become more familiar with the content of each interview. The transcripts were then formatted for use in NVivo 12, a software tool for coding and analysing qualitative data. For this research, a university (site-level) license was used to access NVivo.

As the first step, each interview transcript was coded and the codes were grouped by a predefined subject based on the information provided during the interview. The subjects were then grouped with similar subjects to form themes. This process allowed the researcher to identify patterns and commonalities in the data, which was then used to identify key themes and insights. The use of NVivo software in this process allowed for a more systematic and efficient way of coding and analysing data.

The coding phase in qualitative research is a crucial step in the process of organising, analysing, and making sense of data collected from interviews. The process typically begins with transcribing the interviews. This involves taking the recorded audio or video of the interview and writing down precisely what was said by the participants. Transcribing the interviews allowed the researcher to have a written record of the data that can be easily read and analysed. Once the interviews had been transcribed, the researcher began the process of reading through the answers provided by the participants. This is typically done multiple times in order to identify key themes or concepts that emerge from the data.

After identifying key themes or concepts, the researcher summarises them into a single sentence or ”code” representing the expressed main idea. This allowed the researcher to analyse

and understand the data and identify response patterns or trends. This process is called thematic analysis. The four themes that have been extracted from the data in this study are drivers & inhibitors, DT properties, finance, and next steps & needs. These themes represent the key concepts or patterns that emerged from the data collected from the interviews with experts from different companies. The themes are reported in Table 4.4, which provides more detailed information on each theme and how they were identified from the data.

The tables (Table 4.7 to Table 4.12) include codes (listed in alphabetical order) that have been extracted from the transcripts of the interviews and grouped into themes. The tables also include additional information, such as the source and reference for each code.

- The *'N. interviews'* column in the table indicates the likely number of interviews in which the specific code was mentioned. This allows the reader to see how frequently a particular driver or inhibitor was mentioned across the interviews.

- The *'Count'* column in the table indicates how many times a specific code was mentioned within the interviews. This allows the reader to see how important a particular driver or inhibitor was to the participants.

Drivers & Inhibitors

This subsection focuses on the theme of drivers and inhibitors, the results are presented in two separate tables, one for drivers and one for inhibitors. The tables allow the reader to see the specific drivers and inhibitors that were identified in the study and how frequently they were mentioned. This information provides insight into the benefits and barriers that companies are experiencing when implementing DTs and can be used to understand the current state of DT implementation in the industry.

Various advantages associated with the DT development process categorised into themes based on the codes derived from the interview transcripts, are reported in Table 4.5. These benefits include enhancing the knowledge base and awareness of the twinned asset, as well as the ability of the DT to transition the workload from the physical to the virtual world, providing more flexibility and autonomy for digital engineering tasks. The table also indicates that a DT brings about customer satisfaction. It is noted that DTs facilitate the integration of different systems under the ”system of systems” paradigm, and also contribute to improving a company's overall efficiency. The development of DTs also allows for the optimisation of costs and performance of assets, leads to better safety outcomes, encourages companies to simplify complexity, and supports overall sustainability.

An extract from the interview, Agent A12: *“The opportunity the DT provide us is to make that much more dynamic and real-time and lead to more true predictive maintenance, and better data coming out”*.

Inhibitors to the DT development process, most notably ”interoperability issues” and ”company’s mindset” are reported in Table 4.6. ”Interoperability issues” refer to the lack of standardised approaches in data and information sharing, which can hinder effective communication between different DT systems. ”Company’s mindset” relates to the reluctance to adopt innovative technologies due to decision-makers’ skill gaps or lack of understanding. The data in the table also points out other inhibitors like ”complexity” of managing complex tools, ”quality of data” referring to poor input data, ”uncertainty” indicating discrepancies between data and physical models, and ”expense” as a financial constraint, particularly in the case of large companies initiating a pilot DT project.

An extract from the interview, Agent A08: *“Dealing with uncertainty is a difficult part of the DT. In the work that we have done, we generally have data and a physical model, and we deal with some sort of uncertainty. Bringing those three things together is what gives you something that allows you to make decisions which you can trust. As your physical model gets better, or as the available data gets better, the uncertainty can reduce, and so you are not dependent so much on the sort of math too”*.

Table 4.5: Table of DT drivers.

Drivers	**N. interviews**	**Count**
Autonomy	1	1
Customer satisfaction	1	1
Enable system of systems	1	1
Improve the company’s efficiency	2	2
Improve knowledge and awareness	3	4
Optimisation	2	4
Safety	2	2
Simplify complexity	1	1
Sustainability	1	1

Table 4.6: Table of inhibitors.

Inhibitors	**N. interviews**	**Count**
Company’s mindset, lack of skills, and understanding	4	9
Complexity	4	7
Expensive	1	1
Interoperability issue	8	17
Quality of data	2	3
Uncertainty	2	2

DT properties

This theme provides an in-depth look into the different factors that make up a DT and the processes used to create and interpret them. This information helps understand the current state

of DTs and can be used to improve the design, visualisation, and interpretation of DTs.

The main attribute mentioned in the interview is the set composed of ”portability, scalability and interoperability”, as reported in Table 4.7. ”Portability” refers to the ability of a DT to be transferred or adapted to different platforms, environments, or systems without loss of functionality or performance. This attribute is important for companies as it allows flexibility in using the DT in different situations or contexts. ”Scalability” refers to the ability of a DT to adapt or grow to accommodate changes in the volume or complexity of the data or asset being twinned. This attribute is important for companies as it allows the DT to be used in various applications and handle different complexity levels. ”Interoperability” refers to the ability of a DT to communicate and share data with other systems, platforms, or DTs. This attribute is important for companies as it allows the DT to be integrated into existing systems and share data across different departments or organisations.

The second attribute reported in the table is that a DT does not have to work “necessarily in real-time”. Depending on the purpose and the asset, discrete communication might also work. For example, suppose a DT is used to monitor and analyse the performance of a machine or equipment. In that case, it may only need to update at specific intervals or when certain conditions are met.

The third attribute reported in the table is that DT has to “support the twinned asset through life”; it is not specifically for one part of the life cycle. This means that a DT should be able to support the asset throughout its entire life cycle, from design to disposal, and not just a specific part of it. This is important as it allows companies to gain insights and make decisions throughout the asset’s life cycle, not just during specific stages.

An extract from the interview, Agent A01: *“There is an attribute evaluation during the design process. Attributes that we consider would be portability, scalability and interoperability. So different DT solutions will require to focus on different attributes”*.

A defined procedure for DT design is not universally adopted, with diversity in the design team being highlighted as crucial for innovation and comprehensiveness, as reported in Table 4.8. It has shown that a blend of experienced operators and newly graduated data scientists adds value to the DT design, allowing for diverse expertise to inform the process. The data reported in the table underscores the importance of a clear purpose as the starting point of the DT design process, alongside other strategies such as initiating design through a research project. Other factors like a focus on simulation, safety, and DT for data validation, which refers to the use of DT for verifying collected asset data, are also mentioned.

An extract from the interview, Agent A03: *“We did not apply all this modern AI technology to our engine specifically for DT development, but to optimise operation and costs, and because we have created this kind of digital thread across our organisation. Talking about the DT hierarchy, people do not often know where to start to become a DT organisation and I always say, you just start everywhere. If you do not forecast your fleet, start forecasting your fleet, and*

start creating models. If you do not have digital inspection across everything, including MRO, start implementing it. If you do not look at how to manage individual components through your supply chain, there will be a value proposition associated with just that".

From the interviews, a strong emphasis emerged on the significance of visualisation in the context of value and data in the DT design, rather than focusing on visualisation aesthetics alone, as reported in Table 4.9. "Focus on value," implies that visualisation should facilitate decision-making by providing simplified, standardised, and graphically represented information. "Focus on data," underscores the need to trust the data and the analytical abilities of the DT over the final product. As for the visualisation tools, extended reality technologies (XR) - augmented reality (AR), virtual reality (VR), and mixed reality (MR) - are increasingly popular for DT visualisation, while dashboards remain useful for presenting decision-making information. This insight portrays how interviewees perceive the role of visualisation in DT design and the elements they deem significant.

An extract from the interview, Agent A14: *"I think building up that trust in the data in there and the analytics capability, I think is key. I think the other aspect is being able to visualise the data in the right context, dependent on the user, and ensure that what the data is projecting or saying is truly understood and interpreted in the right way. I have seen examples of dashboards being created to almost tell the story that somebody wants to tell as opposed to undertaking analysis of the data to determine what could be done or should be done"*.

Table 4.7: Table of DT attributes.

Attributes	**N. interviews**	**Count**
Portability, scalability, and interoperability	3	3
Not necessarily real-time	1	1
Through life support of the real entity	1	1

Table 4.8: Table of DT design.

DT design	**N. interviews**	**Count**
Defined procedure	4	4
Diversity of the team	7	9
DT for data validation	1	1
Focus on safety	1	1
Focus on simulation	1	2
No formal procedure	5	5
Start from a research problem	1	2
Start from the purpose	6	8

A range of awareness and application of semantic approaches in the DT design process has emerged, as reported in Table 4.10. The table shows that a significant number of interviewees (nine out of twelve) employ standard approaches in their DT design. These standard approaches include the use of standard ontologies, open-source solutions, or a shared language across

Table 4.9: Table of DT visualisation.

Visualisation	N. interviews	Count
Dashboards	1	2
Focus on data	2	2
Focus on value	7	13
Headsets	3	4

devices. The high count for standard approaches suggests that there is a prevalent trend towards seeking uniformity and interoperability in DT design through established methodologies.

Additionally, three interviewees specifically mentioned the use of ontologies. Ontologies in this context refer to structured frameworks that define the relationships between different concepts within the DT, facilitating a more semantic and interoperable design.

Conversely, one interviewee was not aware of semantic approaches, indicating a gap in knowledge or exposure to these methodologies. Moreover, two participants considered semantics in DT design to still be a research topic lacking a widely recognised standard approach. This highlights discrepancies in understanding and utilising semantic strategies, suggesting that while some practitioners are integrating these methods effectively, others see them as nascent or not yet fully standardised.

These variations underscore the need for further research and standardisation in semantic approaches within DT design to bridge the gap between research and practical application.

An extract from the interview, Agent A08: *"I think the modularisation of ontologies, from an application level to a top-level one is very important, mostly in the space domain. I am less familiar with the temporal side, which I think is a trickier problem. The other problem is how you deal with uncertainty within the context of ontologies"*.

Table 4.10: Table of DT semantics.

Semantics	N. interviews	Count
Not aware	1	1
Ontology	3	3
Standard approach	6	9
Still research topic	2	2

Finance

The focus of this part of the interview regarded the length of time it takes for a company to see a return on their investment in a DT project. A diversity in breakeven points and ROI estimates for DT projects among interviewees, with some expecting a significant ROI or a short breakeven point, while others do not consider ROI in their projects, as reported in Table 4.11. Qualitative insights reveal that several interviewees view DT as an opportunity for uncovering and addressing inefficiencies, optimising processes, and reducing time-to-market, rather than

purely an investment. The breakeven point can vary based on several factors, including industry, company size, and the type of DT project.

An extract from the interview, Agent A04: *"If the DT is working well, you will not necessarily see where it is saving money. You are dealing with an issue before it becomes an issue, so therefore, you are not necessarily having to invest money to solve a problem later down the line, because you are solving it before it becomes a problem. So, it might be difficult to identify that kind of return on investment because it might look like it is quite a process rather than financial reward or a financial benefit"*.

Table 4.11: Results from the return-on-investment questions.

Return of investment	**N. interviews**	**Count**
2-year ROI	1	1
30%-40%	1	1
Accept longer period returns	1	1
Opportunity benefit rather than investment	4	6
Optimise to not waste money later	3	4
Reduce time-to-market	1	1
ROI not considered	2	2
Significant return on the investment	2	3
Usually short	4	4

Next steps and needs

One critical progression identified in this part of the interview is to establish a "system of systems" approach. This aspiration aims for a network of interconnected and semantically interoperable DTs, enhancing the overall knowledge of this network. To achieve this, various methods were proposed, including the use of ontologies, shared language approaches, and standard methods.

Another significant need is a fundamental change in the company's mindset and reskilling of their teams. This change involves not just incorporating data experts and domain experts into the design team but also fostering collaboration both within and outside the company. The goal is to transition from a traditional mindset to a more data-centric one, embracing a system of systems thinking in the process.

A commonly suggested strategy is to "think big, start small," which advises starting with small projects to realise benefits quickly while keeping the larger objective in sight. In parallel, an emphasis on data was recognised, with plans to improve the quality of input data, thereby increasing the consistency and overall value of the DT.

Looking towards a direction of collective growth and learning, the creation of a marketplace or registry for best practices and experiences in the development of DTs was suggested. This resource would serve as a database for companies to learn from past experiences, common challenges, and successful strategies in implementing and using DTs.

In terms of practical tools, the development of plug-and-play tools, customisable for specific applications, was emphasised. Such tools would save time and resources for companies, avoiding the need to start from scratch in developing their DT solutions. Moreover, a marketplace for DT solutions was suggested to connect supply and demand efficiently.

Companies also noted the need for cloud migration, emphasising more cloud storage solutions and the creation of a defence cloud for companies operating in both the defence and commercial sectors. This need for migration aligns with the growing popularity of cloud technology in the commercial sector. Calls for more grants to initiate future DT projects, focusing on customer satisfaction, and staying updated with the field's progress were also highlighted.

Another area where companies identified a need for improvement was in predictive analytics and predictive maintenance capabilities in DTs. Predictive maintenance enables proactive equipment upkeep, which can save costs and improve equipment uptime. This was highlighted by two interviewees, as shown in Table 4.12.

An intriguing concept brought up during the interviews was the notion of "self-defining Digital Twins." This future-oriented idea envisions DTs that can autonomously adjust parameters and settings based on the state and condition of the physical entity they are monitoring, potentially enhancing the efficiency and responsiveness of DTs. This concept was mentioned by one interviewee.

The table also reveals two key areas with a significant number of references: changing the company's mindset and reskilling the team (mentioned by eight interviewees with a total count of 26) and the system of systems, ecosystem, and federation of twins (mentioned by nine interviewees with a total count of 29). These topics are highly relevant to the focus of this thesis. The emphasis on changing the company's mindset and reskilling the team reflects the recognition that successful implementation of DTs requires not just technological advancements but also organisational change and workforce development. This influenced our work by highlighting the need to address human and organisational factors in our methodology.

The importance of the system of systems and the federation of twins aligns directly with the core of this thesis, which explores the development of CDTs and their interoperability within a broader ecosystem. This emphasis guided our selection and filtering of topics to ensure that our research addressed the critical need for scalable and interoperable DT solutions that can operate within a federated system.

All these insights are detailed in Table 4.12, which summarises the next steps and needs identified by the interviewees.

An extract from the interview, Agent A14: *"I personally think that as we get into low-code no-code environments, I think that digital twins almost be self-defining in that based upon all sorts of scans the original entity, how that is interpreted, how I can apply proven analytics models to that capability. I think that is just the future of digital. And I think it is a strap-on and try to hold on for the ride. Because I think it is going to take people much further, a lot quicker.*

It requires things like comms infrastructures to be in place and searchlight".

Table 4.12: Next steps and needs.

Next steps and needs	N. interviews	Count
Change the company's mindset and reskill the team	8	26
Customer satisfaction	2	2
Focus on data	4	8
Improve predictive maintenance	2	2
Keep pace with progress	1	1
Marketplace, registry of best practice	4	4
Migration to cloud	3	3
More grants	2	2
Self-defining DT	1	1
System of systems, ecosystem, the federation of twins	9	29
Think big, start small	7	8

4.5 Discussion

The three main contributions of this paper are presented through answering the three research questions posed in the introduction section.

1) What is the application landscape of DT in manufacturing? According to the sample analysed in this paper, DT projects are primarily applied in the defence, aviation, and maritime sectors. Their pivotal role spans different asset life cycle phases such as in-service optimisation, Maintenance, Repair, and Overhaul (MRO), and mid-life upgrades. Most DT projects, distributed fairly across small, medium, and large scales, focus predominantly on individual products (assets) rather than the broader systems they belong to, like supply chains. Financially, DT project returns typically materialise within two years for smaller projects and under four years for larger ones.

2) What are the current practices in manufacturing? The process of designing DTs in manufacturing begins by determining the purpose and motivation of the DT. It also involves considering its elements, semantic properties, and visualisation attributes. Once the design process is established, the properties of DTs become critical. They include attributes such as portability, scalability, and interoperability, which directly influence their utility and effectiveness in the manufacturing processes. On the financial side, views on the use of DTs differ. Some companies utilise them as tools to identify hidden operational benefits, disregarding the need for a specific ROI for these projects. In contrast, others focus on the ROI and the BEP. This calculation can vary based on factors such as the industry, company size, and nature of the project.

3) What are the future needs? Based on the results presented in the paper, the future needs for DTs in manufacturing industries are as follows:

- **Achieving semantic interoperability between DTs**. Semantic interoperability between DTs is necessary for building an ecosystem or federation of twins. This can be achieved through a ”system of systems” approach.

- **More focus should be placed on the system rather than just the product**. The study found that the focus of DT has primarily been centred on the product (e.g., asset) rather than the system (e.g., processes) that consisted of, among other things, the supply chain. Thus, there is a need to shift the focus towards the system to achieve better outcomes.

- **Reskilling the team**. There is a need to change the company's mindset and reskill the team to enable the effective implementation and use of DTs.

- **Think big, start small**. Adopting a gradual approach to DT implementation can help manage risks and ensure steady progress.

- **Focus on data**. Emphasising the importance of data quality and management is crucial for the successful deployment of DTs.

- **Marketplaces and registries of best practices**. The need for marketplaces and registries of best practices can enable companies to learn from each other and achieve better outcomes. (4 interviews, 4 counts)

- **Cloud migration**. Cloud migration is needed to enable better accessibility and scalability of DTs.

- **Addressing the inhibitors of DTs**. The inhibitors of DTs, such as interoperability issues, the company's mindset, complexity, quality of data, uncertainty, and expense, need to be addressed to promote the implementation of DTs in companies.

- **Improving predictive analytics and maintenance capability in DTs**. This can enable companies to optimise processes and achieve better outcomes.

- **More grants**. More grants can enable companies to invest in DTs and achieve better outcomes.

- **Self-defining DTs**. This approach offers the potential to enhance efficiency and effectiveness, complementing existing methodologies.

Both the survey and interviews highlighted significant challenges and barriers in DT development. The main barriers and solutions to overcome them are depicted in Figure 4.11. Based on the findings presented in the paper, evidence suggests that the use of DTs is likely to increase in the manufacturing industry. The study found that most DT projects are in progress or completed, with an even distribution of sizes. Additionally, the return on investment for DT

projects was generally realised within two years for small projects and under four years for large projects.

Future needs for DTs were identified, including a focus on the system rather than just the product, achieving semantic interoperability between DTs, reskilling the team, cloud migration, marketplaces of best practices, plug-and-play tools, and improving predictive analytics and maintenance capabilities. These needs reflect a growing interest in and recognition of the potential benefits of DTs in the manufacturing industry, as well as efforts to overcome some of the challenges identified in the study. The increasing interest in the literature on DTs and their potential applications in various industries, including manufacturing, suggests a growing awareness and understanding of the benefits of DTs, which is likely to contribute to their increased use in the future.

The most significant impact of DTs is likely to be in improving efficiency and productivity in the manufacturing industry. The study found that DT projects primarily focused on optimising the asset's life cycle during operation, including in-service optimisation, maintenance, repair, overhaul (MRO), and mid-life upgrade and modification, which can lead to better efficiency and productivity. Additionally, the study identified the properties of DTs, such as transparency, traceability, responsiveness, adaptability, and connectivity, as well as the financial aspects of DTs, such as return on investment (ROI) and break-even point (BEP), as important considerations for their implementation. Addressing the challenges identified in the study, such as interoperability issues, mindset shifts, reskilling, and data focus, can also contribute to the successful implementation of DTs and their impact on improving efficiency and productivity in the manufacturing industry.

By addressing these challenges, organisations can more effectively implement DTs and leverage their potential benefits to improve efficiency, productivity, and other key performance indicators in the manufacturing industry. One significant finding in this paper is the current mindset of companies, which, from the authors' perspective, hinders the development of DTs. This mindset needs to change for the development and progression of DTs. The current mindset of companies includes a lack of understanding of the technology and its potential benefits, internal managerial and procedural issues, and a lack of necessary skills among team members. This mindset can hinder the development of DTs by limiting the technology's implementation, adoption, and successful use.

Adopting DTs in the manufacturing industry can be a powerful catalyst for boosting efficiency, productivity, and optimising various other performance indicators. To achieve widespread adoption and maximise its potential benefits, a comprehensive and strategic roadmap has been proposed and illustrated in Figure 4.12. This roadmap is based on the results obtained from both the survey and the interviews.

The roadmap was developed through a systematic process involving the following steps:

1. Conducting a comprehensive assessment of the company's current capabilities and iden-

Barriers	How to overcome barriers
Interoperability issues: Difficulty in integrating DTs with existing systems, software, and data formats can hinder seamless communication and data exchange between different components of a manufacturing process.	Developing standardised frameworks and protocols for interoperability to ensure seamless communication and data exchange between different components of a manufacturing process.
Company mindset: Resistance to change within an organisation or a lack of understanding of the potential benefits of DT technology can slow down adoption and limit its effectiveness.	Investing in employee training and education to reskill the workforce and enable them to effectively implement and use DT technology.
Quality of data: Ensuring the accuracy, consistency, and reliability of the data used in DTs is critical for effective decision-making. Incomplete, outdated, or inaccurate data can compromise the effectiveness of the DT.	Ensuring data quality by using accurate, consistent, and reliable data in DTs, is critical for effective decision-making.
Uncertainty: Uncertainty surrounding the return on investment, potential benefits, and long-term impact of DT technology can make organisations hesitant to adopt it or allocate sufficient resources for its implementation.	Promoting a culture of innovation and collaboration within the organisation to foster a positive mindset towards DT technology and its potential benefits.
Expense: The cost of implementing and maintaining DT technology, including the required hardware, software, and personnel training, can be a significant barrier for companies, particularly small and medium-sized enterprises.	Conducting cost-benefit analyses to better understand the potential ROI of DT projects and identify opportunities for process improvements.
Security concerns: Implementing DT technology can expose companies to potential security risks, such as unauthorised access to sensitive data, intellectual property theft, and cyberattacks targeting critical infrastructure. Ensuring the confidentiality, integrity, and availability of information within DTs is crucial for protecting both the digital and physical assets of a company.	Implementing encryption, secure authentication, and access control mechanisms to protect the confidentiality, integrity, and availability of information within DTs.

Figure 4.11: Barriers and how to overcome them.

tifying potential areas where DT technology can be applied to improve efficiency, productivity, and other key performance indicators (Step 1).

2. Developing a clear vision and strategy for implementing DT technology, including identifying key stakeholders, setting goals, and establishing a timeline for implementation (Step 2).

3. Developing a comprehensive plan for reskilling the workforce and providing the necessary training and education to enable employees to effectively use DT technology (Step 3).

4. Establishing standardised frameworks and protocols for interoperability to ensure seamless communication and data exchange between different components of a manufacturing process (Step 4).

5. Implementing secure authentication, access control mechanisms, and encryption to protect the confidentiality, integrity, and availability of information within DTs (Step 5).

6. Conducting cost-benefit analyses to better understand the potential ROI of DT projects and identify opportunities for process improvements (Step 6).

7. Leveraging cloud migration to enable better accessibility and scalability of DT technology (Step 7).

8. Fostering a culture of innovation and collaboration to promote a positive mindset towards DT technology and its potential benefits (Step 8).

9. Establishing marketplaces and registries of best practices to enable companies to learn from each other and achieve better outcomes (Step 9).

10. Developing plug-and-play tools to enable companies to quickly and easily implement DT technology (Step 10).

11. Increasing the availability of grants to enable companies to invest in DT technology and achieve better outcomes (Step 11).

12. Staying up-to-date with the latest advances in DT technology and continuing to improve predictive analytics and maintenance capabilities (Step 12).

By following this roadmap, organisations can systematically address the barriers and challenges identified in the study and effectively implement DTs to enhance efficiency, productivity, and other key performance indicators in the manufacturing industry.

Figure 4.12: Roadmap for wider adoption of DTs in the manufacturing industry.

4.6 Conclusion and future work

This paper aims to fill the gap in the literature about the current industrial practices of DTs. This paper addresses this gap, providing insights into the scientific literature about the application landscape, current practices, and future needs of DTs directly from the people who work and manage DT projects in the industry. Those insights have been extracted from the analysis of the results coming from an online survey and semi-structured interviews.

The key takeaways from this study on DTs in the manufacturing industry include:

1. **Growing interest and adoption**. The study shows that there is an increasing interest in and adoption of DTs in various sectors, such as defence, aviation, and maritime, with a focus on assets rather than systems.

2. **Identified challenges**. Interoperability issues, company mindset, complexity, data quality, uncertainty, expense, and security concerns are the primary challenges faced while implementing DTs in the manufacturing industry.

3. **DT properties and financial aspects**. Key properties of DTs, such as transparency, traceability, responsiveness, adaptability, and connectivity, are essential considerations for their implementation. ROI and BEP are significant financial aspects that influence the adoption of DTs.

4. **Future needs and developments**. The study identifies several future needs and potential developments in the field of DTs, including a ”system of systems” approach, semantic interoperability, mindset shifts, reskilling, data focus, cloud migration, marketplaces of best practices, plug-and-play tools, and improving predictive analytics and maintenance capabilities.

5. **Roadmap for wider adoption**. A strategic roadmap is proposed for the wider adoption of DTs in the manufacturing industry, including steps such as conducting assessments, developing a clear vision and strategy, reskilling the workforce, standardising frameworks and protocols, implementing security measures, conducting cost-benefit analyses, and fostering a culture of innovation and collaboration.

6. **Collaboration between industry and academia**. The study highlights the importance of collaboration between industry and academia to advance the understanding and implementation of DTs, leading to the development of best practices, innovative solutions, and effective dissemination of knowledge.

One of the main needs arising from the interviews is the concept of the Federation of Twins. The authors have already published a systematic literature review about the concept of the Cognitive Digital Twin [D'Amico et al., 2022]. The Cognitive DT concept is proposed as a DT with

augmented semantic capabilities for identifying the dynamics of the virtual model evolution, promoting the understanding of interrelationships between virtual models and enhancing the decision-making based on DT [Lu et al., 2020b]. The authors see the Cognitive DT model as one of the enablers and facilitators for the creation of the Federation of Twins. The semantic tools (such as ontologies and graph databases) bring to the DT, a more human-like method [Fattahi et al., 2021] of processing data and information. This article empowers the need for research in the field of cognitive DT, thus contributing to the field of cognitive computing.

However, this research is constrained by the limited sample size of companies surveyed, all of which were located in the UK and predominantly engaged in the defence sector. These companies are primarily manufacturing industries within the defence sector, which means that the findings are still relevant to manufacturing, even if the context is defence. It is known that DT represents innovation for many industries, and it has not yet been extensively adopted; therefore, it is always challenging to find industries ready to invest time in this type of study.

A more comprehensive study, employing a larger sample size and incorporating international representation, would yield a broader and more holistic perspective on the subject matter. Future research could encompass a more diverse array of companies and industries to assess the progress of ongoing projects and identify new initiatives that have emerged in the interim. This approach would facilitate the collection of more extensive data and insights into the current state and future potential of DTs across various sectors and geographical locations.

It would be beneficial to conduct further investigation to assess the implementation of DTs across a broader range of companies. As the research progresses, it would be advantageous to delve deeper into the subject matter using a more balanced survey that encompasses all relevant sectors and enables cross-sectoral comparisons. Furthermore, subsequent rounds of surveys and interviews can be carried out to evaluate the progress of current DT projects and identify any new initiatives that have been launched. This would contribute to a more comprehensive understanding of the present state of DT implementation and its ongoing evolution.

This article yields numerous results that can be employed as a foundation for further research projects or to generate new investigative endeavours. The aspects of DTs highlighted in response to the third research question represent key areas companies are keen to explore soon.

Appendix A

Following the questions used in the online survey.

1. What business challenge is your digital twin project looking to address? [Open]
2. What level of maturity is your project? [Multiple choice]
3. What is the theme of the digital twin application? [Multiple choice]
4. What areas are applicable to your digital twin? [Multiple choice]

5. What is the complex system scope or focus? [Multiple choice]
6. Would you like to share any further details? [Open]
7. What sector is your digital twin applied to? [Multiple choice]
8. What is the status of your project? [Multiple choice]
9. What is the scale of the project? [Multiple choice]
10. What is the project start date? [Short answer]
11. What is the forecast return on investment (when will benefits in the business case be realised)? [Short answer]
12. Enter your organisation, department, and role [Short answer]
13. If this is a consortium/team who else is involved? [Open]
14. Who is the end customer, user, or target audience? [Open]
15. Please provide your name and contact details. [Short answer]
16. We may wish to contact you to gather additional information in relation to the creation of the compendium of Digital Twinning Implementation. If you would like to become a member of the Community of Practice. Please provide your consent below: [Multiple choice]

Appendix B

Following the questions used in the interviews.

1. How do you define a Digital Twin?
2. How do you design a DT and how the DT design team is composed, do you have any formal procedure?
3. How long is the ROI that a DT project should have?
4. How did you get the funding and where from?
5. How did you launch the design process? How do you validate the design process? What steps do you go through? Do you have any milestones to reach?
6. Have you considered any ontology/semantic approach in DT development?

7. Do you see a common approach that could be developed across the industry, or do you think it is going to have to be company-driven? (Federation of Twins) Talking about interoperability, and the link between software platforms, how much of a challenge do you see in interoperability, and how do you think that will become addressed over time?

8. How do users make decisions? Where the visualisation (Dashboard, VR, AR, XR, etc) has a role here in terms of improving the process of making a decision?

9. What are the challenges you faced, and why you needed a DT? What business challenge is your digital twin project looking to address?

10. What worked and what did not work?

11. What are the lessons learned from the whole process (business case definition, design, and development)?

12. What do you think are the next steps?

Chapter 5

Methodology

Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins

Published as Rosario Davide D'Amico, Arkopaul Sarkar, Mohamed Hedi Karray, Sri Addepalli, John Ahmet Erkoyuncu; Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins; CIRP Journal of Manufacturing Science and Technology; Volume 52; 2024; Pages 366-385; https://doi.org/10.1016/j.cirpj.2024.06.007

Abstract

In the realm of Digital Twins (DTs), industry experts have emphasised the pivotal concept of the Federation of Twins, envisioning seamless collaboration across sectors driven by shared semantics. In response to this challenge, the Cognitive Digital Twin (CDT) integrates the DT framework with formal semantics, specifically ontologies. This paper introduces a comprehensive five-step methodology for CDT development. Furthermore, it becomes possible to incorporate human expertise into the DT ecosystem by adopting an ontological approach. The CDT enhances DT services with advanced reasoning capabilities, leading to a profound semantic enrichment of the data. The presented methodology has been validated using a use case where the CDT is employed to detect malfunctions, significantly reducing manual intervention. This paper advocates for the adoption of CDTs, which represent a harmonious fusion of formal semantics and human expertise, enhancing system efficiency and operational performance.

Keywords

Basic Formal Ontology (BFO); Common Core Ontology (CCO); Connected Digital Twins; Digital Twin; Industrial Ontology Foundry (IOF); Knowledge Graph; Semantics; Top-Level Ontology

5.1 Introduction

The current industrial direction aims to enhance productivity across production and management systems. Businesses are reconsidering their production tools, automation processes, and fostering collaboration within the supply chain, engineering, and sales [Tao et al., 2019]. In some industries, this transformative shift is seen as the beginning of a cognitive manufacturing era [Bonnaud et al., 2019]. Within this context, cognitive manufacturing systems are emerging with the capability to autonomously recognise, assess, plan, predict, optimise, react, adapt, and enhance their operations. They exhibit decision-making and action-taking abilities akin to real-world cognitive behaviour, especially in dynamic and changing environments [ElMaraghy and ElMaraghy, 2022].

From an industry perspective, the pressing need is to establish more interconnected and collaborative frameworks. Notably, a survey conducted with industrial experts regarding the future requirements of DTs has underlined the significance of the "federation of twins" or the "network of twins" concept [Baek, 2022, D'Amico et al., 2023]. To clarify, this paradigm centres on the development of DTs that facilitate interoperability, allowing them to function seamlessly within and across sectors. Moreover, it offers the advantage of a shared common semantics between the twins. Therefore, the federation of twins' paradigm becomes a priority to tackle complex challenges and optimise the operations in large-scale systems, such as smart cities, national DT programs, and complex engineering assets [Baek, 2022].

This concept of the federation of twins is analogous to the Federated Social Web effort discussed in [Bojārs et al., 2008, Passant et al., 2009, Kim et al., 2008]. Both paradigms emphasise the importance of interoperability and shared semantics to create interconnected and collaborative frameworks. In the Federated Social Web, technologies like SIOC (Semantically Interlinked Online Communities) and FOAF (Friend-of-a-Friend) integrate and enrich data across different social media platforms, promoting seamless communication and data exchange. Similarly, the federation of twins leverages common semantic frameworks to ensure that various DTs can interoperate effectively, enabling them to work collectively in diverse domains and applications.

Furthermore, Kim et al. highlights the need for aligning and federating different tag ontologies to enable seamless tag sharing and data reuse across platforms. This approach is akin to the federation of twins' effort to overcome interoperability issues by adopting shared semantic standards. Both efforts illustrate the importance of creating unified frameworks for efficient data exchange and collaboration across different systems [Kim et al., 2008].

Just as the Federated Social Web addresses the challenges of data isolation and the lack of common standards by adopting semantic technologies, the federation of twins aims to overcome interoperability issues and facilitate collaboration between DTs through shared semantics. Both efforts highlight the need for standardised frameworks to enhance the utility and efficiency of interconnected systems [Passant et al., 2009].

To elucidate further, the concept of "federation of twins" is distinct from "federated learning," which primarily concerns privacy preservation and collaborative machine learning [Treleaven et al., 2022]. In contrast, the federation of twins focuses on the interconnection, interoperability, and shared semantics between individual digital twins, enabling them to work collectively and efficiently in diverse domains and applications. This clarification should help avoid any potential confusion between the two concepts.

There has been an increase in the use of ontologies as a framework for managing knowledge and data in DTs. The systematic literature review conducted by Correia et al. reports the adoption of industrial standards and ontologies to enhance interoperability and to provide the reference architecture to develop DTs [Correia et al., 2023]. Ontologies are tools used to represent knowledge through an "explicit, formal specification of a shared conceptualisation" [Gruber, 1995]. In ISO 21838:2021, ontology is defined as a "collection of terms, relational expressions and associated natural-language definitions together with one or more formal theories designed to capture the intended interpretations of these definitions". They provide a structured and organised way to model and represent knowledge, allowing for better understanding, communication, and information management [Steiner and Albert, 2017].

In the context of knowledge management, ontologies serve as a framework for organising and categorising knowledge assets within an organisation. They help to define and standardise the vocabulary and terminology used in a particular domain, enabling effective knowledge sharing and retrieval [Almeida and Barbosa, 2009]. By explicitly specifying the conceptualisation of a domain, ontologies facilitate the identification, classification, and retrieval of relevant information [Guarino and Welty, 2009]. Preliminary evidence from a case study in Singapore demonstrated that a DT using semantic technologies (ontologies) and knowledge graphs improves the planning of smart cities and the design of future energy systems [Akroyd et al., 2021]. The association between these two paradigms (DTs and ontologies) has led to an evolution of the concept of DT, where additional components have been added to the standard approach (see section 5.2).

Cognition and semantics are closely related, as cognition refers to the mental processes and activities involved in acquiring, processing, and using information. In contrast, semantics is the study of meaning in language and how it is encoded and understood [Hsiao, 2003]. In the Semantic Web context, concerns about its cognitive plausibility exist. Some propose using conceptual spaces, a knowledge representation framework aligned with human cognition, to improve Semantic Web data utility [Raubal and Adams, 2010]. Additionally, it is suggested that the Semantic Web should consider human cognition, including conceptual representations and reasoning mechanisms. This involves incorporating conceptual spaces rooted in cognitive semantics and emphasising factors like similarity and prototype effects in categorisation as a more realistic approach to knowledge representation [Goddard and Wierzbicka, 2011]. An ontological approach might satisfy this need. Recent research shows that the DT framework

has implemented a cognitive or semantic layer. Such DT frameworks with cognitive/semantic aspects are:

- *The universal digital twin* [Akroyd et al., 2021] – a DT which includes semantic web technologies and dynamic knowledge graphs;
- *The next-generation digital twin* [Boschert et al., 2018] – a framework where DTs are connected through a knowledge graph enhancing knowledge sharing;
- *The cognitive twin* [Lu et al., 2020a] - DT that uses ontology for managing interrelationships between models and virtual entities;
- *The cognitwin* [Abburu et al., 2020] – a third stage of evolution of the DT concept implementing interconnection between twins, and expert knowledge and historical data into the DT;
- *The cognitive digital twin* [Al Faruque et al., 2021] – a DT that includes elements of cognitive science, machine learning, and artificial intelligence harnessing implicit knowledge derived from the operational experience of existing manufacturing systems, facilitates the transfer of more effective decision-making and control methods, resulting in enhanced overall performance across the system.

Most of the frameworks presented in the literature include an ontology as the methodology to semantically enrich the knowledge into the DT. The concept chosen in this paper is the cognitive twin, which recently evolved into the cognitive digital twin (CDT). These CDTs include a graph database to manage the data, and a series of ontologies including a top-level ontology (TLO) [ISO 23247-1] to facilitate the semantic interoperability between the CDTs, enabling the creation of the "federation of twins". The CDT is defined as "a digital representation of a physical system that is augmented with certain cognitive capabilities and support to execute autonomous activities; comprises a set of semantically interlinked digital models related to different life cycle phases of the physical system including its subsystems and components; and evolves continuously with the physical system across the entire life-cycle" [Zheng et al., 2022b].

The papers mentioned above highlight several benefits of CDTs, including interoperability, portability, and the incorporation of human knowledge for reasoning. These characteristics are crucial in facilitating the creation of a federation of twins' paradigm. Despite these advantages, few publications offer a comprehensive methodology for constructing CDTs, indicating a significant gap in the existing literature. There is a clear demand for conceptual frameworks that can effectively guide the development of CDTs. Addressing the need for seamless knowledge sharing, transfer, and scalability within and across multiple domains in CDTs is also essential [Al Faruque et al., 2021]. This underscores not only the need for structured approaches and best practices in creating CDTs but also the necessity for these systems to be interoperable and

flexible, capable of adapting to various environments and applications. This situation highlights the growing interest in CDTs and their potential applications while simultaneously identifying crucial areas for further research and development.

This paper addresses the above-mentioned gaps by presenting a methodology for developing and validating a CDT framework, illustrated by creating a CDT for a laboratory setup. The laboratory setup comprises a Festo sorting station. Festo is a German company specialising in industrial automation and control. They are renowned for developing and manufacturing pneumatic and electrical control systems, automation technology, and educational equipment. Festo's products and solutions find widespread application across various industries, including manufacturing, automation, and education, enhancing efficiency and productivity. This system was chosen because it embodies the complexities of real-world material handling systems, featuring multiple sensors and high flexibility. It is also controlled by industry-standard Programmable Logic Controllers (PLCs), aligning the laboratory setup with actual industrial applications.

The novelty of this research lies in the proposed methodology, which incorporates human expertise into the DT domain. Furthermore, the methodology is tool-agnostic, based on system engineering approaches, and aligns well with manufacturing systems.

The research questions this paper aims to answer are the following:

1. Can an ontology approach be used to develop a DT?

2. Can human expertise be transferred to the CDT?

3. Can a CDT support a root cause analysis for the malfunction of a system?

The remainder of the paper is structured as follows. Section 5.2 presents the literature review on CDTs, highlighting the state of the art and the current gaps in addressing them. Section 5.3 presents the methodology. Section 5.4 presents the framework application using the Festo sorting station. Section 5.5 presents the discussion; section 5.6 presents the conclusion and future works.

5.2 Literature review

The number of publications on CDTs is growing, as reported in Figure 5.1. The result comes from a search of the "cognitive digital twin" string on the Scopus database on the 22nd of June, 2023. A total of 38 publications, constituting 19 conference papers and 19 journal articles that directly support this research, have been identified.

Key breakthroughs inspire the CDT concept in cognitive science, machine learning, and artificial intelligence. According to Neisser's classic definition [Solso et al., 2008], cognition involves 'all the processes by which the sensory input is transformed, reduced, elaborated, stored,

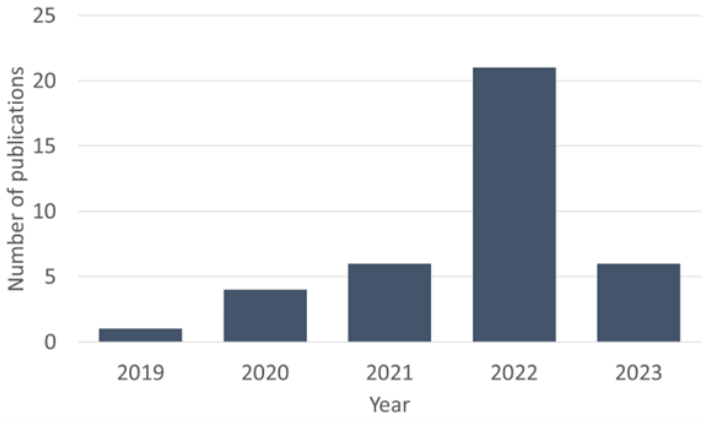


Figure 5.1: Number of publications regarding CDTs.

recovered and used'. The fundamental characteristics of cognition are attention (selective focus), perception (forming practical precepts from raw sensory data), memory (encoding and retrieval of knowledge), reasoning (drawing inferences from observations, beliefs, and models), learning (from experiences, observations, and teachers), and problem-solving (achieving goals) [Al Faruque et al., 2021]. A list of existing definitions of cognitive digital twins and related concepts is reported in Table 5.1.

Table 5.1: Definitions of Cognitive Digital Twins in the literature.

Reference	Concept	Definition
[Fernández et al., 2019]	Associative Cognitive Digital Twin (AC-DT)	'A contextual augmented description that explicitly includes the relevant associated relationships, connections or information channels of the considered entity with the other entities, for the considered application and purpose'
[Eirinakis et al., 2020]	Enhanced Cognitive Twin (ECT)	'Introduce advanced cognitive capabilities to the DT artefact that enable supporting decisions, with the end goal to enable DTs to react to inner or outer stimuli. The ECT can be deployed at different hierarchical levels of the production process, i.e. at sensor-, machine-, process-, employee- or even factory-level, aggregated to allow both horizontal and vertical interplay'
[Lu et al., 2020a]	Cognitive Twin	'Digital Twins 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'
[Abburu et al., 2020]	COGNITWIN - Cognitive Digital Twin	'An extension of Hybrid Digital Twins incorporating cognitive features that enable sensing complex and unpredicted behaviour and reason about dynamic strategies for process optimisation, leading to a system that continuously evolves its digital structure as well as its behaviour'
[Ali et al., 2021]	Cognitive Digital Twin	'An extension of existing digital twins with additional capabilities of communication, analytics, and intelligence in three layers: i) access, ii) analytics, and iii) cognition'
[Al Faruque et al., 2021]	Cognitive Digital Twin	'Digital Twins endowed with the other elements of cognition such as perception, attention, memory, reasoning, problem-solving, etc'
[Zheng et al., 2022a]	Cognitive Digital Twin	'A digital representation of a physical system that is augmented with certain cognitive capabilities and support to execute autonomous activities; comprises a set of semantically interlinked digital models related to different life cycle phases of the physical system including its subsystems and components; and evolves continuously with the physical system across the entire lifecycle'
[ElMaraghy and ElMaraghy, 2022]	Cognitive Digital Twin	'A digital representation, augmentation, and intelligent companion of its physical twin as a whole including subsystems and across all of its life cycles and evaluation phases'

The literature identifies several gaps, research challenges, and open research questions related to cognitive digital twins. These include:

- Lack of consensus on the definition and scope of cognitive digital twins: There is no widespread consensus on the definition and scope of cognitive digital twins, leading to ambiguity and confusion [Zheng et al., 2022a].
- Integration of heterogeneous digital twin models: Integrating digital twin models from different stakeholders with different protocols and standards is a challenging task due to the heterogeneity of data structures [Zheng et al., 2022a].
- Semantic integration of digital twin models: Semantic technologies such as ontology and knowledge graphs are promising solutions for integrating heterogeneous digital twin models, but further research is needed to explore their full potential [Lu et al., 2022, Zheng et al., 2022a].
- Enabling technologies for cognitive digital twins: Research is needed to explore and develop enabling technologies such as knowledge graphs, artificial intelligence, and machine learning techniques to enhance the cognitive capabilities of digital twins [Lu et al., 2022, Zheng et al., 2022a].
- Life cycle representation of cognitive digital twins: Cognitive digital twins should be able to represent the entire life cycle of a system, including its subsystems and components, and evolve continuously with the actual system. Further research is needed to support this comprehensive representation and to develop frameworks and architectures [Zheng et al., 2022c].
- Decision-making capabilities of cognitive digital twins: Research is needed to explore how cognitive digital twins can support decision-making processes, including real-time data analysis, learning, reasoning, problem-solving, and autonomous decision-making [Zheng et al., 2022a].
- Interoperability and connectivity of cognitive digital twins: Research is needed to address the interoperability and connectivity challenges of integrating multiple cognitive digital twins of different systems and domains, enabling seamless communication and collaboration [Zheng et al., 2022c].
- Validation and verification of cognitive digital twins: Research is needed to develop methodologies and techniques for validating and verifying the accuracy, reliability, and performance of cognitive digital twins in real-world manufacturing settings [Yitmen et al., 2021, Zheng et al., 2022c].

Some relevant future works highlighted in the literature include:

- The need for novel conceptual frameworks and algorithmic breakthroughs to develop cognitive digital twins. The facilitation of knowledge sharing and transfer among digital twin models and the scalability of knowledge sharing across multiple domains in cognitive digital twins [Al Faruque et al., 2021].

- Enhancing the cognitive capability of industrial systems by combining ontology-based systems and MBSE methods [Kalaboukas et al., 2021].

- Developing methodologies and tools to enable knowledge capturing considering different expert/user-oriented formalisation stages. Improving the effectiveness and efficiency of the ontology-based engineering system and enhancing knowledge capturing and utilisation [Arista et al., 2023].

- Exploring collaborative behaviours such as value co-creation in the context of tangible and intangible value. Additionally, determining members' contributions in a collaborative venture. These future works aim to enhance further the understanding and implementation of the concept of cognitive digital twins [Adu-Kankam and Camarinha-Matos, 2022].

- Identifying and extending ontology-driven services to enhance digital consistency, interoperability, and reusability in engineering activities. There is a lack of research on integrating ontology as a semantic core for industrial system design [Hu et al., 2022].

- Identify and extend ontology-driven services to enhance digital consistency, interoperability, and reusability in engineering activities [Ali et al., 2021].

- Generalising knowledge sharing across multiple domains and enabling scalable knowledge sharing in digital twin models [Mortlock et al., 2022].

- Explorationing knowledge graph-based analysis in the context of digital twins. There is a lack of tools and methodologies for integrating knowledge graphs into industrial settings [Meyers et al., 2022].

- Developing a top-level ontology-based digital twin to improve human-machine interoperability [D'Amico et al., 2021], enhancing the CDT with optimisation algorithms for decision-making [Eirinakis et al., 2020], and expanding research in the aircraft manufacturing domain to cover additional assembly stations and address system complexity [Zheng et al., 2022b].

- Developing a solid methodological framework for the implementation of DTs in manufacturing and the absence of a systematic approach in the scientific literature that applies the concepts of CDTs in a complex system of systems setup, such as agile supply chain networks [Ashraf et al., 2022, Kalaboukas et al., 2021, Rožanec et al., 2021].

These gaps, challenges, next steps, and open research questions highlight the need for further research and development to fully realise the potential of cognitive digital twins in various industrial domains. There are limited publications that present a methodology to develop CDT. Kalaboukas et al. presented a methodology composed of four steps: scope definition, system modelling, design and operation, and rollout. However, the application of this methodology focuses on the supply chain. The cognitive capabilities are characterised by a semantic representation of data without any detailed explanation or any use of ontology [Kalaboukas et al., 2021]. Most CDT publications focus more on describing the elements and the architecture rather than a general methodology to develop the CDT. This paper aims to provide a methodology to build a CDT framework that addresses most of the aforementioned research gaps. Moreover, the methodology has been validated using a laboratory-based experiment.

5.3 Methodology

In this section, a comprehensive and structured methodology for the development of CDTs is presented. The primary objective of this methodology is to provide a well-defined, step-by-step procedure that encompasses all essential aspects involved in the process of CDT creation. What sets this methodology apart from the existing literature is its emphasis on the procedural dimension rather than delving into architectural or tool-related details. The methodology proposed is the culmination of iterative work, lessons learnt, and research experience gained throughout this project.

This methodology consists of five key steps, each playing a crucial role in the coherent and successful development of CDTs. What characterises a step is the achievement of a complete and tangible result at its conclusion, which then becomes instrumental for the subsequent step. Adhering to this criterion, we have delineated a process comprising five distinct steps. These five steps are not isolated entities but rather intricately interconnected, demonstrating a significant interdependency that ensures a cohesive and effective CDT development process.

The methodology's structure is underpinned by theoretical principles from systems engineering and interdisciplinary insights, ensuring a robust and holistic framework (inspired by the one proposed and discussed in [Lu et al., 2022]. Each step was designed to align with the goals of CDT development, focusing on integrating human expertise, data interoperability, and facilitating informed decision-making in complex asset management. Consequently, this methodology is tailored for CDT with a strong ontology component and is influenced by the recent methodology for developing ontologies, Linked Open Terms (LOT), proposed by Poveda-Villalón et al. [Poveda-Villalón et al., 2022]. This approach ensures that the methodology addresses current challenges in CDT development and lays a foundation for future advancements in the field.

A visual representation of the methodology is reported in Figure 5.2. The presented

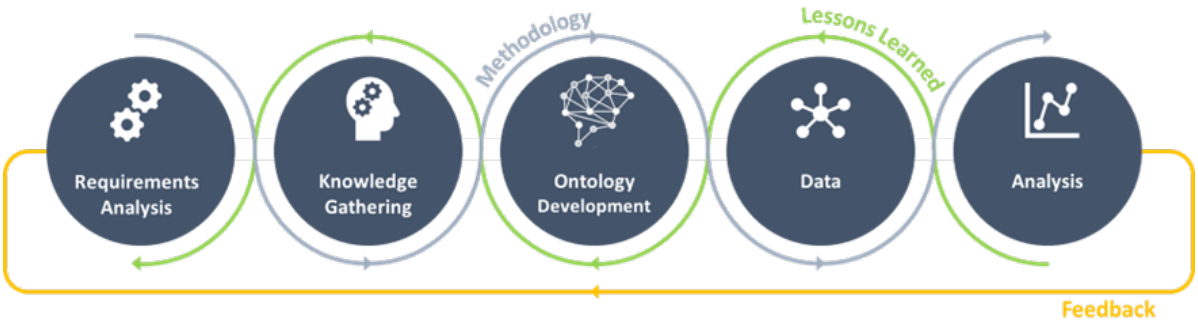


Figure 5.2: Methodology to develop CDT.

methodology provides a structured framework for developing CDTs and emphasises the procedural aspect, ensuring clarity and coherence throughout the CDT creation journey. By following these defined steps, researchers and practitioners can effectively harness the potential of CDTs across various domains and applications.

In particular, the five steps are defined as follows:

1. **Requirements Analysis**. The initial step involves a comprehensive analysis of the requirements. This includes gaining a deep understanding of the system of interest, dissecting its structural and functional components, and defining the scope and purpose of the intended CDT.

2. **Knowledge Gathering**. This step entails the systematic collection of a broad spectrum of essential information. This information encompasses critical elements, such as underlying facts, assumptions, rules, behavioural patterns, periodic observations, and invaluable human expertise related to the system of interest. The knowledge collected in this step serves as the foundation and needs to be encoded into an ontology in the subsequent phase.

3. **Ontology Development**. The development of the ontology begins with the formulation of Competency Questions (CQs) that precisely define the ontology's purpose. This step proceeds with creating the ontology itself, which serves as a formalised structural description. It encapsulates critical aspects like functions, behaviours, processes, and state descriptions of the system.

4. **Data**. In this step, we address the data aspects of the CDT. This includes selecting the appropriate data strategy, which encompasses choosing the suitable graph database technology capable of importing both the ontology and data to the graph database. Additionally, it involves defining the query strategy, middle-ware selection, and establishing the information and knowledge transfer strategy to and from the CDT.

5. **Analysis**. The final step revolves around conducting analytical and simulation activities essential to fulfil the intended purpose of the CDT. As defined by Tao et al., this

element closely relates to the service component of the Digital Twin [Tao et al., 2019]. The outcomes of this step are intricately tied to the specific application and purpose for which the CDT is being developed. The CDT will then provide feedback, enhancing decision-making and operations.

5.3.1 Requirements analysis

This step consists of understanding the components and boundaries of the system of interest. In particular, the classification of the components, qualitative formalisation of the parthood and connection relations between components (mereotopology). Mereotopology combines mereology, the theory of parts and wholes, and topology, the study of spatial properties and relations that remain constant under continuous transformations. In essence, mereotopology explores the relationships between parts and wholes and their spatial properties and connections. Mereotopology helps define what constitutes a 'part' of a system and how these parts are spatially and functionally related to each other and the 'whole' system. This is crucial in systems engineering and design, where understanding the boundaries and interfaces between components is key to effective system integration. Mereotopology provides a framework for modelling and understanding complex systems. It helps in breaking down a complex system into more manageable parts, understanding the relationships and dependencies between these parts, and how changes in one part might affect the whole system.

In the context of requirement analysis, mereotopology can aid in understanding how different system components interact and depend on each other. This understanding is essential for identifying and defining system requirements, especially interoperability, scalability, and modularity. Mereotopology plays a crucial role in the requirements analysis, as it is fundamental for the ontology creation by providing a framework for representing and reasoning over part-whole relations (mereology) and spatial relationships (topology). Mereotopology deals with the concepts of parthood and location, which are fundamental for modelling complex relationships between entities in the world. By incorporating mereotopological relations into ontologies, we can capture the spatial and structural aspects of the modelled domain. This allows for a more accurate and comprehensive representation of knowledge, enabling better reasoning and inference capabilities within the ontology [Keet et al., 2012]. Furthermore, the relation between functions and capabilities of each component, modelling the system's behaviour, the flow of material, energy, and control in the system should all be included in the system boundary.

Capturing the parthood and changes in different situations happening over time in a system can be managed with the Mereotopology theory. Salustri proposes using design mereotopology (DMT) in product modelling, highlighting its potential for improved understanding and manipulation of product knowledge [Salustri, 2002]. Forrest presents a mereotopological system based on interior parthood without relying on classical mereology or the Axiom of Choice [Forrest, 2010]. Hahmann and Grunninger provide an overview of axiomatisations and repre-

sentations of mereotopologies, emphasising the need for practical extensions and integration into comprehensive ontologies [Hahmann and Grüninger, 2012]. Galton introduces discrete mereotopology, which considers entities built from atomic elements, and discusses its applications in qualitative spatial reasoning [Galton, 2014]. These papers demonstrate the relevance and potential of mereotopology in capturing parthood and changes in various contexts, including product modelling, theoretical frameworks, and practical applications. Casati and Varzi discuss the interplay between mereology, topology, and the theory of spatial location in understanding the structures of spatial representation. Moreover, they presented two strategies representing the combination of mereology and topology and the definition of mereotopology [Casati and Varzi, 2003].

Mereotopology is pivotal in ontological analysis [Varzi, 1998]. This theory allows for the qualitative formalisation of two essential concepts: part-whole relationships (i.e., one entity being a part of another) and connectivity (i.e., the relationship between entities being connected). The current obstacle in applying mereotopology-based theory in engineering design is the need to account for how a product is perceived in the real world [Demoly et al., 2012]. The mereotopology theory is well explained in the literature [Varzi, 1996] and has also been associated with the formal ontology domain [Smith, 1996]. It must be noted that the explanation of the underlying mereotopology theory is out of the scope of this paper. However, the application of the mereotopology theory through the use case falls within the scope of this paper and is presented in section 5.4.

The representation of the mereology structure of a system is reported in Figure 5.3. The figure represents a system composed of three artefacts (a, b, and c) where two components b and c are part of a super component a. The Venn diagram at the bottom shows how the two subcomponents belong to a. In this paper, the authors follow the concept of functions and capabilities defined by Merrell et al. [Merrell et al., 2022]. It should be noted that other schools of thought also exist and slightly differ from what will be described below [Borgo et al., 2021]. A capability is defined as a disposition in whose realisation some organism or group of organisms has/had an interest ([Merrell et al., 2022]. A function is a capability that explicitly highlights why an entity exists (teleological argument). All designed artefacts have functions because that is the definition of the word artefact. They are entities which were designed to perform a particular process. Moreover, that means they were intended to have a function realised in that process [Merrell et al., 2022].

Also, to verify the system's mechanical design, the Axiomatic Design (AD) theory has been used. The AD is a framework developed by Professor Nam P. Suh at MIT in the late 1970s and early 1980s. It is a methodology used in engineering and design processes to systematically develop solutions that meet defined objectives while minimizing complexity and improving quality [Suh, 1998]. It is used to break down the system into its components and their functions. This theory is applicable across various types of systems, encompassing ma-

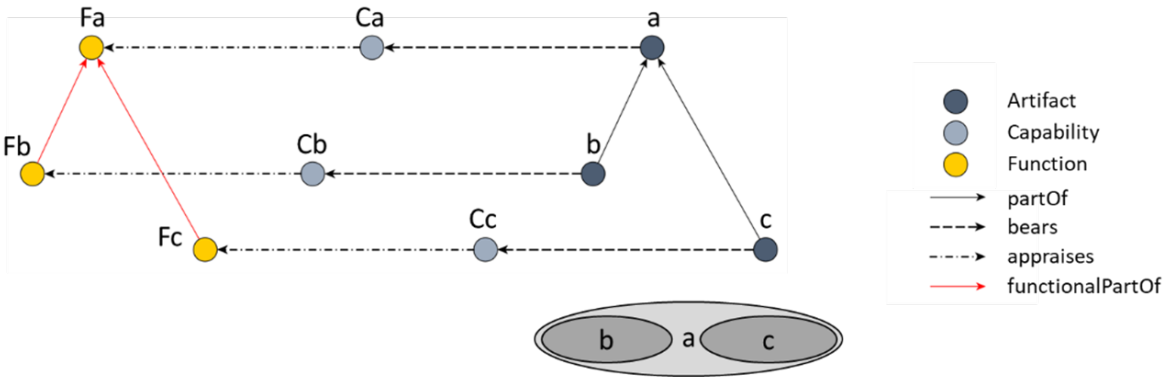


Figure 5.3: Mereotopology diagram that shows the relations between artefacts, their functions, and their capabilities.

chines, complex systems, software systems, organisational structures, and hybrid systems combining hardware and software elements. Systems are represented using a system framework, which includes the {*Functional Requirements*}/{*Design Parameters*}/{*Process Variables*} hierarchies, a 'junction-module' diagram, and a 'flow diagram'. This theory fits well with the methodology introduced in this paper as analyses and formalises the relationship between the objects and their functions. This analysis perfectly links with the ontological analysis.

The representation shown in Figure 5.3 is a result of the combined use of AD theory and mereological composition principles. In this figure, the design and functional relationships within the system are systematically mapped out, reflecting the structured approach of AD theory [Suh, 1998]. This approach emphasizes the clear delineation of functional requirements and design parameters, ensuring that each component of the system is effectively aligned with its intended purpose.

Moreover, the integration of mereology, which focuses on the relationships between parts and wholes, provides a comprehensive view of the system's hierarchical structure. This allows for a deeper understanding of how individual components interact within the overall system, ensuring that the design is both cohesive and logically sound.

A similar approach to system representation, combining these methodologies, has also been adopted by other researchers in the field. For instance, Sarkar et al. [Sarkar et al., 2020] have utilized a comparable framework to model complex systems, highlighting the utility and effectiveness of this integrative method in capturing both the functional and structural aspects of engineering systems. This combined approach not only enhances the clarity of the system's design but also provides a robust foundation for further analysis and development.

The 'flow diagram' within the system framework succinctly illustrates the system's design, the interconnections among its modules, and the control sequence in operating systems. This flow diagram serves many purposes, such as aiding in system design, construction, operation, modification, and maintenance. Furthermore, it is a valuable tool for facilitating distributed design and operation of systems, diagnosing system failures, and maintaining archival docu-

mentation [Suh, 1998].

With this approach, at each level of decomposition, if specific processes achieve the functions of each component, then the parent component's function is also effectively realised [Sarkar et al., 2020]. Axiom 5.1 sets this limitation by employing the predicate functionalPartOf, which acts as a specific form of parthood relation among functions. This axiom helps formalise the concept of systems and their functional relationships within a more extensive system or context.

$$\begin{aligned}\forall s, s' \ \text{functionalPartOf}(s', s) \rightarrow \text{System}(s) \wedge \text{System}(s') \wedge \\ \exists f, f' \ (\text{bears}(s', f') \wedge \text{bears}(s, f) \wedge \text{decompose}(f', f))\end{aligned} \tag{5.1}$$

∴ 5.1 For any entities s and s', if s' is a functional part of s, then both s and s' are systems, and there exist functions f and f' such that s bears f, s' bears f', and f' can be decomposed into f

Sarkar et al. have restructured the independence axiom from AD theory, focusing on how the associated components' capabilities back functions at various decomposition levels [Sarkar et al., 2020]. From the interplay between functions and the abilities of system components, two valid configurations can be discerned: the summation junction and the control junction, both considered types of modulo-junction. As originally outlined in AD theory, these configurations align with the diagonal and triangular design matrices.

The predicates SummationJunction 5.2 and ControlJunction 5.3 are established between a component and its two sub-components. Specifically, within the ControlJunction predicate, the component occupying the second position exercises control over the component in the third position.

$$\begin{aligned}\forall c, c', c'' \ (\text{SummationJunction}(c, c', c'') \leftrightarrow \forall p, f \ (\text{bears}(c, f) \wedge \text{realise}(p, f) \rightarrow \\ \exists f', f'', p', p'' \ (\text{bears}(c', f') \wedge \text{bears}(c'', f'') \wedge \text{decompose}(f', f) \wedge \\ \text{decompose}(f'', f) \wedge \text{realise}(p', f') \wedge \text{realise}(p'', f'')))))\end{aligned} \tag{5.2}$$

$$\begin{aligned}\forall c, c', c'' \ (\text{ControlJunction}(c, c', c'') \leftrightarrow \forall p, f \ (\text{bears}(c, f) \wedge \text{realise}(p, f) \rightarrow \\ \exists p', f', p'', f'' \ (\text{bears}(c', f') \wedge \text{bears}(c'', f'') \wedge (\text{realise}(p', f') \rightarrow \text{realise}(p'', f'')))))\end{aligned} \tag{5.3}$$

∴ 5.2 For all configurations c, c', and c'', the 'Summation Junction' relation holds if and only if, for all processes p and functions f, if c bears f and p realises f, then there exist functions f' and f'' and processes p' and p'' such that c' bears f', c'' bears f'', f' decomposes into f, f'' decomposes into f, p' realises f', and p'' realises f''

∴ 5.3 For all configurations c, c', and c'', the 'Control Junction' relation is valid if and only if, for all processes p and functions f, when c bears f and p realises f, there exist processes p' and p'', as well as functions f' and f'', such that c' bears f', c'' bears f'', and if p' realises f', then p'' realises f''

The elements of the AD theory serve as essential guides throughout the stages of system design, development, servicing, and troubleshooting. Furthermore, they offer a valuable means of documenting the system design effectively. When inspecting the flow diagram, it becomes possible to identify vulnerable design connections by recognising the interconnected components within the system's design. This insight can help pinpoint areas of potential weakness in the overall system design. Moreover, the theory should offer clear instructions for identifying the components that need to be examined when the system fails to execute specific functions [Suh, 1998].

In this step of the methodology, the applicability criteria of methods that include mereotopology, mechanical, and system engineering analysis should be aligned with the project's scope and objectives, taking into account the complexity and scale of the project.

5.3.2 Knowledge gathering

This section aims to understand the system's behaviour, model the processes that occur within a specific system and then integrate that knowledge into the CDT. This process involves several key steps:

- Defining Processes within the System: The first step is to identify and describe the system's various activities, actions, or transformations. This is essential because systems can be pretty complex, and clearly understanding what is happening is crucial for effective modelling.

- Transferring Knowledge into the CDT: CDT refers to a digital model of a real-world system that incorporates data, knowledge, and algorithms to simulate and understand how that system behaves. So, part of the process is taking the knowledge about the system's processes and incorporating it into this digital twin for analysis and simulation.

- Using First Order Logic (FOL) Axioms for Definitions: To make these processes more precise and formal, they are expressed using First Order Logic (FOL) axioms. FOL is a mathematical language that provides a structured way to represent knowledge. Each process is defined in terms of logical statements, making it clear, unambiguous, and suitable for computer-based reasoning.

- Thematic Roles - ”Instrument“ and ”Patient”[1]: To better understand and model the processes, it is essential to distinguish between the various entities or elements involved. Here, two thematic roles are introduced:

    1. Instrument: This role represents the entity or element that directly causes or initiates a process. It is the actor or agent responsible for carrying out the action.
    2. Patient: In contrast, the patient is the entity that is impacted or affected by the processes. It is the entity undergoing a change as a result of the process. These roles help define the roles of different entities within the system clearly.

- Gathering and Storing Human Expertise: Often, domain experts possess valuable knowledge about the system. They may have insights that are not necessarily expressed in a formalised way. During this phase, the aim is to collect this knowledge from domain experts and store it for later use.

- Conversion of Human Expertise to Formal Logical Statements: Human expertise, especially from experts, tends to be in natural language or an informal format. This knowledge is converted into formal logical statements to make it compatible with the formal model. This step ensures that the gathered knowledge can be seamlessly integrated into the model, maintaining consistency and compatibility with the rest of the formal representation.

This approach systematically understands, models, and simulates a specific system. It emphasises the need for precise formalisation, introduces thematic roles to clarify entity roles in processes, and stresses the importance of gathering, storing, and formalising human expertise from domain experts for a comprehensive system analysis.

In this step of the methodology, where the methods focus on systematically understanding, modelling, and simulating a specific system, the applicability criteria are centred around the thoroughness and precision of the information-gathering process. This includes emphasising the need for formalisation to ensure accuracy in the representation of the system. The method calls for clearly defined thematic roles to elucidate the functions of various entities within the processes being modelled. Furthermore, a significant emphasis is placed on collecting, storing, and formalising human expertise, particularly from domain experts, to ensure a comprehensive and insightful system analysis. This approach necessitates a meticulous and systematic process of knowledge acquisition, prioritising depth and quality of information to facilitate a robust and accurate modelling and simulation of the system. The criteria, therefore, revolve around the precision of data, clarity in role definition, and the depth of expert insights incorporated into the system analysis.

[1] `https://semantic-annotation.uvt.nl/LIRICS_semroles.htm#LIRICS_semrole`

5.3.3 Ontology development

The information gathered in the previous two steps (subsections 5.3.1 and 5.3.2) is encoded and used to develop the ontology in this phase. In general, an ontology is a set of logical statements. A guide that will help the user formalise what has been found in the previous steps in a set of consistent logical statements is presented below. The result will then be converted into a file that can be used consecutively to enhance the collected data's quality and perform more advanced analytics. The methodology to develop an ontology takes inspiration from the Linked Open Terms (LOT) methodology published by Poveda-Villalón et al. It is composed of 4 main steps: (i) ontology requirements specification, (ii) ontology implementation, (iii) ontology publication, and (iv) ontology maintenance [Poveda-Villalón et al., 2022]. This methodology emphasises the importance of ontology construction for sharing a common understanding of information structure and enabling query retrieval, highlighting its role in facilitating ontology reuse, knowledge representation, and information sharing in various domains.

The ontology development process is initiated by identifying its requirements, a pivotal step achieved by formulating competency questions (CQs). These CQs are carefully crafted queries that encapsulate the specific types of information and knowledge the ontology is meant to handle. They are derived from practical scenarios, such as the analysis phase in a CDT, which dictate what kind of information the ontology should effectively manage [Grüninger and Fox, 1995].

In conjunction with the formulation of CQs, it is essential to articulate the answers the ontology should provide when these questions are posed. This step ensures that the ontology is constructed with a clear purpose and delivers the necessary information and knowledge when called upon.

Following this, a rigorous verification process is carried out by querying the ontology with these CQs to validate that it aligns with the defined requirements. This verification process systematically compiles a glossary of essential terminology, serving as the foundational framework for constructing the ontology. This terminology helps organise key concepts and relationships within the ontology and informs the development of a hierarchical taxonomy, ensuring that the knowledge representation is structured and efficient for its intended application. In essence, this comprehensive approach ensures that the ontology is tailored to represent, manage, and provide relevant knowledge for its specific domain of application effectively.

In the context of CDT, ontology development holds a pivotal role. One key aspect of this development is the careful selection of upper-level ontologies. The different types of ontologies divided by level of abstraction are reported in Figure 5.4. These upper-level ontologies, often called top-level ontologies (TLOs), serve as the foundational building blocks for defining more specific domain ontologies within the CDT framework. TLOs like Basic Formal Ontology (BFO), Suggested Upper Merged Ontology (SUMO), and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) are generic frameworks that describe high-

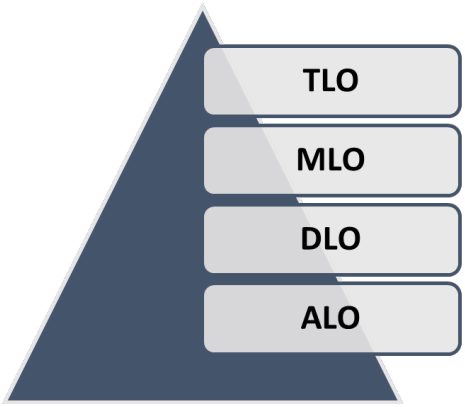


Figure 5.4: Levels of abstraction in ontology theory.

level concepts or universals. They are critical in ensuring interoperability between domain ontologies, facilitating seamless integration and knowledge reuse. In the context of CDTs, the choice of TLOs and mid-level ontologies (MLOs) is guided by the nature of the system and the knowledge required for analysis, enabling the development of domain-level ontologies (DLO) and application-level ontologies (ALO) tailored to the specific context of a CDT, taking into account factors such as functional composition and processes [Arp et al., 2015].

In addition, the design of ontologies within the CDT framework is guided by a set of eight fundamental principles, with a particular emphasis on domain ontologies [Arp et al., 2015]. These principles encompass essential aspects such as realism, perspectivalism, fallibilism, adequatism, the principle of reuse, the balance between utility and realism, open-ended ontology design, and addressing low-hanging fruit. Adhering to these principles is paramount to ensure the development of ontologies that not only address immediate utility but also accommodate revisions and updates in light of evolving scientific knowledge and technological advancements, which are crucial in the dynamic context of CDTs.

Moreover, in the context of CDTs, it is imperative to avoid developing new ontologies from scratch, as doing so could hinder interoperability. Instead, a comprehensive review and assessment of existing ontology resources related to the CDT's domain should be conducted. Ontologies are intended to facilitate communication between information resources across various domains and disciplines, aligning with the holistic nature of CDTs.

Furthermore, within the CDT framework, designing a domain ontology is just the initial step in an ongoing process. This process involves continuous maintenance, evaluation, updates, and alignment with neighbouring ontologies, considering advancements in scientific knowledge and our evolving understanding of ontology and its associated logical and computational technology. These principles underscore the importance of creating ontologies in a manner that allows for expansion and amendment over time. The best practice guidelines within this framework are intended to achieve this dynamic and adaptable ontology development approach.

In developing domain ontologies within CDTs, a preliminary phase involves classifying fundamental universal concepts and relationships. A recommended approach is to begin by

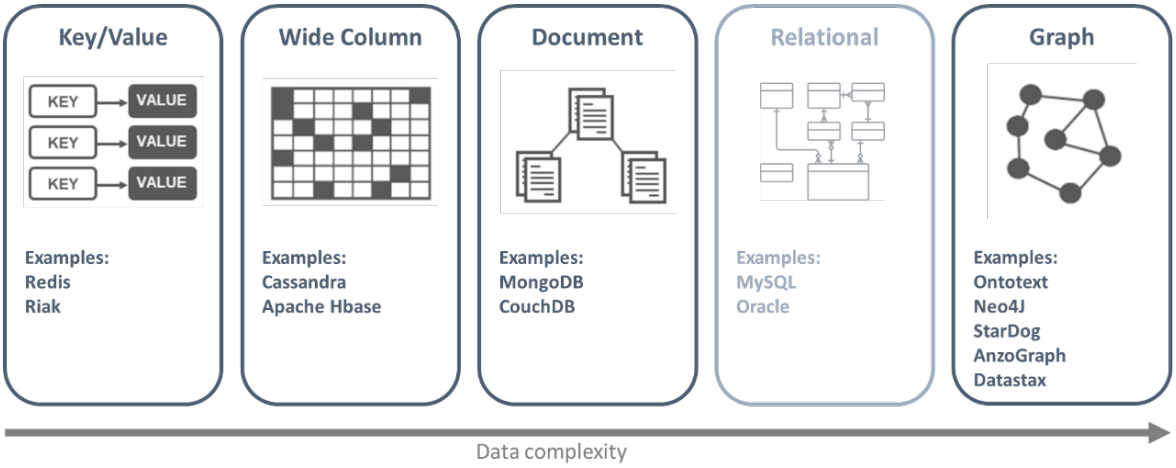


Figure 5.5: Database approaches based on data complexity.

identifying overarching terms commonly found in foundational literature. Subsequently, the process entails systematically representing increasingly intricate entities within the domain. The five general steps to follow when designing a DLO and ALO are [Arp et al., 2015]:

1. Delimit the ontology's subject matter.
2. Collect data: identify the general terminology used in literature and other ontologies.
3. Sort these terms in a hierarchy.
4. Regulate the outcome to ensure (i) logical, philosophical, and scientific coherence; (ii) coherence and compatibility with neighbouring ontologies; and (iii) clarity, mainly through the development of human-readable definitions.
5. Formalise the outcome in a computer-usable language to be implemented in some computable framework.

5.3.4 Data

In this phase, a data management strategy is selected to ascertain how to store the data, which kind of databases need to be used, how to send the data into those databases, and where to store those databases (locally or in the cloud). Moreover, this step includes modelling the data, which consists of transforming raw data into the knowledge graph using the ontology vocabulary developed in the previous step (subsection 5.3.3).

The ontology maps human expertise and provides a formal way to express the knowledge of the system of interest. It establishes common semantics throughout the system of systems environment across its entire life cycle by using a TLO approach. The TLO connects all elements of the system, linking various parts and their temporal aspects. One of the ontology elements is a map that represents the relationships between objects, typically in the form of a

graph. From a data management perspective, data might enter the system in different formats and be managed with different strategies depending on the data's complexity. Various types of databases that handle different levels of data complexity are illustrated in Figure 5.5, with the non-relational database systems highlighted in dark blue [Robinson et al., 2015].

Data complexity is assessed in terms of size, structure, variety, and abstraction [D'Amico et al., 2023]. It is important to note that graph databases may not always be the most effective solution due to this approach's complexity and overhead. The optimal solution could be using a combination of databases, with each approach being better suited for handling specific areas of data complexity. Typically, a relational database is preferable when dealing with entities, whereas a graph database is better suited for examining the relationships between entities. Further information on these approaches can be found in the literature, such as Bechberger and Perryman [Bechberger and Perryman, 2020].

The graph database is the perfect fit for the user to shape the data based on an ontology. Furthermore, some advantages of graph databases are [Robinson et al., 2015]:

- Flexible data modelling: Graph databases allow for flexible and dynamic data modelling that can easily handle complex, interconnected data.
- High performance for complex queries: Graph databases are designed for complex queries, which can be executed quickly, even on large datasets.
- Scalability: Graph databases can easily scale horizontally by adding more nodes to the cluster.
- Better handling of unstructured data: Graph databases are good at handling unstructured data like social media data or IoT data.

Applications where graph databases are more efficient than relational databases are [Bechberger and Perryman, 2020]:

- Fraud or anomaly detection
- Knowledge graph
- Network/asset monitoring
- Identity and access management
- Social networks
- Machine learning (pathfinding, inferred properties, etc.)

In this methodology step, the applicability criteria revolve around selecting an appropriate data management strategy, which is crucial for determining storage methods, database types,

transmission techniques, and whether data should be stored locally or in the cloud. This phase also includes data modelling, transforming raw data into a knowledge graph using the ontology vocabulary developed in an earlier phase. The complexity of the data, measured by size, structure, variety, and abstraction, dictates the choice of database. While graph databases offer flexibility, high performance for complex queries, scalability, and adeptness with unstructured data, they are not universally optimal due to their complexity and overhead. A blend of database types is sometimes more effective, with relational databases favoured for entity management and graph databases for analysing relationships. This decision-making process is also shaped by the data's nature and application areas, with graph databases excelling in fields such as fraud detection and network monitoring. Thus, this step involves a nuanced assessment of data characteristics and system requirements to identify the most suitable data management and modelling approach.

5.3.5 Analysis

In this phase, the purpose that led to the development of the CDT is converted into a validation strategy. Figure 5.6 illustrates a conceptual framework that connects the physical and digital realms. It is partitioned into two principal sections: "Physical World" and "Digital World", demarcated by a dotted line. On the left, under "Physical World", there is an "Entity" linked to a "User" and "Sensors". The entity and sensors connect to a "Database" in the "Digital World" segment, delineated by dotted lines, suggesting data transfer or communication pathways. Within the "Digital World", several components are interlinked:

- "Ontology" is linked to "KG", denoting Knowledge Graph, and to "CDT," standing for Cognitive Digital Twin.
- "Database" is connected to "Analytics", which in turn links to both "Model" and "DT", representing Digital Twin.
- "Model" has connections to "Algorithms".

This configuration indicates a system wherein sensors gather data from the physical environment, which is then stored in a database. This information is analysed and utilised to develop models and algorithms that underpin the DT. The CDT is a more sophisticated iteration that integrates a KG and ontology for enhanced data analysis and decision-making processes. The intent is to clarify the roles of various components in developing a CDT and how it differs from a standard DT.

In the "Physical World", the tangible entity is to be mirrored digitally, which could be an asset, a system, or an organisation. The users and sensors serve as interfaces, collecting data and insights to be transmitted into the "Digital World".

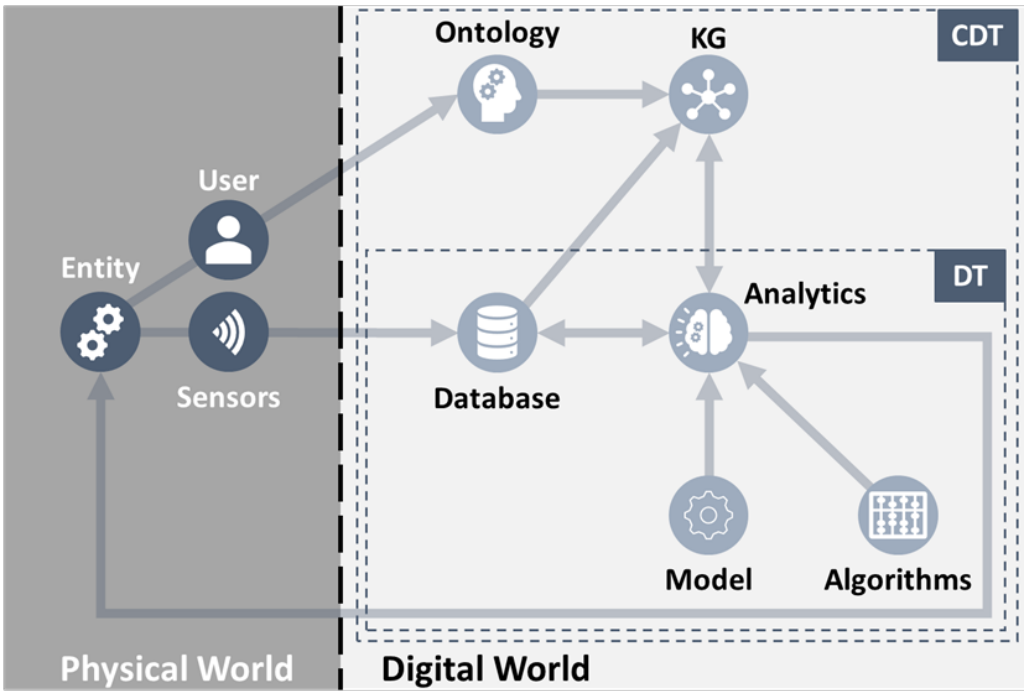


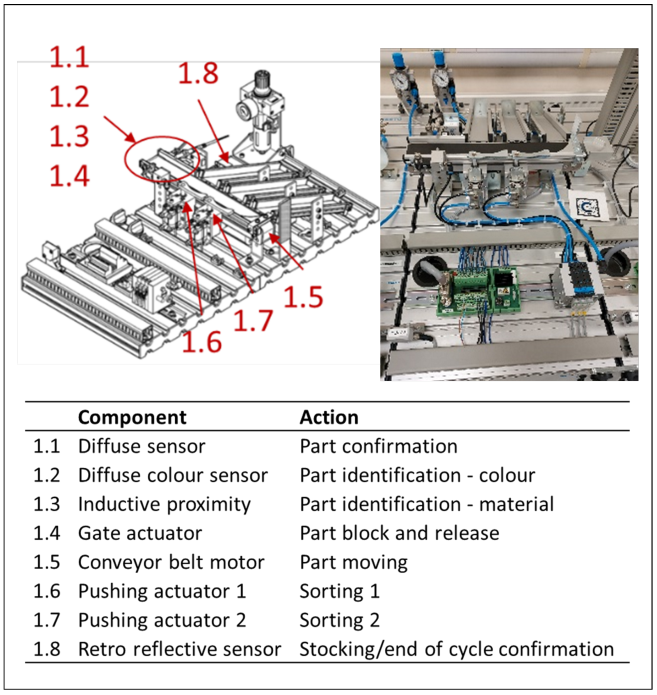
Figure 5.6: Relation between the components of the CDT.

Within the ”Digital World”, the data gathered by the sensors is held within a database structure. Simultaneously, the insights collected from user and entity interactions are organised within an ontology. This ontology encapsulates the knowledge and shapes the KG, into which the data is assimilated.

The ”Analytics” component illustrates the core of the CDT’s capabilities. Herein, the CDT analyses data, executes algorithms, and conducts analytics within the KG. This process entails comparing the digital model to its physical counterpart as represented in the CDT and supplying feedback directly to the entity (to refine operations) or the user (to support decision-making).

This procedure is reported in Figure 5.6, which also visually differentiates the CDT from the DT. The CDT is an enhanced system because it can create a digital representation of a physical entity and comprehend and use the knowledge concerning the entity to enhance operations and support decision-making. This contrasts with a DT, which may not possess the same level of cognitive functionality and typically functions primarily as a digital duplicate, lacking advanced analytical and decision-support capabilities.

In this step of the methodology, the applicability criteria revolve around converting the development purpose of the CDT into a validation strategy. This involves linking an entity in the physical world, such as an asset or system, with users and sensors to gather data. This data is then transferred to the digital realm, where it is stored in a database and integrated into an ontology and KG. The CDT’s role is to compare the digital model with its physical counterpart and provide feedback for enhancing operations and decision-making, highlighting its advanced analytical and decision-support capabilities over the DT.



	Component	Action
1.1	Diffuse sensor	Part confirmation
1.2	Diffuse colour sensor	Part identification - colour
1.3	Inductive proximity	Part identification - material
1.4	Gate actuator	Part block and release
1.5	Conveyor belt motor	Part moving
1.6	Pushing actuator 1	Sorting 1
1.7	Pushing actuator 2	Sorting 2
1.8	Retro reflective sensor	Stocking/end of cycle confirmation

Figure 5.7: Festo Sorting Station, schema, components, actions.

5.4 The use case: Festo sorting station

In this section, the methodology described in section 5.3 above is validated using the Festo Sorting Station module, a part of the Modular Production System (MPS). This asset has been chosen as a testbed because it is an excellent example of a material handling system that can represent multiple real-world applications, such as warehouses, baggage handling systems, etc. Additionally, the asset is reprogrammable according to the application of interest and can also be used to import the human experience into the CDT. This example represents an attempt to validate the methodology mentioned above. The main focus was on developing the framework, the graph database, and the import of human expertise. The CDT worked well for the characteristics assessed and responded well to detect malfunctioning components. However, some aspects need further investigation and are highlighted in the next section (Section 5.5) under the limitations paragraph.

5.4.1 Requirements analysis

Initially, the asset has been studied and reverse-engineered. Classification of the components, mereotopology, relations between artefacts, functions, and capabilities have been carried out. The classification of the components of interest and their actions is reported in Figure 5.7. The working procedure for the Festo Sorting Station is described below:

1. Upon entry of the workpiece into the sensing module, the data is transmitted to the CDT. The CDT then pairs the workpiece ID and the spatial area with the provided timestamp.

2. The sensing module determines the workpiece's colour.

3. Once the colour is identified, a command is sent to the gate actuator to release the workpiece.

4. As the gate actuator opens, the colour findings from the sensing phase and the corresponding timestamp are conveyed to the CDT. The CDT then links the workpiece ID, spatial area, and colour data with the provided timestamp.

5. The primary actuator is engaged if the workpiece appears red, directing the piece to the initial collection point. Conversely, a metallic workpiece triggers the secondary actuator, directing it towards the second collection point. When the workpiece is black, no actuator is engaged. Instead, a physical barrier guides the black workpiece via the conveyor system to the third collection point. For such cases, data points for spatial zones are matched with the appropriate workpiece ID in the CDT.

6. The data is relayed to the CDT whenever the retro-reflective sensor identifies a workpiece. The CDT then pairs the timestamp and spatial area with the corresponding workpiece ID.

Considering the list of components reported in Figure 5.7, the Festo sorting station system has been then divided into three subsystems: the red sorting station (Red SS), the metal sorting station (Metal SS), and the black sorting station (Black SS). The Red SS includes the conveyor, the red transferring unit (Red TU), together with the sensing unit (SU) and the first actuator (A1). Similarly, the Metal SS also includes the conveyor and the Metal TU, with the associated SU and the second actuator (A2). The Black SS contains the conveyor and the SU.

Regarding functions and capabilities, as per 5.3.1, the functions are capabilities that specifically highlight the reason why an entity exists. Therefore, only the functions need to be specified for this application; this paper does not report the connections between the components' capabilities and their functions. The artefact compositions and functional composition of the system are presented in Figure 5.8.

The AD theory has been applied to verify the mechanical design and delineate how the functions are realised at every level of the artefacts' decomposition. As represented in the artefacts' decomposition, the sorting station comprises a "SummationJunction" made up of the 'Black SS', 'Red SS', and 'Metal SS'. The first two units function as a "SummationJunction" combining a section of the 'conveyor belt' with a 'transferring unit'. Furthermore, the 'transferring unit' operates as a "ControlJunction", where the 'actuators' are governed by the 'sensing unit'. It is important to note that the final slope lacks an actuator, as reported in Figure 5.8. The AD theory allows building matrices and verifies the relations between artefacts and their functions for each uncoupled (SummationJunction 5.2) and decoupled (ControlJunction 5.3) case in the system. The theory is explained by Suh [Suh, 1998] and is not included in this paper for brevity. The resulting matrices are reported in Figure 5.9.

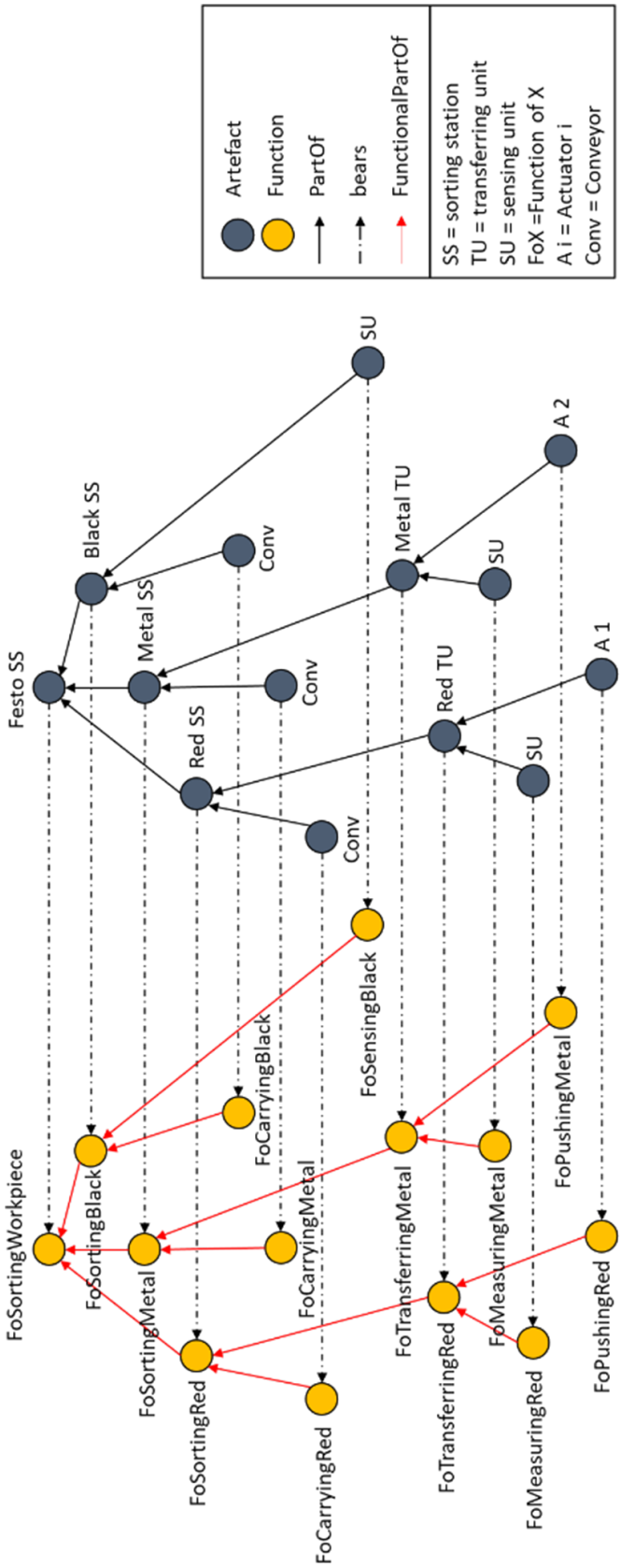


Figure 5.8: Mereotopological representation of the system along with functions.

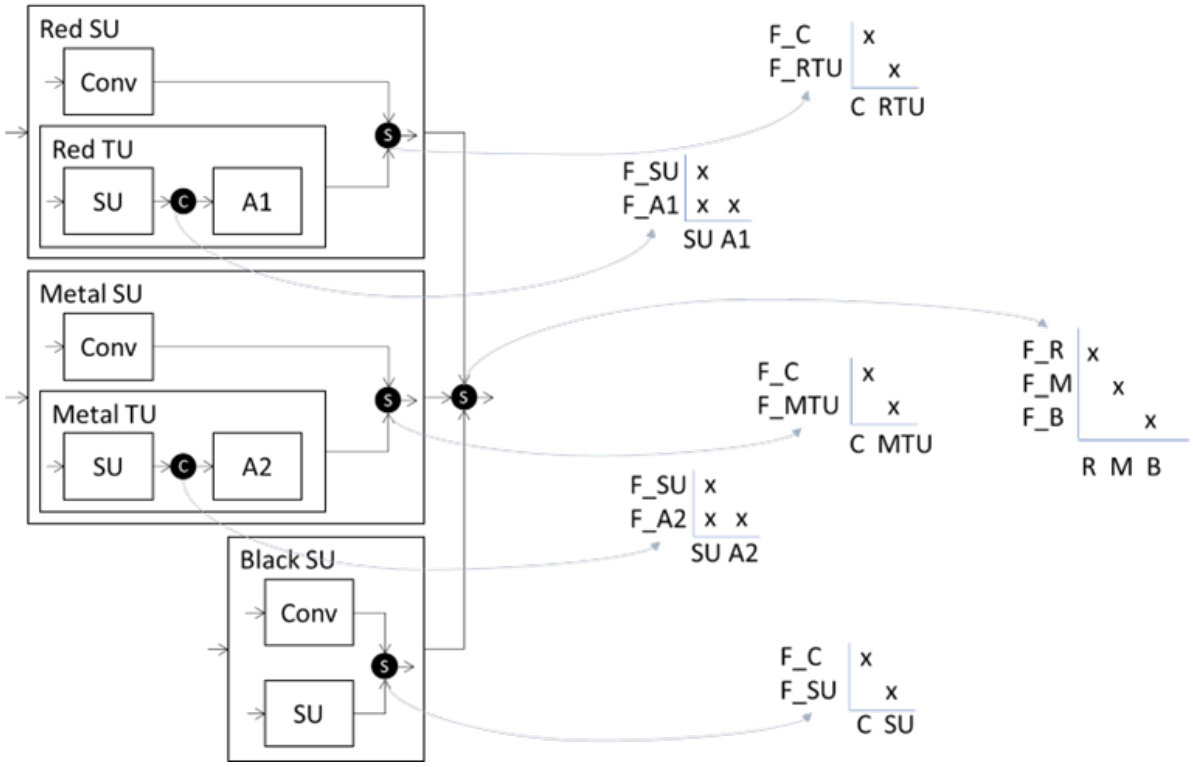


Figure 5.9: Modulo junction diagram from AD theory.

5.4.2 Knowledge gathering

This section presents all the relevant processes and knowledge to detect malfunctions.

Processes

Whenever a certain process occurs, distinct entity types are involved, each performing specific roles. As anticipated in section 5.3.2, to differentiate between the system and the materials it processes, we classify two key thematic roles: the ”instrument” (serving as the direct cause) and the ”patient” (which is affected). In this context, the material in question is the workpiece that the Festo sorting station classifies based on its colour.

The Festo sorting station consists of three main processes. The workpiece functions as the patient in all these processes, though its interaction varies. For instance, as outlined in one of the subsequent axioms, sensors operate to discern colour through a process termed ”ActOfMeasuringColor” 5.4. At the culmination of this process, the colour data of the workpiece is rendered and sent to the system. The processes ”ActOfCarrying” 5.5 and ”ActOfActuating” 5.6 both target the relocation of the workpiece. While ”ActOfCarrying” alters the workpiece’s spatial position as it remains stationary on a moving conveyor belt, ”ActOfActuating” relocates the workpiece from a conveyor site to a collection point, which serves as its terminal storage.

Based on the functional diagram of the Festo Sorting Station provided in Figure 5.8, there exist three distinct ”ActOfMeasuringColor” subtypes: ”ActOfMeasuringBlack”, ”ActOfMea-

suringRed", and "ActOfMeasuringMetal". These are contingent on the colour variations they identify and the associated functions they execute, e.g., realizes(ActOfMeasuringRed, FunctionOfMeasuringRed). There are also three distinct "ActOfCarrying" types, each corresponding to the terminal locations of the workpiece. However, the black workpiece, transported directly to 'collection point 3' by the conveyor, does not require an actuation process. As such, only two "ActOfActuating" subtypes exist: "ActOfActuatingRed" and "ActOfActuatingMetal".

For analytical purposes, understanding which component malfunctions and the structure of these processes is essential. These operations are executed by higher-tier components, which inherently incorporate these processes. For example, the "RedTransferringUnit" functions as the instrument for "ActOfTransferringRed", which inherently consists of processes like "ActOfMeasuringRed" and "ActOfActuatingRed".

$$\begin{aligned}
&\forall p, f\, \text{ActOfMeasuringColor}(p) \wedge \text{realises}(p, f) \leftrightarrow \\
&\quad \text{ActOfMeasurement}(p) \wedge \exists s\, (\text{Sensor}(s) \wedge \text{isInstrumentOf}(s, p)) \wedge \\
&\quad \exists w\, (\text{Workpiece}(w) \wedge \text{isPatientOf}(w, p)) \wedge \\
&\quad \forall t\, (\text{occursOn}(p,t) \rightarrow \exists c, ci\, (\text{Color}(c) \wedge \text{ColorName}(ci) \wedge \text{hasOutput}(p, ci) \wedge \\
&\quad \text{isMeasurementOf}(ci, c) \wedge \exists te\, (\text{hasLastInstant}(t, te) \wedge \text{inheresIn}(c, w, te)))
\end{aligned} \tag{5.4}$$

$$\begin{aligned}
&\forall p, f\, \text{ActOfCarrying}(p) \wedge \text{realises}(p, f) \leftrightarrow \\
&\quad \text{ActofLocationChange}(p) \wedge \exists c\, (\text{Conveyor}(c) \wedge \text{isInstrumentOf}(c, p)) \wedge \\
&\quad \exists w\, (\text{Workpiece}(w) \wedge \text{isPatientOf}(w, p)) \wedge \\
&\quad \forall t\, (\text{occursOn}(p,t) \rightarrow \exists l, s1, s2, tb, te\, (\text{locatedIn}(w, l, t) \wedge \text{partOf}(l, c) \wedge \\
&\quad \text{occupiesSpatialRegion}(l, s1, tb) \wedge \text{occupiesSpatialRegion}(l, s2, te) \wedge \\
&\text{hasFirstInstant}(t, tb) \wedge \text{hasLastInstant}(t, te) \wedge (s1 \neq s2))
\end{aligned} \tag{5.5}$$

$$\begin{aligned}
&\forall p, f\, \text{ActOfActuating}(p) \wedge \text{realises}(p, f) \leftrightarrow \\
&\quad \text{ActofLocationChange}(p) \wedge \exists a\, (\text{Actuator}(a) \wedge \text{isInstrumentOf}(a, p)) \wedge \\
&\exists w\, (\text{Workpiece}(w) \wedge \text{isPatientOf}(w, p)) \wedge \\
&\quad \forall t\, (\text{occursOn}(p,t) \rightarrow \exists l, l', s1, s2, tb, te, n, c\, (\text{Slope}(n) \wedge \text{Conveyor}(c) \wedge \\
&\text{partOf}(l', s) \wedge \quad \text{partOf}(l, c) \wedge \text{occupiesSpatialRegion}(l, s1, tb) \wedge \\
&\text{occupiesSpatialRegion}(l, s1, tb) \wedge \text{hasFirstInstant}(t, tb) \wedge \text{hasLastInstant}(t, te))
\end{aligned} \tag{5.6}$$

∴ 5.4 If an action of measuring colour p results in the realisation of function f, it is the same as an action of measurement p. Additionally, there must be a sensor s that acts as an instrument for p, and there must be a workpiece w that is the patient of p. Furthermore, for all instances of time t when p occurs, there must be a colour c and a corresponding colour name ci. These colour measurements ci are linked to actual colours c and associated with the workpiece w at specific time instances te

∴ 5.5 When an action of carrying p results in the realisation of function f, it is the same as an action of location change p. There must be a conveyor c that p uses as an instrument, and there must be a workpiece w that is the patient of p. Moreover, for all instances of time t when p occurs, there must be a location l, two distinct spatial regions $s1$ and $s2$, time intervals tb and te, and a part l of the conveyor c. The workpiece w must be located in l at time t, and l must occupy spatial regions $s1$ and $s2$ during time intervals tb and te, respectively

∴ 5.6 If an action of actuating p results in the realisation of function f, it is the same as an action of location change p. There must be an actuator a that acts as an instrument for p, and there must be a workpiece w that is the patient of p. Furthermore, for all instances of time t when p occurs, there must be a location l, another location l', two distinct spatial regions $s1$ and $s2$, time intervals tb and te, a slope n, a conveyor c, and a part l of the conveyor. The workpiece w must be part of l at time t, and l and l' must occupy spatial regions $s1$ and $s2$ during time intervals tb and te, respectively. The slope n and the conveyor c must be part of the scenario

Malfunction

If a component fails in a process where it serves as the instrument, resulting in an inability to achieve its designated function, its state is malfunctioning. The repercussions of a component's malfunction can be evident both at the individual component and broader system levels. For example, if Actuator 1 fails to activate, the red workpiece might mistakenly be directed to collection point 3. Similarly, a sensing unit failure could misdirect a Metal workpiece into collection point 1. One can employ a series of rules to pinpoint a system-level error down to a specific component that aligns with the system's modulo-junction structure, as shown in Figure 5.9.

We introduce the predicate *"malfunctionedFor"*. This predicate establishes a relationship between a component, the specific instance of a process where the error transpired, and the workpiece the component unsuccessfully managed. This means that the process, where the workpiece was the patient and the component functioned as the instrument, did not deliver the component's intended function.

To illustrate the application of this methodology, four rules geared towards identifying errors when sorting red workpieces have been outlined. Note that different error types could necessitate their distinct rule sets.

$$\begin{aligned}&\text{malfunctionedFor}(c,p,w) \wedge \text{RedSortingUnit}(c) \wedge \text{ActOfSortingRed}(p) \wedge \\ &\text{Workpiece}(w) \wedge \text{Conveyor}(c') \wedge \text{ActOfCarryingRed}(p') \wedge \\ &\text{isPatientOf}(w,p') \wedge \text{isInstrumentOf}(c',p') \wedge \\ &\neg\text{realizes}(p',\text{FunctionOfCarryingRed}) \rightarrow \text{malfunctionedFor}(c',p',w)\end{aligned} \tag{5.7}$$

$$\begin{aligned}&\text{malfunctionedFor}(c,p,w) \wedge \text{RedSortingUnit}(c) \wedge \text{ActOfSortingRed}(p) \wedge \\ &\text{Workpiece}(w) \wedge \text{RedTransferringUnit}(c') \wedge \text{ActOfTransferringRed}(p') \wedge \\ &\text{isPatientOf}(w,p') \wedge \text{isInstrumentOf}(c',p') \wedge \\ &\text{Conveyor}(c'') \wedge \text{ActOfCarryingRed}(p'') \wedge \\ &\text{isPatientOf}(w,p'') \wedge \text{isInstrumentOf}(c'',p'') \wedge \\ &\text{realizes}(p'',\text{FunctionOfCarryingRed}) \rightarrow \text{malfunctionedFor}(c',p',w)\end{aligned} \tag{5.8}$$

$$\begin{aligned}&\text{malfunctionedFor}(c,p,w) \wedge \text{RedTransferringUnit}(c) \wedge \text{ActOfTransferringRed}(p) \wedge \\ &\text{Workpiece}(w) \wedge \text{SensingUnit}(c') \wedge \text{ActOfMeasuringRed}(p') \wedge \\ &\text{isPatientOf}(w,p') \wedge \text{isInstrumentOf}(c',p') \wedge \\ &\neg\text{realizes}(p',\text{FunctionOfMeasuringRed}) \rightarrow \text{malfunctionedFor}(c',p',w)\end{aligned} \tag{5.9}$$

$$\begin{aligned}&\text{malfunctionedFor}(c,p,w) \wedge \text{RedTransferringUnit}(c) \wedge \text{ActOfTransferringRed}(p) \wedge \\ &\text{Workpiece}(w) \wedge \text{Actuator}(c') \wedge \text{ActOfActuatingRed}(p') \wedge \\ &\text{isPatientOf}(w,p') \wedge \text{isInstrumentOf}(c',p') \wedge \neg\text{realizes}(p',\text{FunctionOfPushingRed}) \wedge \\ &\text{SensingUnit}(c'') \wedge \text{ActOfMeasuringRed}(p'') \wedge \\ &\text{isPatientOf}(w,p'') \wedge \text{isInstrumentOf}(c'',p'') \wedge \\ &\text{realizes}(p'',\text{FunctionOfMeasuringRed}) \rightarrow \text{malfunctionedFor}(c',p',w)\end{aligned} \tag{5.10}$$

$\therefore$ 5.7 If a sorting unit c malfunctions for an action of sorting red p concerning a workpiece w, and c is a red sorting unit, and p is an action of sorting red related to w, and there exists

another conveyor c' for carrying red p' related to w, and p' is an action of carrying red related to c', and c' is an instrument for p', and p' does not realise the function of carrying red, then the malfunction also applies to c' for the action of carrying red p' concerning w

∴ 5.8 If a sorting unit c malfunctions for an action of sorting red p concerning a workpiece w, and c is a red sorting unit, and p is an action of sorting red related to w. There exists a red transferring unit c' for transferring red p' related to w, and p' is an action of transferring red related to c', and c' is an instrument for p', and there exists a conveyor c'' for carrying red p'' related to w, and p'' is an action of carrying red related to c'', and p'' realises the function of carrying red. The malfunction also applies to c' for the action of transferring red p' concerning w

∴ 5.9 If a sorting unit c malfunctions for an action of sorting red p concerning a workpiece w, and c is a red transferring unit, and p is an action of transferring red related to w, and there exists a sensing unit c' for measuring red p' related to w, and p' is an action of measuring red related to c', and c' is an instrument for p', and p' does not realise the function of measuring red, then the malfunction also applies to c' for the action of measuring red p' concerning w

∴ 5.10 If a sorting unit c malfunctions for an action of sorting red p concerning a workpiece w, and c is a red transferring unit, and p is an action of transferring red related to w. There exists an actuator c' for actuating red p' related to w, and p' is an action of actuating red related to c', and c' is an instrument for p', and p' does not realise the function of pushing red. There exists a sensing unit c'' for measuring red p'' related to w, and p'' is an action of measuring red related to c'', and c'' is an instrument for p'', and p'' realises the function of measuring red, then the malfunction also applies to c' for the action of actuating red p' and to c'' for the action of measuring red p'' concerning w

Rule 1 5.7 states that if there is an identified failure with the RedSortingUnit (a preliminary fact signifying a red workpiece sorting error), the fault is ascribed to the Conveyor when the related event, ActOfCarryingRed, does not execute properly. The negation of ”realizes” indicates that the desired outcome depicted in the right part of 5.5 is not met, meaning the required output state is not achieved.

If Rule 1 5.7 is not activated, implying that the Conveyor is functioning as expected, then Rule 2 5.8 attributes the malfunction to the RedTransferringUnit. If the RedTransferringUnit is at fault, Rule 3 5.9 and Rule 4 5.10 are employed to determine if the SensingUnit or RedTransferringUnit, respectively, are the root causes of the malfunction.

As the SensingUnit is the control mechanism for the RedTransferringUnit, it is not necessary to carry on with the rules and verify if the Actuator also malfunctions (as per 5.4).

Conversely, the core malfunction can only be linked to the Actuator if the ActOfActuatingRed does not execute (meaning it does not fulfil the Actuator's function), while the prior process, ActOfMeasuringRed, works as expected (meaning it achieves the SensingUnit's function).

5.4.3 Ontology development

The first step to creating an application ontology, as per 5.3.3, is to delimit the subject or formalise the ontology's purpose. In order to formalise the purpose, a set of CQs have been created to formalise the pieces of information that the CDT can manage. The CQs created for this use case are the following:

- CQ1. What are the components of the Festo sorting station?
- CQ2. How long does the conveyor take to carry the workpiece X?
- CQ3. What is the time that the workpiece X entered the sensing unit?
- CQ4. What is the colour of workpiece X sensed by the sensing unit?
- CQ5. How many workpieces are at the collection point X at time T?
- CQ6. Is the workpiece X sensed by the sensing unit within the time interval T?

The next step is to develop the terminology based on the terms used in the CQs (as per 0 and 0) and the terms used in the existing ontologies. Regarding the latter, existing ontologies are TLOs and MLOs, which are widely recognised among the scientific and industrial community. The authors, following careful research, have chosen the ontology model for the Festo sorting system by extending the predicates from the Common Core Ontology (CCO) (available at `https://www.cubrc.org`, accessed on 01/08/2023) and IOF-Core [Karray et al., 2021]. These ontologies are grounded in a TLO known as the Basic Formal Ontology (BFO) (published by Smith et al. [Smith et al., 2007]). Given BFO's recognition as an ISO standard (ISO 21838-2:2021), employing it ensures that the application ontology is both universal and interoperable.

A few examples of terminologies inherited from those TLOs and MLOs are terms such as "bears," "realizes," "participatesIn," "OccurrentPartOf," "occursOn," "hasFirstInstance," and "hasLastInstance" and are derived from BFO. Terms such as "hasOutput," "ActOfLocationChange," and "ActOfMeasurement" are borrowed from CCO. The term "System" is sourced from IOF-Core. The authors ensure a rigorous and comprehensive approach to structuring and detailing the Festo sorting system's ontology model by aligning with these foundational ontologies.

The following step is to create the taxonomy, starting from the identified TLOs and the MLOs. The ontology is created using Protégé (`https://protegewiki.stanford.edu/`

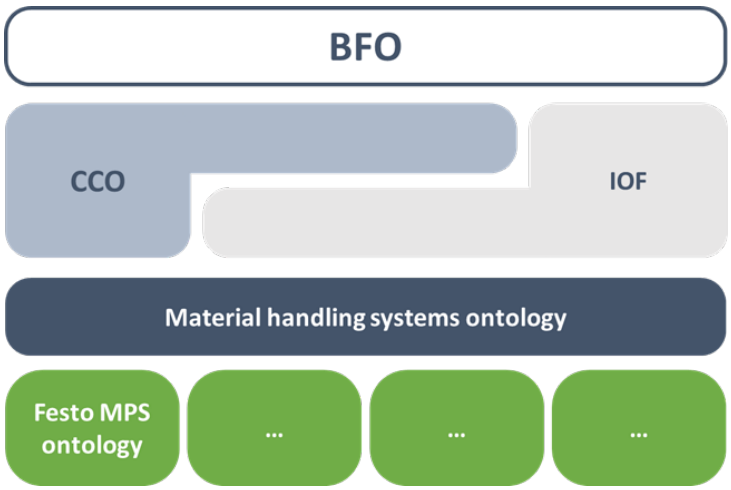


Figure 5.10: Ontology classification based on the level of abstractions.

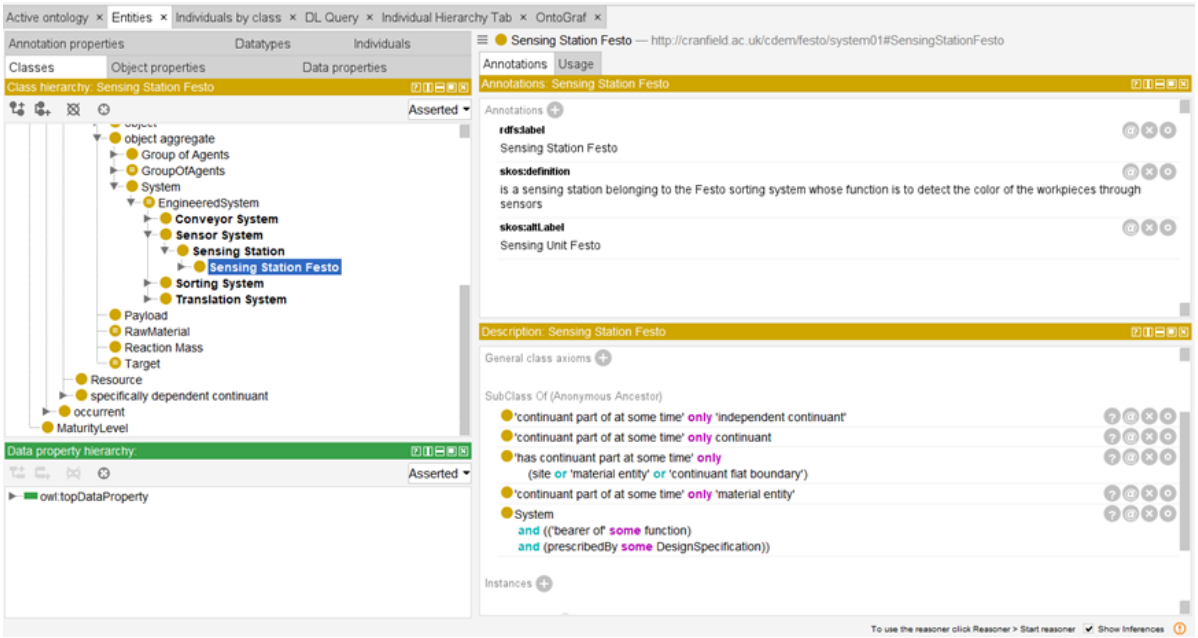


Figure 5.11: Screenshot of the Protégé tool focusing on the class Sensing Station Festo.

`wiki/Main_Page`, accessed on 01/08/2023). This tool allows the importation of existing ontologies (TLO and MLOs) and creates the new DLO and ALO, extending the previous ones and exporting them into a file that formalises the outcome in a computer-usable language. Moreover, it is possible to reuse and create data and object properties and annotations about the new class using Protégé. The schema of the ontologies implemented is reported in Figure 5.10.

Starting from BFO, CCO, and IOF-Core, a DLO and ALO have been created, as reported in Figure 5.14. BFO serves as the TLO, ideally suited for use across various domains, while CCO and IOF-Core are MLOs that encompass terminologies and properties more closely aligned with the manufacturing and industrial sectors.

In the model representation of the ontology (Figure 5.14), DLO represents the domain-level ontology or, rather, the ontology that can be applied in the material handling systems domain,

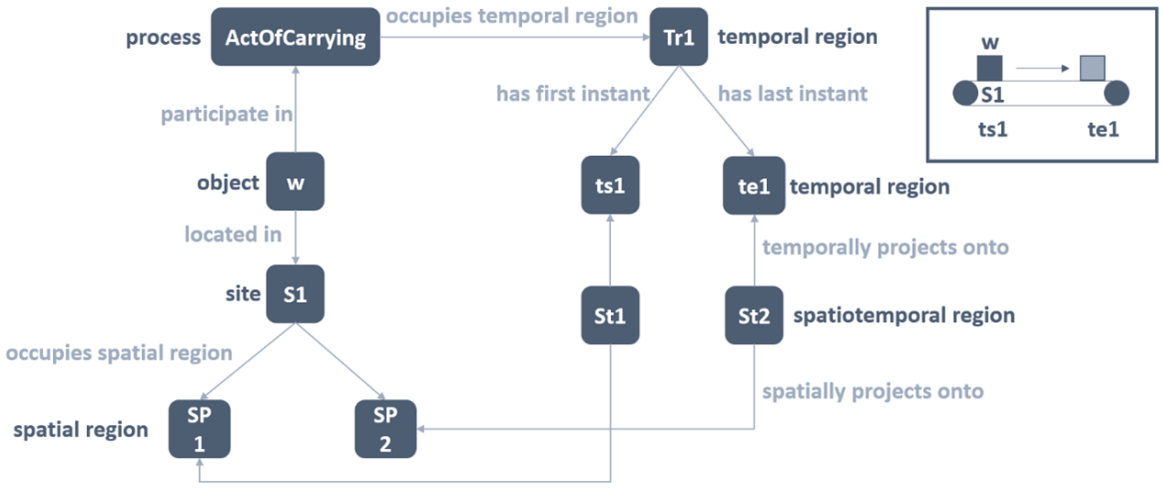


Figure 5.12: Diagram that represents the process ActOfCarrying using BFO properties.

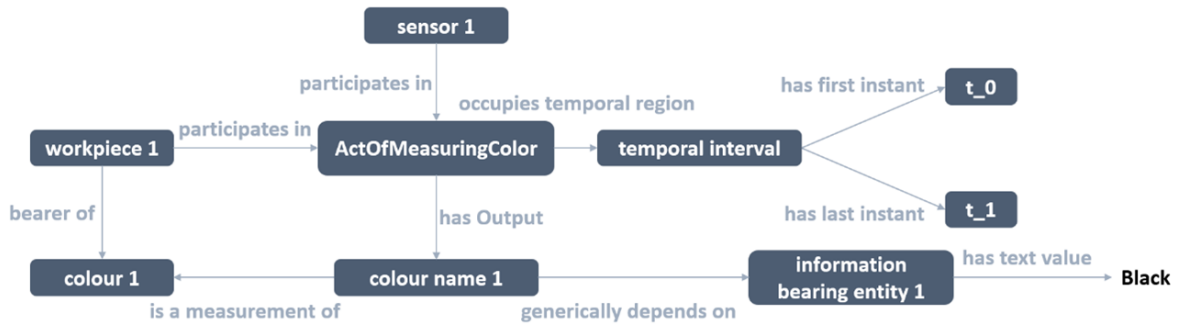


Figure 5.13: Diagram that represents the process ActOfMeasuringColor using BFO and CCO properties.

while ALO represents the application ontology or, rather, the ontology developed for this specific Festo sorting station module. Each class has been defined mainly using two methodologies for the definitions, i.e., the Aristotelian and the non-root terms one [Arp et al., 2015]. An example is reported in Figure 5.11.

The processes can be expressed using the BFO properties. An example is the process ActOfCarrying 5.5, as reported in Figure 5.12. The process represents the workpiece moving in the conveyor belt at a specific time range. The properties in light blue are taken directly from BFO, considering the domain and range for each property, whereas the classes in dark blue are the parent classes belonging to BFO, to reflect the domain and range of each property. Another example is the process ActOfMeasuringColor 5.4, reported in Figure 5.13, represented using the properties and classes from BFO and CCO. In particular, the connection between the literal (string: “Black”) and the instance (“measurement colour 1”) is made through the information-bearing entity class of CCO.

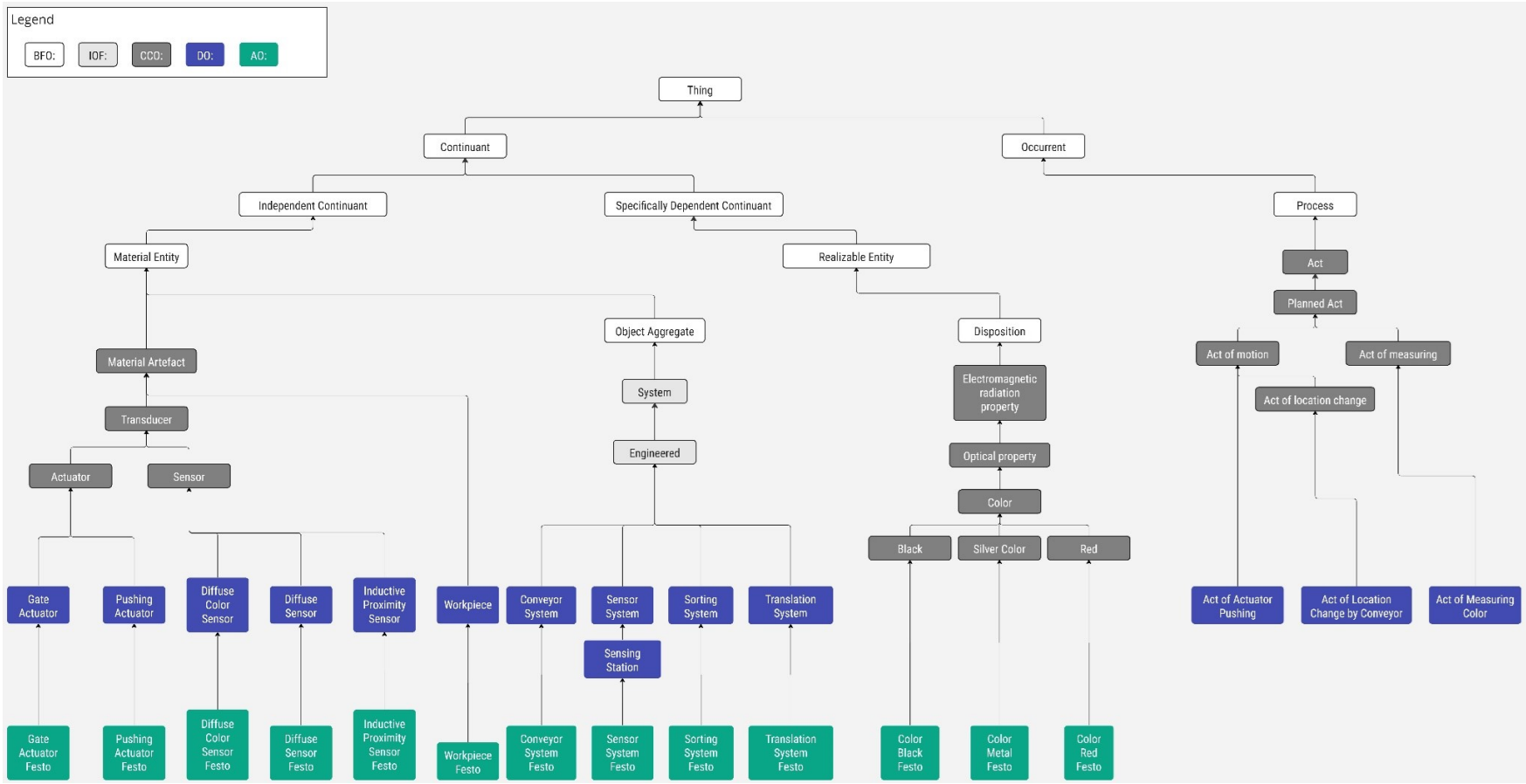


Figure 5.14: Model representation of the ontology.

Table 5.2: Colour detection from sensors' output combination. (DS = Diffuse sensor, DCS = Diffuse colour sensor, IPS = Inductive proximity sensor)

Colour	DS	DCS	IPS
Black	1	0	0
Red	1	1	0
Metal	1	1	1

5.4.4 Data

The outcome of this phase is a file in the Web Ontology Language (OWL) format. The OWL file created will then be imported into the data management tool and all the other TLOs and MLOs used. The data is collected and managed using the same data management tool. The data management tool adopted for this project is GraphDB by Ontotext. This open-source graph database software allows the importation of multiple ontologies and the receipt of data through an HTTP connection. SPARQL is the query language used within this environment. Data has been included in SPARQL queries and used to send information about the structure and functions of the system and its components, the related processes, and the reasoning for the malfunctions, as reported in Figure 5.15.

The sensing unit is composed of three sensors: i) a diffuse sensor (DS), ii) a diffuse colour sensor (DCS), and iii) an inductive proximity sensor (IPS). The combination of those sensors' output determines the workpiece's colour, as reported in Table 5.2.

An example of the query to import the data into the KG is reported in Figure 5.17. The numbers from the sensors will be placed where the red underlined placeholders are. This query is enclosed in a script sent via an HTTP transmission to the CDT. Data is transferred into the software (GraphDB) shaped as a knowledge graph. The extraction of the KG is reported in Figure 5.16. The ontology imported in the KG shapes the data. All the instances are then appended to the relative object imported from the ontology, as reported in Figure 5.16 (right).

In this paper, the only database used is the KG. Other kinds of databases (Figure 5.5), such as relational databases, can be used in association with the KG but are not considered for this project, as they are not needed in this particular case.

5.4.5 Analysis

Competency questions

One of the methods to verify the operation of the CDT, especially the ontology, is to query the CQs to the CDT. The six CQs are converted into SPARQL queries and launched into the GraphDB tool. The six CQs are listed below; for brevity, only the first and the last ones include the SPARQL query.

CQ1. What are the components of the Festo sorting station?

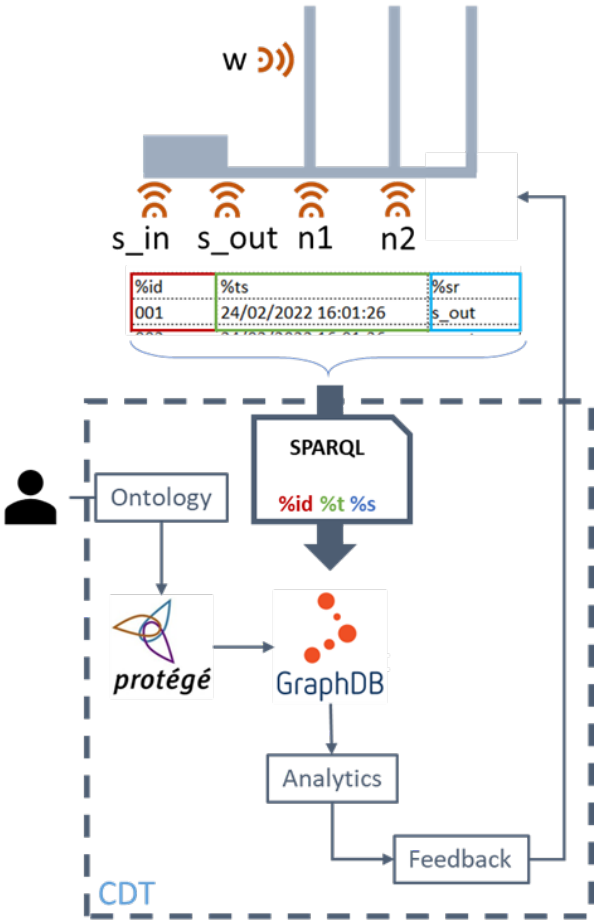


Figure 5.15: Data flow in the CDT.

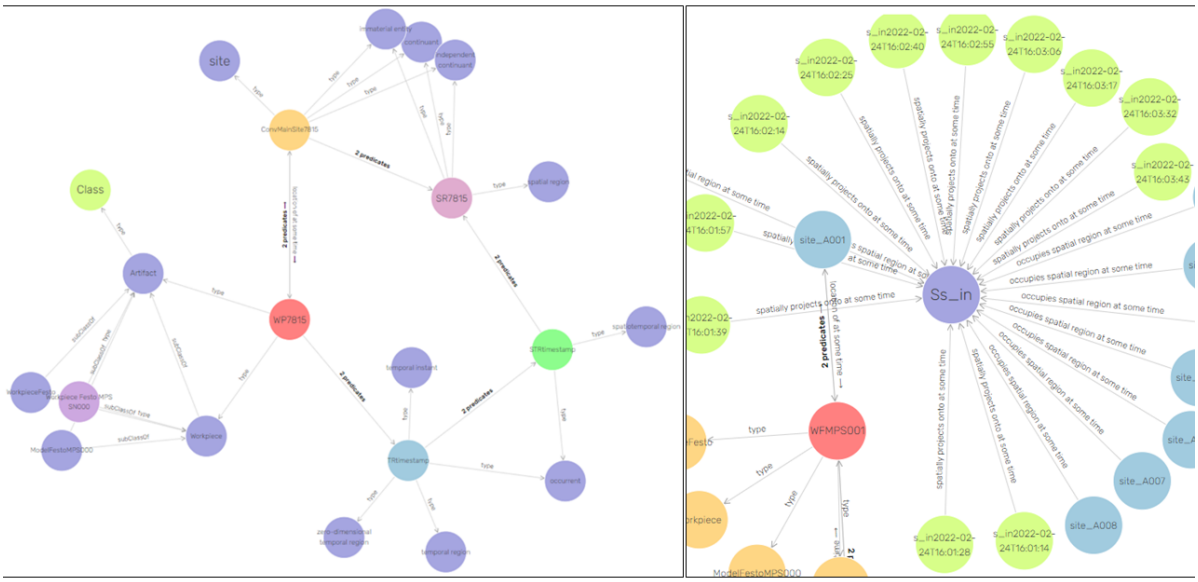


Figure 5.16: Snapshot of the KG populated. Part of the ontology (left), and data population containing data related to workpiece ID001 in space $s_in(right)$.

```
INSERT{
    ?WP bfo:BFO_0000108 _:b1; #exists at ?t
        bfo:BFO_0000082 ?site. #located in at all times ?site
    ?site bfo:BFO_0000210 ?s. # site Occupies Spatial Region
    _:b1 a  bfo:BFO_0000203; #?t a temporal instant
         cco:designated_by _:b2. #ice_timeValue
    _:b2 a  cco:TemporalInstantIdentifier;
         ro:RO_0010001/cco:has_datetime_value   ?time.
    ?st  a  bfo:BFO_0000011; #?st1 a spatiotemporal region
         bfo:BFO_0000153 _:b1; #st1 temporally projects onto t
         bfo:BFO_0000216 ?s. # st1 spatially project onto sp
}
  WHERE{
      BIND(IRI(CONCAT(STR(fes:WFMPS), STR(?WPID))) as ?WP)
      BIND(%id as ?WPID)
      BIND(IRI(CONCAT(STR(fer:site_A), STR(?WPID))) as ?site)
      BIND(IRI(CONCAT(STR(fer:S), %sr)) as ?s)
      BIND(%tr as ?time)
      BIND(IRI(CONCAT(STR(fer:s_out),STR(?time)))  as ?st)
  }
```

Figure 5.17: Example of SPARQL query to send data point to the KG.

```
SELECT DISTINCT ?cl ?c1
WHERE {
    ?c fes:hasSubsystemOf* ?c1.
    ?c1 rdfs:label ?cl.
    FILTER(?c = fes:SOUFSS000)
}
```

	cl
1	"Sorting Unit Festo Sorting Station SN000"
2	"Translation Unit Festo Sorting Station SN000"
3	"Sensing Unit Festo Sorting Station SN000"
4	"Conveyor Festo Sorting Station SN000"
5	"Pushing Actuator Festo Sorting Station SN000"
6	"Diffuse Color Sensor Festo Sorting Station SN000"
7	"Diffuse Sensor Festo Sorting Station SN000"
8	"Inductive Proximity Sensor Festo Sorting Station SN000"
9	"Gate Actuator Festo Sorting Station SN000"

CQ2. How long does the conveyor take to carry the workpiece X?

	w	p	time1	time2
1	:WFMPS001	_:node26516	"2022-02-24T16:01:26"	"2022-02-24T16:01:29"

CQ3. What is the time that the workpiece X entered the sensing unit?

	w	time	s
1	:WFMPS001	"2022-02-24T16:01:14"	http://cranfield.ac.uk/cdem/festo/Ss_in

CQ4. What is the colour of workpiece X sensed by the sensing unit?

	w	s	color
1	:WFMPS001	http://cranfield.ac.uk/cdem/festo/Ss_out	"red"

CQ5. How many workpieces are at collection point X at time T?

	result
1	"12"^^xsd:integer

CQ6. Is workpiece X sensed by the sensing unit within the time interval T?

```
ASK {
    ?p  a
fes:ActOfMeasuringColor;
            bfo:BFO_0000057 ?w;
#process has participant
```

Run

keyboard shortcuts

Query took 0.1s, moments ago.

YES

Detection of malfunction

A test scenario has been presented to showcase the reasoning process for identifying a malfunctioned component in the event of a system error. In this scenario, a discrepancy is observed at collection point 1, where the actual number of workpieces present (CQ 5) is fewer than the expected quantity. Consequently, the comparison script flags a system-level error, as reported in Figure 5.18. Specifically, the workpiece with ID 004 is identified as one of the missing items from the initial collection point.

In order to perform this comparison script, the malfunction function has been created 6.1. The malfunction function has three parameters: the component (instrument) c, the process p, and the workpiece (patient) w.

$$malfunctionedFor(c,p,w) \tag{5.11}$$

The function with three parameters cannot be managed in OWL. Hence, reification has been applied to adopt this function. This procedure decomposes the function into three functions with two parameters each: malfunctionedComponent, malfunctioningProcess, and ma-

	RedSortingUnit	ActOfSortingRed	workpiece	malfunction
1	_:node26827	_:node26826	:WFMPS004	_:node26828

Figure 5.18: Malfunctioning of the red sorting unit.

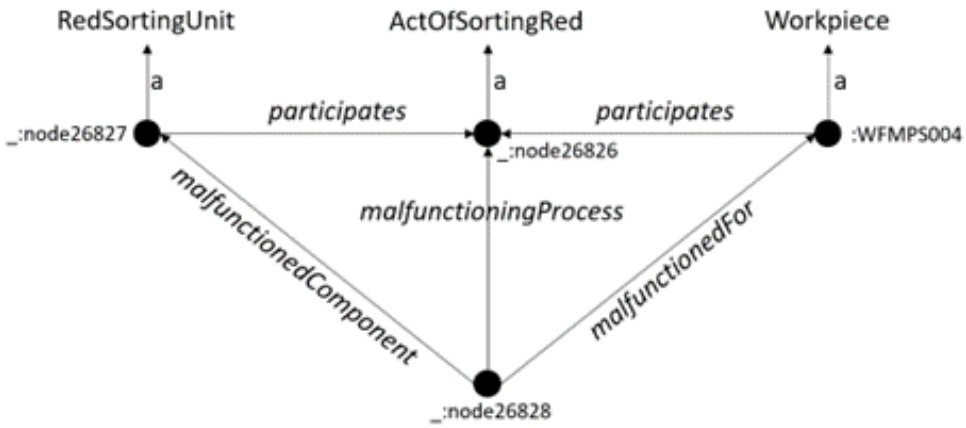


Figure 5.19: Schema for the reification.

functionedFor. The schema (reported in Figure 5.19) for the component, the process, and the workpiece has been created, reporting the values found in the result of the query in Figure 5.18.

A component malfunctions if the process in which the component plays the role of the instrument fails to realise its corresponding function. An error/malfunction at the system level can be narrowed down to a component by applying the rules 5.7 to 5.10 in order to find the process in which a particular workpiece was a patient and the component an instrument did not realise the function of the component.

The rules are contained in the ontology in the form of Semantic Web Rule Language (SWRL), inserted in Protégé after completing the ontology. SWRL is a rule language for the Semantic Web that extends the OWL. It allows users to specify rules in the form of logical expressions that can be used to make inferences about the data in an OWL ontology. SWRL rules are written using a combination of OWL constructs and a rule syntax that allows the creation of conditional expressions, using constructs such as ”if-then” statements, ”and” and ”or” operators, and mathematical comparisons. SWRL rules can be used to automate reasoning tasks and to infer new knowledge from existing data.

In this instance, upon applying the rules outlined from 5.7 to 5.10 to the KG, rule 5.8 is activated. This occurs because the process ActOfCarryingRed successfully fulfils the Conveyor’s function, as detailed by the state description in Figure 5.20 (specifically, Workpiece 004 is transported from s_{out} to w). As a result of rule 5.8 being triggered, the assertion is made that the RedTransferringUnit is experiencing a malfunction. The RedTransferringUnit is the instrument of some ActOfTransferringRed, for which some ActOfMeasuringRed, ActOfCarryingRed, and ActOfActuatingRed are occurrent processes.

Following this assertion, only rule 5.9 is triggered. This is due to the discovery that the out-

	Workpiece	Time	Location	Site
1	:WFMPS004	"2022-02-24T16:01:57"	http://cranfield.ac.uk/cdem/festo/Ss_in	http://cranfield.ac.uk/cdem/festo/site_A004
2	:WFMPS004	"2022-02-24T16:02:12"	http://cranfield.ac.uk/cdem/festo/Ss_out	http://cranfield.ac.uk/cdem/festo/site_A004
3	:WFMPS004	"2022-02-24T16:02:16"	http://cranfield.ac.uk/cdem/festo/Sw	http://cranfield.ac.uk/cdem/festo/site_C004

Figure 5.20: State description of workpiece ID 004.

	RedSensingUnit	ActOfMeasuringRed	Workpiece	Malfunction
1	_:node26846	_:node26847	:WFMPS004	_:node26848

Figure 5.21: Malfunctioning for the Sensing Unit.

put measurement indicates 'black' instead of 'red', a discrepancy that violates 5.6, specialised for ActOfActuatingRed. This violation leads to the non-realisation of the SensingUnit's function.

Rule 5.9 will assert the SensingUnit as the malfunctioned unit as reported in Figure 5.21. The succession of the rule applied to the Red Sorting Unit component is reported in Figure 5.22. This kind of feedback from the analytics section goes back to the user and aims to support the decision-making process at this stage.

5.5 Discussion

This paper presented a five-step methodology to develop CDTs. The methodology consists of understanding the composition of the system of interest, the definition of the processes and knowledge base around the system, developing the ontology, defining the data strategy, and performing the analytics according to the purpose of the CDT. It provides practical guidance, promotes a holistic perspective, offers flexibility, and encourages interdisciplinary collaboration. Moreover, through a formal ontological approach, the successful transfer of human expertise into the CDT has been achieved. To demonstrate and validate this, this paper has used the detection of the malfunctioned component in the case of a system-level error as a validation of the CDT. Furthermore, as per their intrinsic properties, the use of ontology, particularly the TLO approach, promotes reusability, interoperability, and interactions among the next-generation CDTs.

This work demonstrates the capability of the proposed methodology to develop a CDT and transfer human expertise into it. The results presented in this paper answer the research questions posed in the introduction section.

1) Can an ontology approach be used to develop a DT? Yes, an ontology approach can be used to bring benefits to the DT. As seen in previous sections, a new paradigm has been created that brings semantic technologies into the DT, which is the CDT. Ontologies are formal representations of knowledge that define concepts and relationships. Combining these two can

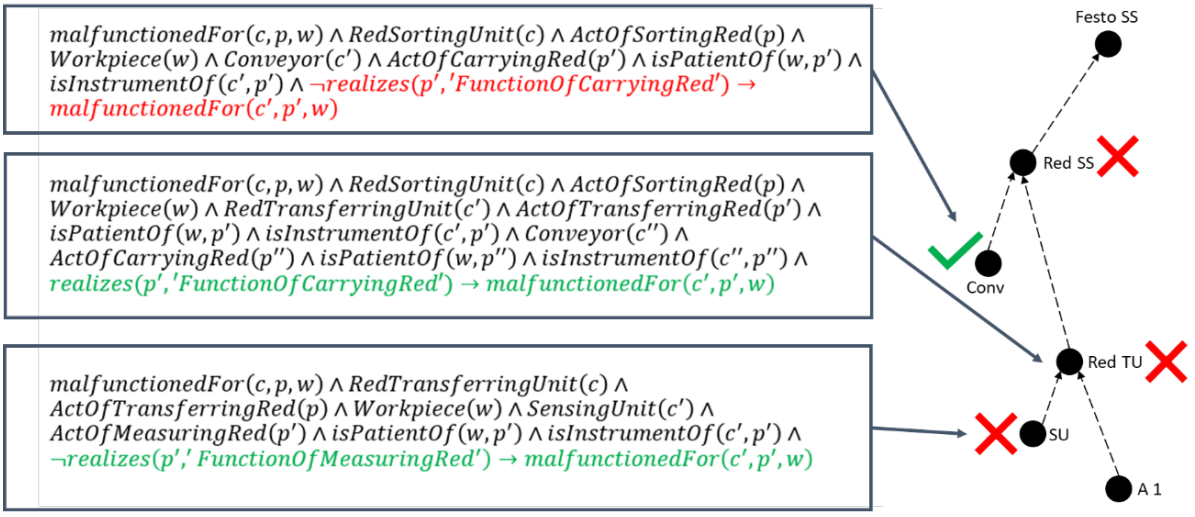


Figure 5.22: Malfunctioning for the Sensing Unit.

offer several benefits in creating and managing DTs:

- Semantic Interoperability: Ontologies provide a standardised way to define concepts, relationships, and properties within a domain. Using ontologies can ensure that different DT components (e.g., sensors, models, and simulations) share a common domain understanding, enabling semantic interoperability between different systems and tools.
- Data Integration: Digital Twins often involve data from various sources and formats. Ontologies integrate heterogeneous data by providing a common vocabulary and structure, allowing data to be transformed and interpreted consistently across the DT's life cycle.
- Knowledge Sharing and Reuse: Ontologies capture domain knowledge in a structured manner. Different stakeholders can easily share and reuse this knowledge in the development, operation, and maintenance of the DT.
- Reasoning and Inference: Ontologies enable automated reasoning and inference, which can be valuable for detecting anomalies, predicting behaviour, and optimising the performance of the DT based on the defined domain knowledge.
- Adaptability: As the physical entity or system evolves, the ontology-based DT can be updated to reflect these changes more efficiently, ensuring the virtual representation remains synchronised with the real-world counterpart.
- Life cycle Management: Ontologies can provide a foundation for managing the entire life cycle of a DT, from its initial design and development to deployment, operation, and eventual decommissioning.
- Collaboration: An ontology-based approach encourages collaboration among experts from different disciplines by offering a shared framework for understanding and discussing the DT's domain.

However, while ontologies offer numerous benefits, they also require careful design and maintenance to ensure they accurately represent the domain and remain aligned with any changes in the real-world system, see section 5.3.3. There are some drawbacks to this approach as well. Effective ontology development relies on domain experts who understand the intricacies of the specific field. The availability of such experts can be a limiting factor. In some cases, using ontologies to represent knowledge may be considered over-engineering, especially for simple systems, those with limited knowledge representation needs, or those systems that are siloed and do not require interaction with other systems.

2) Can human expertise be transferred to the CDT? Yes, it is possible to transfer human expertise into the CDT by adopting a formal ontology approach. Throughout the paper, this knowledge transfer concept, from humans to the CDT, has been validated using the Festo system as an example. The human understanding of the system has been converted from natural language into FOL rules, from (A. 7) to (A. 10) in section 0 and then converted into SPARQL queries able to run into the KG, as reported in section 5.4.5. This Festo example has been selected for its simplicity, and through this, the authors could demonstrate the procedure to convert human expertise into CDT knowledge.

In general, to import existing human expertise into the CDT, the following procedure is recommended:

1. A formal ontology approach must be selected and followed, especially with the definition of components, functions, properties, and processes.

2. The natural language knowledge should be converted into FOL axioms, following the properties defined in the previous point.

3. The FOL axioms can then be converted into SWRL and loaded into the ontology file using a tool similar to Protégé.

4. Upload the ontology file containing the SWRL rules into KG software and perform the reasoning within the KG.

In this way, human expertise is imported into the CDT, reasoning can be performed into the KG, and new knowledge can be inferred from existing knowledge. The answer outlines a structured procedure for transferring human expertise to a CDT using formal ontology approaches. However, it does not address the practical challenges and complexities associated with each process step. For example, converting natural language knowledge to FOL axioms can be challenging, especially when dealing with ambiguous human language. The answer also mentions using tools like Protégé for ontology editing and SWRL rules. While these tools are valuable, it is important to acknowledge that their effective use requires a learning curve and expertise. Additionally, tool compatibility and versioning issues may arise.

3) Can a CDT improve the process of detection of a failing component? Yes, this paper has demonstrated that the CDT is able to detect which component is malfunctioning, starting from

a general system error. Having imported human expertise translated into rules, the CDT has the capability to autonomously go through these rules and detect the malfunctioning components, like human thought, to the CDT. Without these rules, the CDT would have launched the system-level error, and the human operator should have done the whole procedure manually to understand which component was malfunctioning. This technology is valuable for industries where timely and accurate fault detection is critical for system performance and safety.

Although the answer mentions the translation of human expertise into rules, it does not address the potential complexity of creating and managing these rules, especially in large and intricate systems. Rule maintenance can be a challenging task. Moreover, the answer highlights the importance of translating human expertise into rules for the CDT to operate effectively. The accuracy and comprehensiveness of this translation process can significantly impact the CDT's detection capabilities.

Moreover, another critical limitation arises from the difficulties associated with dynamically integrating expertise knowledge into a CDT. While the dynamic adaptiveness of data is a fundamental characteristic of standard DTs and, by extension, of CDTs [D'Amico et al., 2022], this aspect of integrating expertise knowledge presents a complex challenge and necessitates further research and study [Akroyd et al., 2021, Graessler and Poehler, 2017, Mohammadi et al., 2020]. It involves the initial translation of expertise into rules and the ongoing adaptation and updating of these rules to keep pace with evolving expert knowledge and system changes.

The methodology presented in this paper differs from the other methodologies in the literature as this one does not focus on the architecture but rather on the procedure to develop CDTs. Traditional methodologies in the field often prioritise the architectural aspects of CDTs. These methodologies tend to delve deeply into the design and structure of the CDT, including its components, data flows, and system framework. In contrast, the methodology presented in the paper emphasises the procedural aspects of CDT development. It concentrates on the step-by-step processes, methodologies, and best practices for creating CDTs, considering how to effectively define, model, and operationalise them. Rather than diving immediately into technical architecture, it guides practitioners through the process of conceptualisation, knowledge integration, and operationalisation of CDTs within a specific domain. Moreover, the methodology presented is tool-agnostic and can be adapted to various applications within the industry domain. The case study used to validate the methodology serves as a proof-of-concept and does not fully cover all the aspects explained in the methodology. Therefore, additional challenges, out of the scope of this paper, need to be considered in the implementation on a complex engineering asset. The methodology has been applied to a real-world application (Section 5.4) and can be used as guidance to develop new CDTs.

The use of ontologies for developing CDTs introduces limitations such as the complexity of ontology creation and maintenance, requiring domain expertise for accurate representation;

challenges in scaling up ontologies for more extensive and evolving systems; potential gaps in formalising all forms of human expertise; the need for ongoing maintenance and alignment as the real-world system changes; potential difficulties in achieving standardised interoperability and integration across systems; technical implementation requirements and potential data integration challenges; limitations in addressing exceptionally complex scenarios; and dependence on the accuracy of human expertise translated into rules, impacting CDT performance if initial knowledge is flawed or incomplete.

Further limitations, specifically on the use case in relation to the automation and autonomy of the CDT are presented below. The focus of this project was more on the validation of the semantic capability of the DT, such as the ontology, the KG, and the capability to transfer human expertise into the CDT. The automation aspect considers the capability of the CDT to make queries systematically. It can be improved using additional sensors and scripts with control checks. The second aspect to consider is the autonomy of the CDT. It looks at the capability of autonomously running the rules, inferring relationships, and supporting decision-making. This aspect can be improved over time by working more with graph algorithms on the KG. Implementing these aspects has to be considered in the design phase of the CDT or as an iterative implementation that optimises the operation of the asset.

5.6 Conclusion

The paper outlines a methodology comprising five steps for creating CDTs leveraging ontologies to enhance their development. These ontologies offer benefits like standardised understanding, integrated data, knowledge sharing, reasoning, adaptability, life cycle management, and collaboration. The paper demonstrates the transfer of human expertise into CDTs through formal ontological approaches, exemplified by converting human understanding into rules and integrating them into ontologies for reasoning in the KG. Additionally, the paper showcases how CDTs, enriched with human expertise, can autonomously detect failing components when system errors occur, reducing the need for manual intervention and improving overall performance in complex systems. The fundamental benefit of DTs is their ability to predict and simulate physical states. Since CDTs encapsulate all the attributes of DTs, they retain this predictive and rehearsal capacity and enhance it through their cognitive dimension.

Integrating human expertise into the digital domain is set to bolster the predictive analytics capabilities of DTs significantly. Our approach augments the traditional reliance on physical models embedded within DTs. By incorporating the insights and experience of experts and operators familiar with the asset, we provide a more nuanced understanding of its behaviour. Hence, this methodology improves DTs' predictive precision and self-sufficiency, enabling them to act more effectively and autonomously in foreseeing and managing potential issues within complex systems.

Future research in this field is needed to address several critical aspects. The CDT concept requires further elucidation, necessitating the adoption of a unified definition to mitigate inconsistencies. Additional applications are essential to thoroughly validate various dimensions of this paradigm. Conducting case studies across diverse domains holds significant promise for assessing the practical implementation, advantages, and challenges of ontology-based CDTs, offering valuable insights for real-world adoption. Further exploration is warranted into mechanisms that ensure the precision and comprehensiveness of human expertise translated into ontologies, thereby fortifying the reliability of CDTs. Advancing reasoning and inference techniques to manage intricate scenarios marked by complexity and uncertainty would substantially bolster the capabilities of CDTs in handling intricate systems. Another promising avenue is delving into mechanisms to enhance the synergy between human experts and CDTs, enabling continual ontology refinement and knowledge transfer. Efforts to automate the translation of human expertise into formal axioms and rules could streamline the process and minimise dependence on manual conversion. Seamless integration of ontologies with existing systems, databases, and data formats is pivotal for the pragmatic implementation of ontology-based CDTs. Research on real-time ontology updates, aligning with changes in the physical system, holds immense potential for enhancing the adaptability and relevance of CDTs over time. Formulating standards and protocols for ontologies within the context of CDTs is vital to augment interoperability between different systems and stakeholders. The authors do not claim that BFO is the best TLO existing today. In fact, this methodology can be followed by selecting another TLO. The authors believe that a choice must be made to make some progress in this field.

Chapter 6

The CDT concept and its role in interoperability

This section discusses the evolution of CDT, its concept, and its role in manufacturing systems. It breaks down the CDT's framework into five main components, emphasising the integration between the physical and digital domains. The document highlights the importance of a TLO approach for achieving semantic interoperability across domains. It details the reason for ontology development, the benefits of a flexible interface in the CDT, the introduction of advanced semantic capabilities, and how to measure interoperability.

6.1 The CDT framework

This framework is the culmination of an extensive systematic literature review detailed in Chapter 3, specifically adapted from the insights presented in Figure 3.17, the literature reviews in Chapter 4 and Chapter 5, and the experience gained throughout the PhD. The process involved collecting and classifying existing digital twin models from the literature, serving as foundational references for developing this comprehensive CDT framework. Central to this framework are several key aspects:

- *Bidirectional Communication*: A core feature of this framework is the seamless bidirectional communication between the physical and digital worlds.

- *Interoperability Needs*: The framework addresses both internal and external interoperability. Internally, the CDT is designed to facilitate modular interaction, allowing different components within the CDT to seamlessly integrate and evolve over time, inspired by the framework discussed in [Lu et al., 2020a]. Externally, the framework supports a federated approach, enabling the CDT to interact with other CDTs and external systems, as represented by the thick grey line in the model. This dual-layer interoperability is crucial for the CDT's adaptability and scalability in complex ecosystems.

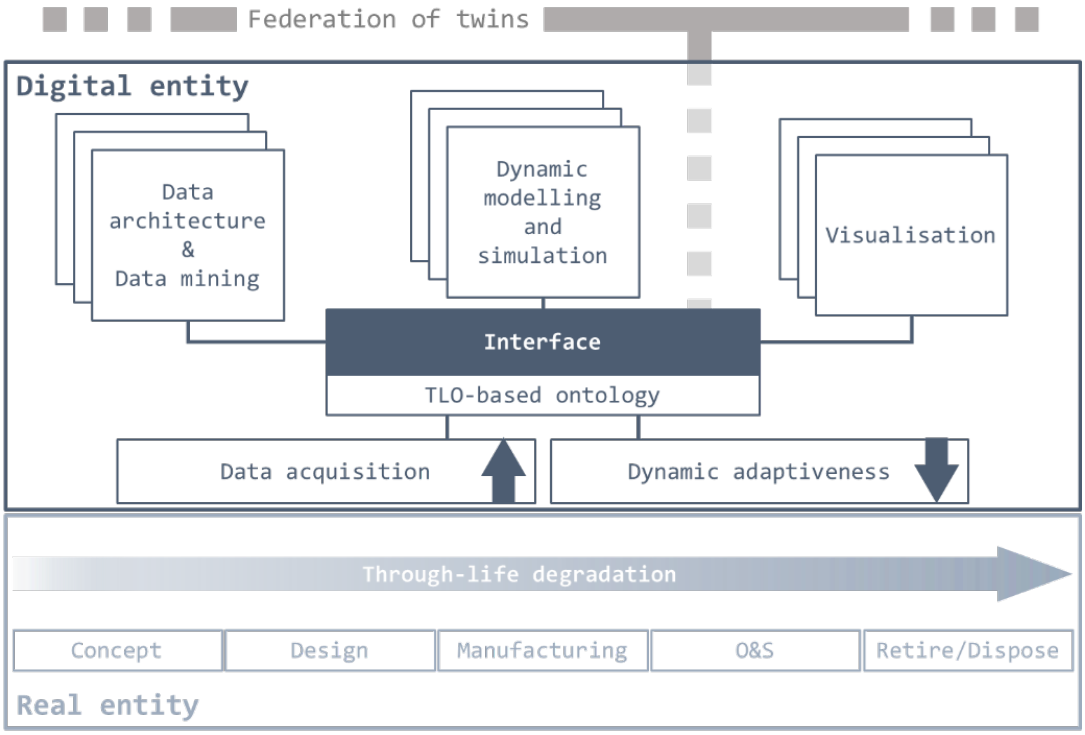


Figure 6.1: Cognitive Digital Twin framework based on a Top-Level Ontology.

- *Central Interface and Ontology*: At the heart of the framework is a central interface grounded in a robust ontology based on a universally recognised TLO. This ontology serves as the semantic backbone of the CDT, ensuring that data and interactions within the CDT are semantically consistent and interpretable both internally and in federated environments.

- *Influence of Systematic Literature Review*: The internal modules of the CDT are directly derived from the findings of the systematic literature review. This review identified prevalent approaches and features in existing DT models, which were then synthesised and integrated into the framework, ensuring that the CDT encapsulates proven methodologies and capabilities identified in academic and industry research.

- *Life Cycle Integration*: A distinctive characteristic of this CDT framework is its design to accompany the asset throughout its entire life cycle. This longitudinal approach ensures that the CDT provides continuous value from conceptualisation through decommissioning, offering insights and decision-support at every stage of the asset's life.

The framework presented Figure 6.1 is not merely a static model but represents the latest evolution of an iterative design process that spans the entire PhD journey. This iterative approach involved continuous refinement based on feedback, empirical testing, and alignment with emerging technologies and methodologies, ensuring the framework remains at the forefront of CDT development.

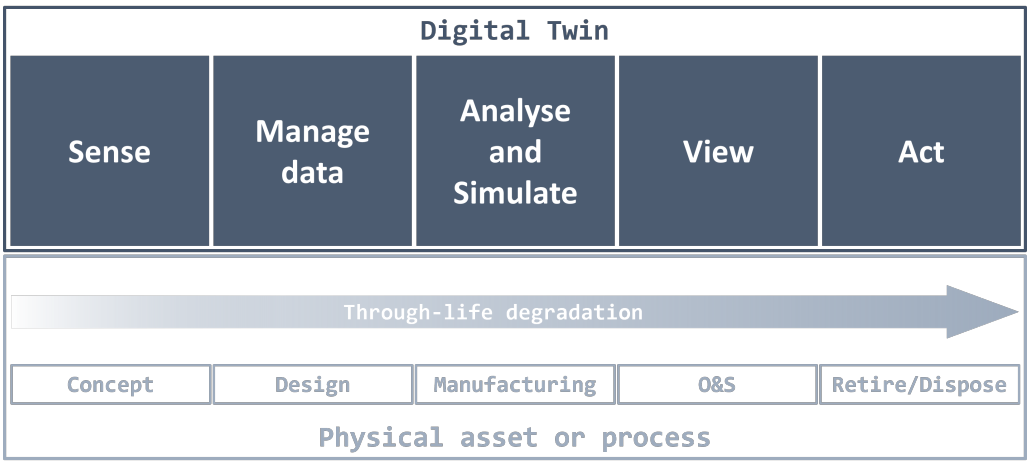


Figure 6.2: The 5D DT model.

The digital world is constituted by blocks, the two at the bottom represent the bidirectional interconnection with the real world. Not only the DT will be able to collect data from the real world, but also dynamically interact with it, changing parameters, and providing action plan scenarios to support decision-makers. The three blocks above represent the main parts of the DT: the data architecture block represents the characteristic of the twin to have a flexible data architecture as well as a reliable mining strategy, with local databases (edge computing), remote databases (cloud computing), or a combination of them. The dynamic modelling and simulation block represents the intelligence of the twin in terms of dynamic models and simulation capability. This part could contain the degradation models, for example, simulation features to investigate several real-world scenarios, and artificial intelligence algorithms to predict remaining useful life. The visualisation block represents the characteristic of the twin to provide a graphical user interface to operators containing all the helpful information to make better decisions. The visualisation can be performed through applications, web applications, and eXtended Reality (XR) technologies. In the core part, the interface, which design relies on a TLO. This flexible and reliable interface, enhanced with advanced semantic capabilities, enables interoperability within the twin. It allows the flexible integration of different agents and provides an open framework to the CDT. The framework of the CDT, reported in Figure 6.1, can be further abstracted, and the features of the CDT can be summarised in five main components: sense, manage data, analyse and simulate, view, and act. These five components represent the five functions that constitute the framework of a CDT, as reported in Figure 6.2.

- *Sense* is one of the two connections to the physical world. This part is the one-way connection from the physical to the digital world. This part of the DT takes care that sensors' data and general information from the real world are transferred to the DT.

- *Manage data* is the function of the DT that manages the data coming from the physical world. Another aspect considered in this step is the architecture, or the rules governing how data will be stored, arranged, integrated, and used in the DT.

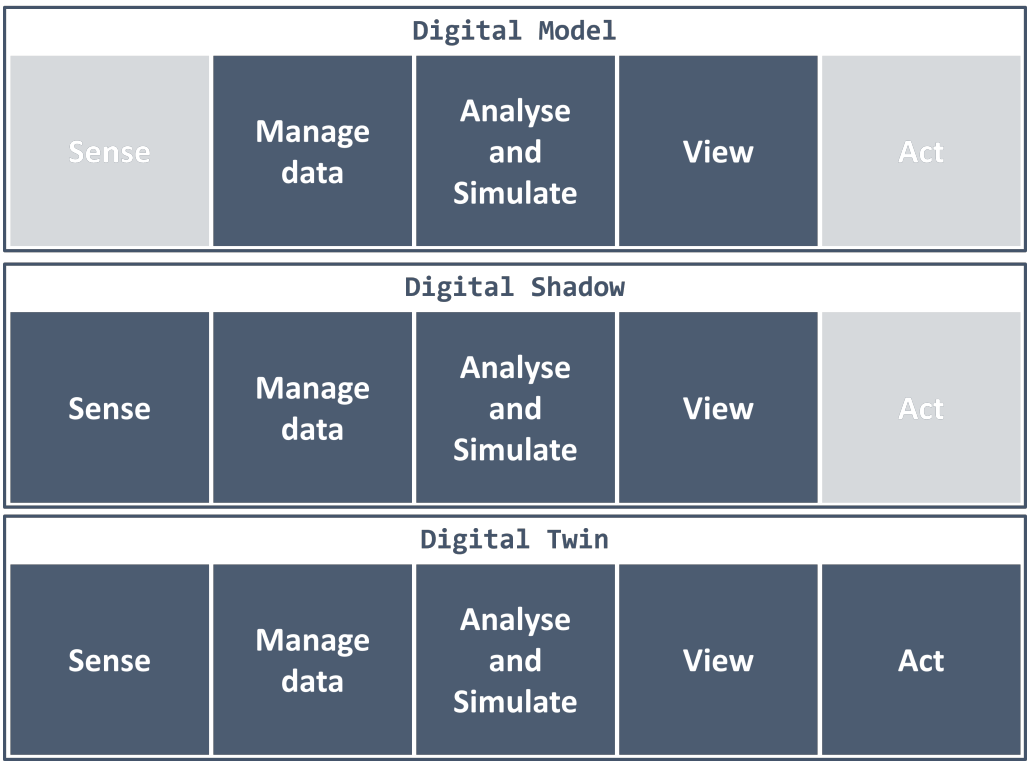


Figure 6.3: The graphical difference between model, shadow, and twin according to the basic model presented in this paper.

- *Analyse and simulate* is the function of the DT that contains models, simulations, predictive analytics algorithms, etc. This feature allows the DT to evaluate all the what-if scenarios and select the optimum operation point.

- *View* is the function of the DT that visualises the information. It can be through dashboards, AR/VR headsets, or a combination. This can be represented through the 3D model of the asset augmented with helpful information or through charts and graphs.

- *Act* is the second connection with the physical world. This part is the one-way connection from the digital to the physical world. It allows the DT to dynamically interact with the asset, optimising the operation in the real world.

The five-functional framework can also showcase the three modalities of digital representation [Kritzinger et al., 2018]: digital model, digital shadow, and DT, as reported in Figure 6.3. The digital model, which represents a category of assets, is not a one-to-one replica of a single asset; it does not gather sensors' data from one specific asset. It represents a kind of asset that can be valid for all the assets in the same model. In the middle, the digital shadow, representing one single asset, is a replica of a single asset. The digital world will be fed by sensors' data of that specific asset, and the only missing thing is the feedback from the digital world to the physical world (the dynamic adaptiveness in Figure 6.1). Ultimately, the DT represents a specific asset through life. In this case, the data is fed to the digital world and elaborated, and actions

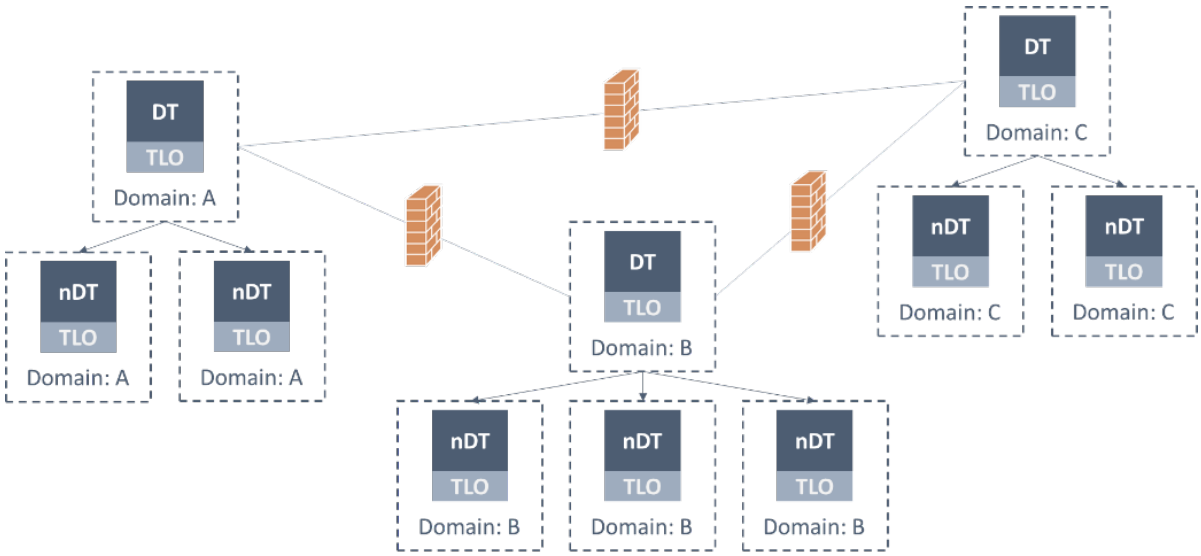


Figure 6.4: Example of representation of Connected DTs.

can be taken autonomously from the digital world to the physical world. This distinction has been adopted in manufacturing [Bergs et al., 2021] also from the 'ecosystem of shadows' point of view [Brecher et al., 2021].

6.2 The connected DTs using the CDT concept

An ontological approach ensures semantic interoperability within the CDT. A TLO approach extends this semantic interoperability across a network of CDTs, forming a Federation of Twins. The vision is to create CDTs that can be easily connected in a network and share information with similar twins within the same domain and across different domains. This interoperability is achievable if they share the same TLO, which significantly enhances the training, experience, and knowledge of the overall system of systems. Figure 6.4 provides an example representation of the Federation of Twins.

A DT can be composed of sub-DTs (nDTs). For instance, the DT of a complex engineering asset might consist of nDTs representing its subsystems, or the DT of a fleet of assets might be composed of nDTs representing each individual asset. Different domains, such as different application fields, can be connected with consideration for security measures (the walls). For example, a generator used in both a wind turbine and a nuclear power plant can have similar characteristics. The CDTs of these generators, though in different domains (wind energy and nuclear energy), can share data patterns, information, knowledge, and experiences.

In the context of a smart city, the CDTs of various components must interoperate, allowing seamless information flow across different CDTs. By adopting a shared semantics approach across all CDTs of each subsystem within the smart city, this flow of information is facilitated. Linking domains such as assets, fleets, and subsystems within these overarching frameworks

ensures that even diverse application fields can benefit from shared insights and interoperability.

6.3 The TLO approach

In the last two decades, ontologies have been used in computer science as a tool to transfer human knowledge into machine instructions. An ontology aims to enable knowledge sharing and reuse by means of a definitive classification of entities in specific domains constructed based on a controlled vocabulary with logical definitions of its terms. To achieve this goal, the ontology content should be validated by human experts in the corresponding domain [Smith et al., 2008].

Ontologies can be categorised into 3D and 4D ontologies, which derive from differing philosophical views on objects' nature and continued existence over time. The primary difference between the two is that 3D ontology considers objects as self-contained entities with all their properties at a single moment in time, whereas 4D ontology views objects as collections of temporal parts with properties distributed across time [Sider, 1997]. Although both ontologies can represent the same concept, they offer distinct perspectives and methods of comprehending it. 3D ontology emphasises the fixed properties of objects, while 4D ontology highlights the dynamic nature of objects over time. The choice of ontology depends on the research question and the objectives of the analysis.

One central goal of the annotation of data using ontologies is to enable what is called "semantic interoperability" between heterogeneous computer systems, defined as the ability of two or more systems capable of exchanging information in such a way that the meaning of the information generated by any one system can be automatically interpreted by each receiving system and accurately enough to produce results beneficial to its end users. Using a common top-level ontology for describing the data generated by information systems increases the likelihood that these conditions can be met [Arp et al., 2015]. Interoperability at the software level can be achieved by using open standard formats [Government, 2012], such as W3C standards.

An ontology will (i) map human knowledge, (ii) provide common knowledge of the system/context of interest, and (iii) establish common semantics over the system of systems environment through life. One of the elements of an ontology is the map that represents the relations between objects. This is in the form of a graph. The graph database is the perfect fit for the user to shape the data based on an ontology.

6.4 Flexibility

The methodology presented in the previous chapter (Chapter 5) places a strong emphasis on flexibility, which is achieved through the use of graph databases. Graph databases are known for their ability to easily accommodate new data and knowledge without the need for extensive redesigns or restructuring of the database. Once the initial graph is built, new elements can be

integrated without any significant overhaul of the existing architecture. This is because graph databases are designed to be highly flexible and adaptable. When new data and knowledge are added to the system, the only effort required is to find the linking point with the existing ontology graph and append the data at the dataset level. This is a significant advantage of graph databases, as it eliminates the need for complex database restructuring and redesigns, thereby saving time and effort.

Another advantage of using graph databases is their efficiency in managing unstructured data. The flexibility of graph databases makes them ideal for handling unstructured data inputs, which may not fit easily into traditional database formats. By leveraging the power of graph databases, this methodology can effectively manage and integrate unstructured data inputs into the ontology graph, resulting in a highly flexible and adaptable system.

Graph databases are particularly well-suited to handling unstructured data because of their ability to represent complex relationships between data points. Unlike traditional relational databases, which store data in tables with pre-defined columns and rows, graph databases store data in nodes (or vertices) and edges, allowing for more flexible data modelling. This structure enables the representation of intricate and dynamic relationships among data points, which is often a challenge for traditional databases.

Moreover, graph databases excel in the area of entity extraction from unstructured data or blocks of text. Entity extraction involves identifying and classifying key elements within text, such as names of people, organisations, locations, dates, and other significant terms. For example, in a block of text from a social media post, an entity extraction algorithm can identify and extract entities like the names of users, mentioned products, and relevant locations. This process transforms unstructured text into a structured format, where each entity is a node in the graph, and the relationships between entities are represented as edges.

Some ways that graph databases can better handle unstructured data are set out below:

- **Flexible data modelling:** Graph databases allow for more flexible data modelling compared to traditional databases. New types of relationships can be created and defined as they emerge, and new nodes and edges can be added as needed without disrupting the existing data structure. This flexibility is crucial for managing unstructured data, where the structure and relationships may not be known in advance.

- **Fast and efficient querying:** Graph databases excel in querying and navigating complex relationships among data points. This makes it easier to find patterns, connections, and insights in unstructured data that would be difficult to discover with other database systems. The graph structure allows for efficient traversal and querying of data, even as the dataset grows in complexity.

- **Natural language processing (NLP):** Graph databases can be used to store and process unstructured text data, such as social media posts, customer feedback, or product reviews.

By combining graph databases with NLP algorithms, it is possible to extract meaning and insights from text data. For example, entity extraction involves identifying and classifying key elements within text, such as names of people, organisations, locations, and dates. This allows for the creation of a structured representation of unstructured text, which can then be analysed for sentiment, topics, or other relevant information.

- **Machine learning (ML):** Graph databases can be integrated with ML algorithms to improve prediction accuracy and recommendation systems. By analysing the relationships among nodes, patterns and trends that would be difficult to find with other database systems can be identified. For instance, in recommendation systems, graph databases can help identify connections between users and products, improving the accuracy of recommendations based on the relational data.

6.5 Interoperability

A TLO can be used as a basis for all application ontologies to ensure interoperability in the development of ontologies. This approach enables semantic interoperability. TLO is a domain-neutral ontology that can be applied to different domains. All application ontologies can inherit the upper-level semantic properties by following the same TLO, or upper-level semantic rules (including MLOs and DLOs). This methodology ensures that all application ontologies developed within the system are consistent and follow the same rules, resulting in semantic interoperability across the system.

Selecting a suitable TLO is critical to achieving semantic interoperability. For instance, if the selected TLO has already been used in a specific domain, such as medicine, this can facilitate integration between medical application ontologies and other ontologies, such as machine or asset application ontologies. Thus, carefully selecting a TLO and transferring the upper-level semantic model into lower-level ontologies ensures semantic interoperability across the system.

By ensuring semantic interoperability across the system, we can achieve a common understanding of the data and knowledge represented by the ontology. This enables diverse systems, organisations, and individuals to collaborate and exchange information, regardless of technology. It also makes integrating new data and knowledge into the system easier without requiring extensive modifications.

Furthermore, using a TLO can also improve the consistency and accuracy of the application ontologies. By inheriting the upper-level semantic properties, we can avoid duplication of effort and ensure that the ontologies are consistent with each other. This can reduce the risk of errors and improve the quality of the data and knowledge represented by the ontology.

Overall, the use of a TLO and the adoption of semantic interoperability principles can provide a solid foundation for the development of an ontology-based system. By adopting a

reference model, the system can attain greater coherence and compatibility, enhancing its effectiveness in capturing and utilising information. Additionally, using a standardised reference model can simplify the process of exchanging information between different organisations and systems, enabling better communication and collaboration.

6.6 Reusability

The development of an ontology can be driven by several motivations [Noy and McGuinness, 2001]:

- Establishing a unified framework for understanding information structures across humans or software entities.
- Facilitating the reuse of specialised domain knowledge (such knowledge can be unstructured, textual, hidden in data, and even sometimes only in the employee/operator's experience - know-how).
- Clearly expressing and delineating domain-specific assumptions.
- Providing a means for rigorous analysis of domain knowledge.
- Extracting domain knowledge from data and information. This is crucial as much of today's data remains unused due to a lack of metadata. Ontological analysis offers a solution by providing a structured framework that defines relationships and common vocabulary within a domain, enhancing data understanding, integration, and retrieval. This approach transforms raw data into valuable insights, facilitating knowledge sharing and enabling organisations to achieve wisdom through informed decision-making.

The Web Ontology Language (OWL) is a language adopted to express the ontologies. OWL's foundation in Description Logics (DLs), a family of logics that are expressively weaker than standard FOL but have some computational advantages for ontology-based reasoning and data validation, is one of its main features. Originating in the mid-1980s within the artificial intelligence realm, DLs have formal semantics often articulated through model theory, a concept commonly associated with FOL [Arp et al., 2015]. Every ontology is a set of vocabularies (sometimes called the signature of the ontology) and a set of statements that can be used as axioms (for what you described) or for various other documentation and organisation purposes. OWL provides a formal representation of a specific domain of interest. There are three distinct syntactic categories, namely Entities, Expressions, and Axioms:

- Entities: These are the foundational elements of an ontology, representing classes, properties, and individuals, each identified by an IRI (Internationalised Resource Identifier). They act as the primitive terms of an ontology.

- Expressions: These are utilised to convey complex concepts within the defined domain. A class expression, for instance, delineates a group of individuals based on specific characteristic constraints imposed on the group members.

- Axioms: These are declarative statements postulated as truths within the domain of interest. They provide a formal mechanism to define relationships and constraints between different entities.

Any TLO provides a set of vocabularies that are domain-neutral and defined so that any domain can use them. When opting for a TLO like BFO, one inherits a structured framework comprising entities, expressions, and axioms, but the OWL version of BFO is just a part of it. This structured inheritance is pivotal because the foundation has already been set when one proceeds to design more specific ontologies, be it domain-specific (domain-level ontology) or tailored for particular applications (application-level ontology).

The classes within these specific ontologies (domain or application) can be systematically developed as subclasses of the overarching classes present in the upper-level ontology, whether it is a TLO or an MLO. This hierarchical structuring ensures a clear lineage and connection between broader concepts and their specific instances. This is key to ensuring interoperability among different domains, as the vocabularies in different domain-level ontologies are defined based on the same primitive (broader) concepts.

Furthermore, every domain and application ontologies entity is assigned a unique IRI. This ensures that each entity is distinctly identifiable over the network, eliminating ambiguity.

One of the significant advantages of this approach lies in the properties. When defining relationships or attributes within domain or application ontologies, it is possible—and often advantageous—to adopt properties directly from the TLO or MLOs. This adoption promotes reusability, which is a vital feature in ontology development. The domain or application ontology inherits a degree of compatibility by reusing established properties from upper-level ontologies. This reusability streamlines the development process and enhances the ontology's ability to work in different contexts or applications—what we call 'interoperability'.

The advantage of this interoperability is its simplification of queries. For instance, when querying using a property from an upper-level ontology, the system can fetch results from any domain or application ontology that employs that specific property. This cross-referencing capability streamlines data retrieval, making the system more efficient and cohesive.

6.7 Interoperability index

In developing an application ontology, all the classes are constructed as subclasses deriving from the upper-tier ontologies, whether they are TLO, MLO, or DLO. Consequently, each class in the application ontology is assigned a unique IRI (Internationalised Resource Identifier) dis-

Figure 6.5: Representation of ontology dependence with two levels of abstractions.

tinct from the IRIs of the classes in the higher-tier ontologies. While this unique identification ensures clarity, classes are not directly interoperable due to these different identifiers.

However, a positive aspect of this approach is in property inheritance. Even though classes in the application ontology have distinct IRIs, they still inherit the properties of the upper-level classes from which they are derived. These inherited properties are pivotal in establishing interoperability for the application ontology. For maximum interoperability, an ontologist must incorporate as many properties from the upper-tier ontologies as feasible.

When integrating a TLO, for instance, BFO, into a framework, it also inherits all the defined entities, expressions, and axioms defined at the TLO level. All properties within the TLO's documentation are systematically detailed with their respective 'domain' and 'range'. These parameters indicate the specific classes to which each property relates to. Therefore, while developing a class at the application level as a subclass of an upper-level ontology, this new class seamlessly inherits all properties associated with its parent class. This intricate design ensures a harmonious blend of distinct identification and inherited functionality.

Assuming that two application ontologies exist deriving from a TLO, as reported in Figure 6.5. The interoperability index is defined as:

$$I^2_{ALO}(TLO) = \frac{P_{TLO}(ALO)}{(P_{ALO}(ALO) + P_{TLO}(ALO)} \tag{6.1}$$

Where:

- P_{ALO} (ALO) is the total amount of properties of the application ontology (created from scratch) at the application ontology level.
- P_{TLO} (ALO) is the total amount of properties inherited from the TLO at the application ontology level.
- I^2_{ALO} (TLO) is the interoperability index of the application ontology at the TLO level.

The interoperability index defines the degree of interoperability (from 0 to 1) of an application ontology at the upper ontology level. Therefore, the application ontology is interoperable with

all the other application ontologies that inherit the same properties from the same TLO used as a reference. This procedure can be applied at any level of abstraction. It can also be used to calculate the degree of interoperability among application ontologies created from a specific domain ontology. Moreover, it can also be applied when more than two levels of abstraction are used. Using the level of the ontology, we are calculating the interoperability index as a subscript, and at which level we are considering the interoperability, inside the brackets. (e.g., having four levels of abstractions, I^2_{ALO} (DLO) represents the interoperability index of the application ontology at the domain level).

Chapter 7

Discussion

The work during this PhD aimed to develop a CDT architecture to facilitate the transferability of human expertise and support the management of complex engineering assets by transforming low-semantics data into semantically enriched data. This paradigm moves away from the current siloed approach, enhancing interoperability and allowing all current and future components of the DT to access data through life. This chapter discusses the key findings that are aligned with the research objectives developed to achieve the aim. Section 7.1 critically discusses the findings in the context of DT technologies, outlining the intricate relationship between research findings and the industry's current state. Following this, section 7.2 outlines the study's contributions to the knowledge, particularly the new insights provided by the CDT methodology. Section 7.3 acknowledges the study's limitations, offering a transparent version of the scope and potential areas for further study.

7.1 Critical discussion of the results

This study started in 2020 when the number of publications about DTs doubled in just one year. The DT is gaining huge success both in academia and industry, and the very success of this technology has led to a proliferation of standalone solutions that are not interoperable with each other [D'Amico et al., 2022]. This is one of the main reasons that triggered this research.

7.1.1 Analysis of the systematic literature review paper

The initiation of this research with a systematic literature review on DTs within the manufacturing domain sets a foundational stage. The analysis included a detailed examination of DTs' technological foundations, application domains, and operational functions, with a particular emphasis on maintenance and degradation concerns. The discovery that most current maintenance procedures alternate between preventive and reactive schedules, which frequently result in unplanned downtimes and unnecessary maintenance tasks, calls into question the effective-

ness of current tactics and the incorporation of DTs as a solution that works.

Further investigation is necessary into the claim that DTs have the potential to transform the maintenance paradigm by replacing time-based methods with condition-based ones. Although there is agreement that DTs have the potential to be sophisticated diagnostic instruments, there is disagreement over how realistically they can be integrated into larger systems and networks. Supporters of a single, unified semantic framework as the foundation for this integration—possibly achieved by adopting a TLO—provide an idealised picture that can ignore the complexity and variety in manufacturing environments.

The focus on interoperability as a crucial component, similar to the radical effects of gene ontology in the medical field, stimulates a critical discussion about the viability and scalability of this strategy in the manufacturing industry. These difficulties are made worse by recognising non-standard language in the DT domain as a significant barrier, pointing to the necessity of a more sophisticated strategy for standardisation and semantic clarity.

Drawing comparisons with the advances in the biomedical field, in particular the success credited to the ontological approach and the fundamental function of the Basic Formal Ontology (BFO) in the Open Biomedical Ontology (OBO) foundry, offers a promising model for manufacturing replication. Nonetheless, this comparison calls for a careful evaluation of the transferability of such frameworks as well as the contextual variations between the sectors. While the desire to emulate the biomedical industry's success and establish a DT environment with seamless human-machine interaction is valuable, assessing the underlying presumptions, scalability, and practical implications of this line of thinking is important. The idea of using a TLO-based DT as a benchmark for future industrial landscapes is forward-thinking, but given the heterogeneous and ever-changing nature of manufacturing environments, it necessitates careful analysis of its feasibility.

7.1.2 Analysis of the survey and interview paper

The second chapter delves into the topic of DTs in the manufacturing domain, providing industry viewpoints via a survey and interviews with experts, primarily from the UK defence industry. Although the article presents a strategy for DT integration in manufacturing, it does not cover initiatives like cloud migration and security improvements. Industry and academics must work together to advance the knowledge and implementation of DT. However, more research has to be done on the realities of these relationships, especially regarding intellectual property and commercial interests.

The paper introduced the concept of the Federation of Twins, proposing a more interconnected digital ecosystem. However, the scope of the study underpinning this concept, predominantly rooted in the defence sector, narrows its broader applicability across diverse industrial landscapes. This limitation underscores the imperative for a more comprehensive empirical investigation that spans various business sectors. Such investigation might assess the viability

and adaptability of the proposed Federation of Twins framework across different domains, ensuring its relevance and effectiveness in addressing the challenges and requirements of various industries.

Upon critical evaluation of this article's results, it is clear that although the research contributes to the definition of the CDT framework and highlights the limitations of the traditional DT framework, it also creates opportunities for discussion regarding the potential, scalability, and economic feasibility of advanced DT implementations in the manufacturing industry.

7.1.3 Analysis of the methodology paper

CDTs are an advancement in DTs that includes a cognitive/semantic layer meant to improve DT functionality. This conceptual shift toward a "federation of twins" indicates an encouraging future for DT applications. However, there are significant concerns over scalability, integration complexity, and the suitability of existing technological infrastructures to enable this advanced level of interoperability regarding the feasibility of creating such a cognitive layer across various DT systems.

The five-step methodology for developing a CDT highlights ontologies' crucial role in achieving semantic interoperability into the DT ecosystem. Although ontologies are a formal way to standardise concepts and relationships, there are practical obstacles in creating ontologies, such as the necessity for domain-specific expertise and the potential of overcomplicating simpler systems, which should be considered. Additionally, the methodology, which focuses on converting human expertise into formal logic rules for integration into CDTs, has substantial language and technical challenges that may compromise the accuracy of knowledge transfer. Especially appealing is the possibility of CDTs recognising and reacting to system faults on their own, reducing the need for human intervention. However, the realities of putting such sophisticated capabilities into practice, such as the difficulties of managing ontologies and the challenging learning curves associated with essential tools, limit this potential.

Looking ahead, the need for additional study in areas like cross-domain applicability, real-time ontology updates, unified CDT definition, and improved reasoning methodologies illustrates the significant obstacles as well as the promise that CDT advancement will bring. These upcoming research areas highlight the necessity of a cooperative strategy that closes the knowledge gap between theoretical developments and real-world applications, guaranteeing that CDTs may be successfully applied in various industries to improve the management of complex engineering assets.

7.1.4 Analysis of the CDT framework contribution

A complete approach to digital twinning is presented by breaking down the CDT architecture into five main components: Sense, Manage Data, Analyse & Simulate, View, and Act. Notably,

TLOs are recognised as playing a crucial role in ensuring semantic interoperability; nonetheless, this raises concerns regarding the practical difficulties in attaining this degree of semantic coherence across disparate systems in the manufacturing industry. The focus on a TLO as the foundation of seamless integration and communication presents an idealised view that does not adequately consider the complexity and heterogeneity in actual manufacturing ecosystems. Although fundamentally valid, bidirectional communication between physical and digital entities requires a careful analysis of how resilient and adaptable the dynamic relationship is to changing operational settings.

The idea of a ”federation of twins” offers an opportunity for DT applications, but the feasibility of creating a networked ecosystem with individual twins interacting and working together in synergistic networks requires further investigation of the underlying organisational, technological, and cultural issues.

Ontologies play a triple role in mapping human knowledge, establishing common semantics, and providing systematic overviews. However, acquiring the tacit knowledge and complicated decision-making processes of human specialists is challenging when attempting to translate the complexities of human cognition and understanding into machine-readable formats.

The literature surrounding graph databases and their structural divergence from conventional relational databases to depict intricate data networks underscores their capacity to support the CDT architecture. However, important factors to consider while switching to graph databases include data transfer, integrating old systems, and upskilling employees to utilise these new technologies properly.

A significant development in knowledge representation and sharing is highlighted by investigating the paradigm change that ontologies, specifically the TLO approach, represent nowadays in computer science. The adoption of such ontologies, however, adds a layer of complexity to the creation, maintenance, and application of ontologies, especially when it comes to OWL and DLs. This requires critically evaluating the trade-off between expressiveness, computational efficiency, and usefulness.

7.2 Contribution to the knowledge

One of the main contributions of this research is the five-step methodology for the development of CDTs. This methodology emphasises the procedural aspects of CDT development, offering a comprehensive approach that spans from understanding the system's composition to performing analytics tailored to the CDT's purpose. Another significant contribution is the emphasis on the role of ontologies in enhancing the development and functionality of CDTs. By highlighting benefits such as semantic interoperability, data integration, and knowledge sharing, the research deepens the understanding of how ontologies can be effectively integrated into the

realm of DTs.

An advancement presented in the research is the approach to transferring human expertise into the CDT. By converting human understanding from natural language into formal logic rules and subsequently integrating them into a knowledge graph, the research bridges a gap between human experience and automated systems. While highlighting the benefits of this approach, the research also offers a balanced perspective by shedding light on the inherent challenges associated with ontologies. This acknowledgement of potential limitations, intricacies, and the essential need for domain expertise adds depth to the study, providing a holistic understanding of the subject.

Furthermore, the research delves into the practical implications of these insights. Demonstrating how CDTs, enriched with human expertise, can autonomously detect malfunctioning components, offers a glimpse into the real-world benefits and applications of CDTs. Additionally, the paper sets the stage for future explorations in the domain by suggesting potential avenues for research, such as the need for a unified definition of CDT, the importance of further case studies, and the exploration of real-time ontology updates. While presenting a novel methodology to develop CDTs, this research points out three important aspects that further enhance the value of the proposed framework: flexibility, reusability, and interoperability.

Flexibility. The first distinguishing feature of the methodology is its inherent flexibility, which is essential for modifying CDTs to fit the dynamic and changing environments of diverse industries. The procedural approach to CDT development embodies this flexibility by facilitating the integration of diverse data streams and complex systems without adding to their complexity. Through the provision of a framework that is both modular and flexible, the methodology allows CDTs to progress in line with both operational demands and technological breakthroughs. This flexibility ensures CDTs stay applicable and valuable in rapidly evolving industrial settings, increasing their usefulness and longevity.

Reusability. In this work, ontologies are essential, especially for improving the reusability of CDT components. A blueprint that goes beyond particular use cases is created by developing ontologies based on standardised definitions and relationships, enabling their application across various DTs and sectors. This reusability, which significantly reduces redundancy and promotes a common understanding across various applications, is evidence of the methodology's efficacy. The research enhances the overall efficiency and coherence of DT systems by promoting ontology reusability, which in turn supports a sustainable knowledge management ecosystem where insights and data structures can be leveraged multiple times.

Interoperability. This research's emphasis on interoperability—highlighted by adding a cognitive/semantic layer to the CDT framework—may be its most important contribution. By allowing DTs to share a semantic understanding, this layer essentially dismantles the silos that have historically impeded system integration. The research offers protocols that enable smooth data exchange and communication between different DT systems, presenting a progressive

approach to interoperability. In order to build a cohesive digital ecosystem where DTs can cooperate, exchange insights, and contribute to a comprehensive understanding of operational procedures and asset management, interoperability is essential.

7.3 Limitations

This study has highlighted certain limitations. The primary constraint was the exclusive focus on manufacturing, which might limit the broader applicability of the findings to other industries where DTs are being utilised, such as healthcare, oil and gas, finance, or agriculture. Additionally, the methodologies used for data collection brought their challenges. While surveys and interviews provided invaluable insights, they also introduced potential biases from the questions' framing or the participants' personal biases. Furthermore, the rapid pace of technological change means that the landscape of DTs and the technologies that support them is ever-evolving.

Consequently, the insights derived from this research may be particularly tied to the study's period. Another dimension to consider is the inevitable tension between depth and breadth in research. While the study aims for a comprehensive view of CDTs in manufacturing, particular challenges within specific manufacturing sub-domains might not have been given attention. The transfer of human expertise into a digital realm, at the heart of this study, is full of complexity. Some levels of human experience and intuitive understanding could prove elusive when attempting to represent them digitally. The study's foundation on existing technological solutions also poses potential limitations, as not all conceivable advanced features or possibilities for DTs might have been explored, given the technological landscape at the time. Lastly, one must consider the external, dynamic factors such as economic shifts or political changes that can significantly impact the manufacturing sector and, consequently, the application of DTs, even though they were beyond the research's scope.

While pioneering its approach to developing CDTs, the research acknowledges several criticisms and limitations. One of the primary challenges highlighted is the inherent complexity associated with creating and maintaining ontologies. Although ontologies offer a structured way to represent knowledge, their development requires a deep understanding of the domain, making the process complicated and demanding. This complexity is further compounded by the need for domain experts who understand the specific field. The availability, or lack thereof, of these experts can be a limiting factor, potentially hindering the widespread adoption of the methodology. In certain scenarios, particularly with simpler systems or those with limited knowledge representation needs, the research concedes that using ontologies might introduce unnecessary complexity, leading to over-engineering.

Another critical challenge is the conversion of human expertise into the CDT. While the paper presents a structured procedure for this transfer, it does not shy away from highlighting

the practical challenges at each step. For instance, converting natural language knowledge into formal logic axioms can be a complex effort, especially when dealing with the ambiguities of human language.

To address these challenges, the research suggests exploring the use of large language models (LLMs) for ontology generation or population. LLMs, with their advanced natural language processing capabilities, can assist in automating the creation and maintenance of ontologies, reducing the reliance on domain experts. They can help in translating natural language descriptions into structured ontological representations, thereby mitigating some of the complexities involved in the process. This integration of LLMs could streamline the ontology development process, making it more accessible and efficient, and potentially enhancing the adoption of CDT methodologies across various domains.

This research also discusses the limitations of tools like Protégé for ontology editing and SWRL rules. While these tools are undeniably valuable, they come with their own set of challenges, from compatibility and versioning issues to the steep learning curves associated with their effective use. The study also delves into the intricacies of rule maintenance. As it showcases the translation of human expertise into rules for the CDT's operation, it simultaneously underscores the potential challenges of creating, managing, and maintaining these rules, especially in expansive and intricate systems. The accuracy of translating human expertise into rules is another area of concern. This research emphasises that the efficacy of this translation can significantly influence the CDT's detection capabilities. If the foundational knowledge is flawed or incomplete, it could detrimentally impact the performance of the CDT.

Furthermore, while the methodology's validation is anchored in a case study, this research acknowledges that this case study serves more as a proof of concept. It does not encompass all aspects elucidated in the methodology, suggesting potential challenges when applying it to more complex engineering assets. Lastly, this research's focus predominantly lies in the validation of the semantic capability of the DT. However, it recognises further limitations related to the automation and autonomy facets of the CDT. Enhancements in these domains, such as more systematic queries and rule inferences, would necessitate additional considerations and efforts.

One criticism concerns the scalability of the approach. Although this research and some of the scientific literature have laid a convincing foundation on how CDTs can evolve and integrate with complex systems, other parts of the scientific literature argue that real-world applications may introduce unforeseen challenges that could test the limits of scalability. These challenges may arise from large amounts of real-time data, unexpected system behaviour or intricate interactions that laboratory setups have not considered.

Another point of discussion is the reusability of the ontology. Although the ambition to create standardised ontologies that can be seamlessly reused in different DTs is laudable, it may encounter practical obstacles. Given the diversity of domain-specific details and requirements,

part of the scientific community questions whether a truly universal ontology can exist without compromising the nuanced needs of individual domains. Though ideal, some suggest that this aspiration may oversimplify the intricate complexities inherent in specialised domains.

While interoperability is a cornerstone of the proposed framework, it also brings numerous challenges. The integration of different systems, each with its own protocols, data structures and operational nuances, may require more work than expected. Part of the scientific community often worries about the tangible costs, both in terms of time and resources, associated with achieving true interoperability. Furthermore, one wonders how well these systems can communicate without errors or misunderstandings, especially in high-risk scenarios where accuracy is non-negotiable.

Moreover, while promising, transferring human expertise within the framework of CDT is not without its pitfalls. The process of converting human understanding from natural language to formal logical rules has plenty of challenges. Human experience and cognition are deeply nuanced and context-sensitive, and reducing such experience into structured, machine-understandable logic may risk omitting subtleties.

Finally, although tools such as Protégé have been fundamental to this research, they have learning curves and limitations. Reliance on third-party tools always introduces an element of dependency and potential vulnerabilities that could affect the robustness and efficiency of the CDT framework.

7.4 The research question

The two main contributions of this thesis are (i) the application of the proposed methodology in the development of CDTs and (ii) the scalability, ontology reusability and interoperability of the system. They represent significant steps towards realising a framework in which the transfer of human expertise and the management of complex engineering assets are not just aspirational goals, but practical outcomes. These contributions have been shown to enhance the capability of the DTs, enabling the scalability of the system and ensuring that the ontology developed can be reused and applied in various domains, thus responding positively to the core research question.

Can a CDT approach facilitate the transferability of human expertise and the associated semantic interoperability between digital twins, improving the management of complex engineering assets?

Yes, a CDT approach can facilitate the transferability of human expertise and ensure semantic interoperability between DTs, which has been shown to improve the management of complex engineering assets. The application of the CDT approach encapsulates and translates human skills into a format that is not only accessible but also operational within the DT environment. The interoperability promoted by this framework ensures that this expertise is not

isolated but is a shared asset that enhances the collective knowledge of the system.

The explorations of this thesis aligned consistently with established theories, thus underlining the robustness of the insights gained and affirming the significant contribution of research to the field of digital systems. This alignment reinforces the credibility of the thesis and accentuates its role in shaping the future of DTs.

In conclusion, the perspectives illustrated by this research outline a trajectory for the evolution of DTs, emphasising the potential inherent in the strategic application of ontologies. However, a critical perspective is indispensable in moving towards this future, which conscientiously addresses the criticisms and limitations identified here. Only through critical evaluation and deliberate action can we ensure that the innovative directions proposed by this research produce robust, practical and sustainable solutions. The future course, while full of promise, requires thoughtful navigation to ensure that the DTs of tomorrow not only reflect our current ambitions, but are also resilient and adaptable to future needs.

Chapter 8

Thesis conclusions and future work

This section contains the conclusive insights and future research directions deriving from the extensive study conducted in this PhD project. It describes the novel applications of CDTs in asset management and shows how top-level ontology-based methods can facilitate the transfer of human knowledge while also enhancing flexibility, semantic interoperability, and reusability. The main conclusions and contributions of the thesis are summarised below. This section also outlines a roadmap for future research, outlining possible directions like expanding into new industries, using CDTs in different domains, and investigating cutting-edge methods for integrating human expertise into CDTs.

8.1 Conclusions

This PhD thesis has addressed the research question concerning whether a CDT approach can enhance the transferability of human expertise and semantic interoperability among DTs, ultimately leading to improved management of complex engineering assets. This thesis addresses the posed question by making the following contributions:

- A systematic literature review delineates the state-of-the-art in this field, effectively identifying the research gaps in the existing academic literature (Chapter 3). Additionally, an industrial review assesses current industry practices and highlights the relevance and necessity of this PhD work from an industrial standpoint (Chapter 4).

- A novel methodology for the development of CDTs grounded in a top-level ontology approach. This methodology not only facilitates the integration of human expertise into CDTs but also ensures semantic consistency throughout. Moreover, the methodology has been validated through a laboratory investigation (Chapter 5).

- A novel CDT framework that promotes flexibility, interoperability, and reusability of CDTs was developed via the proposed methodology (Chapter 6).

The contributions detailed in this thesis comprehensively fulfil all the objectives outlined in Chapter 2. They methodically delineate the steps this research has undertaken to achieve the overarching research goal. Central to this achievement is the methodology presented in this research, which articulates a systematic procedure for formalising human expertise using formal ontologies and formal logic and subsequently integrating this expertise into the CDT framework. The research objectives for enhancing DTs have been addressed in this PhD thesis, focusing on facilitating semantic interoperability across DTs and integrating human experience to enhance the management of complex engineering assets.

Objective 1: Define the context and elements of the CDT framework. It is understanding the current limitations and challenges of the traditional DT architecture. Conceptualise the fundamental elements of a CDT that can effectively integrate human expertise.

The thesis commenced with a systematic literature review and an industrial review, incorporating surveys and interviews, to define the context and fundamental elements of the CDT framework. This initial phase identified the limitations of traditional DT architectures and conceptualised the essential components required for a CDT to effectively integrate human expertise, addressing the first research objective comprehensively.

Objective 2: Develop methods for converting low-semantics data into semantically enriched information, ensuring the CDT can effectively understand and process the data it receives.

A novel 5-step methodology for CDT development was introduced, focusing on converting low-semantics data into semantically enriched information. This methodology, grounded in a top-level ontology approach, ensures that the CDT can effectively comprehend and process data, thus fulfilling the second research objective.

Objective 3: Design interoperability protocols. Break down the siloed approach by developing standardised methodologies that facilitate seamless data exchange and communication between different systems. Investigate how those standardised methodologies can be incorporated into the CDT framework.

The research delineated the design of interoperability protocols within the proposed CDT framework to break down the siloed approach prevalent in traditional DT systems. The thesis addresses the third research objective by developing a formal methodology that facilitates semantic interoperability, enhancing data exchange and communication, and ensuring that these methodologies can be incorporated effectively into the CDT framework. 8

Objective 4: Outline expertise transfer mechanism. Understand the different modalities of human expertise in complex engineering assets management. Incorporate that expertise within the CDT framework.

The thesis elaborates on mechanisms for transferring human expertise into the CDT framework, addressing the fourth research objective. This was achieved by developing a methodology that utilises formal ontologies and formal logic to formalise human expertise and integrate

it into the CDT, ensuring a comprehensive understanding of different modalities of expertise in complex engineering asset management.

Objective 5: Verify and validate the CDT framework. Implement a prototype based on the designed CDT framework to verify and validate it in a laboratory setup.

Finally, the thesis presents the verification and validation of the CDT framework by implementing a prototype in a university laboratory setup. This practical application of the CDT framework demonstrates its robustness and applicability, effectively meeting the fifth research objective.

The research acknowledges the limitations and challenges associated with developing and applying ontologies in the context of complex engineering assets. It highlights the complexity of creating and maintaining ontologies, the difficulties in converting human expertise into formal logic, and the limitations of tools like Protégé. Despite these challenges, the research concludes that a CDT approach can facilitate the transferability of human expertise and improve the management of complex engineering assets. However, it emphasises the need for critical evaluation and deliberate action to ensure the robustness and sustainability of digital systems in the future.

These contributions represent a leap forward in the realm of CDT, offering both a theoretical framework and practical reference that have the potential to change the management and operation of complex engineering assets radically. Challenging the dominance of relational databases, the emergence and potential supremacy of graph databases represent a notable turn. Their capability to handle intricate and interlinked data places them at the forefront of a potential database revolution. In wrapping up this exploration into ontologies, it is pivotal to realise that this is not a mere academic pursuit. Ontologies will play an increasingly fundamental and revolutionary role as the impact of AI, IoT, and Big Data analytics grows.

8.2 Future works

The proposed framework and thesis contributions have demonstrated the potential to improve the efficiency and effectiveness of the maintenance of complex engineering assets by integrating human expertise in digital systems. Moreover, such a top-level ontological approach ensures semantic consistency and enables interoperability among DTs. Human expertise can now be formalised and transferred into CDTs. Future works can be summarised as follows:

Cross-Industry and International Research Expansion: Acknowledging the anticipated increase in the adoption of DT technologies across various sectors, future research aims to extend its scope to encompass a broad array of industries and international settings. Such an expansive approach is expected to yield a richer understanding of how different industries adopt, adapt, and innovate with DT technologies, subsequently influencing the broader application and development of ontologies within these systems.

Implementation Assessment in Various Domains: Through surveys and interviews across a spectrum of industries, future research will aim to assess how companies are implementing DTs and ontologies, what challenges they face, and how they measure success. This could include studies on the organisational impact of DTs, their ROI, and how they are integrated with existing workflows and systems. Cross-sectoral comparisons will be particularly valuable in identifying universal versus industry-specific challenges and opportunities.

CDT Development Using BFO: Research in this area will delve into the practical aspects of developing a CDT using BFO as the foundational ontology. This includes exploring how such twins can be applied in various scenarios, particularly in environments where human-machine interactions are crucial. The effectiveness of BFO in translating and representing human cognitive processes within the DT framework will be a key focus. This will also involve evaluating how these CDTs can enhance decision-making, learning, and adaptability in complex systems.

Enhancement of Ontology-Based CDTs: Research will improve the process of embedding human expertise into ontologies. This includes developing more advanced reasoning and inference techniques to handle complex, nuanced scenarios, and automating the translation of human expertise to reduce manual effort. Ensuring these ontologies can be seamlessly integrated with various systems and updated in real-time to reflect changes in their physical counterparts will also be crucial.

Standards and Protocols for Ontologies in CDTs: Developing and proposing standards and protocols specific to ontologies in CDTs will be critical for ensuring interoperability between different DT systems. This research will aim to create a common framework that can be adopted across industries, facilitating easier data exchange and integration of DTs into diverse technological ecosystems.

Federation of Twins and Network Creation: This line of research will explore the creation of interconnected networks of CDTs, where each twin can share data and learn from the others in a federated system. Such a network underpinned by standardised TLOs could amplify AI training datasets, leading to more sophisticated and accurate predictive models within CDTs. The research would examine the technical and logistical challenges in establishing and maintaining these networks, such as data privacy, synchronisation of updates across the network, and ensuring efficient communication protocols among different CDTs.

Inferencing New Knowledge from Federation of CDTs: Another promising research area is the exploration of knowledge inference within a network or federation of CDTs. This research would investigate how interconnected CDTs can collectively analyse data, learn from each other, and generate new insights that might not be possible from a single CDT. By leveraging the power of a federated network, CDTs could potentially identify patterns and correlations across different datasets, leading to breakthroughs in predictive analytics, anomaly detection, and system optimisation. This approach would also examine the challenges and opportunities

in data sharing and privacy within a federation of twins, ensuring that collaborative learning respects data security and confidentiality norms. Additionally, research could explore the use of distributed ledger technologies like blockchain to securely and transparently manage data exchanges and inference processes within the federation.

Exploration of Alternative TLOs: While BFO offers a solid starting point, future research will also consider other TLOs, examining their potential advantages and suitability for various DT applications. This will involve a comparative analysis of different TLOs, evaluating factors like ease of integration, adaptability to specific industry needs, and their ability to represent complex data and processes. The goal will be to provide a comprehensive guide on selecting appropriate TLOs for developing DTs.

Advanced Techniques for Transferring Human Expertise into CDTs: Future research should focus on developing more sophisticated methods for capturing and encoding human expertise into CDTs. This could involve exploring advanced AI techniques such as machine learning algorithms that learn from human input, natural language processing for understanding and translating expert knowledge into machine-readable formats, and neural networks that mimic human thought processes. The goal would be to create a more seamless and accurate transfer of expertise, ensuring that CDTs can replicate human decision-making processes more effectively. Research could also delve into developing interactive platforms where experts can continuously update and refine the knowledge base of CDTs, ensuring that the DTs evolve and stay relevant over time.

This PhD thesis has investigated the impact of employing top-level ontology in DTs, particularly CDTs, for transferring human expertise into the digital world and enhancing flexibility, semantic interoperability, and reusability in managing complex engineering assets. The research highlights the potential to bridge the gap between the real and digital worlds, significantly improving the operation, availability, and management of these assets. Moreover, it sets the stage for a comprehensive framework that integrates human knowledge into digital information systems.

Appendix A

Appendix A Ethical approval letter

16 December 2020

Dear Mr D'Amico ,

Reference: CURES/12518/2020
Title: Developing digital twins of through-life degradation for high-value systems

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/12518/2020 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

References

[Abburu et al., 2020] Abburu, S., Berre, A. J., Jacoby, M., Roman, D., Stojanovic, L., and Stojanovic, N. (2020). Cognitwin - hybrid and cognitive digital twins for the process industry.

[Adu-Kankam and Camarinha-Matos, 2022] Adu-Kankam, K. O. and Camarinha-Matos, L. M. (2022). A framework for the integration of iot components into the household digital twins for energy communities. volume 665 IFIP, pages 197–216. Springer Science and Business Media Deutschland GmbH.

[Agostinelli, 2021] Agostinelli, S. (2021). Actionable framework for city digital twin-enabled predictive maintenance and security management systems. volume 205, pages 223–233. WITPress.

[Aivaliotis et al., 2019] Aivaliotis, P., Georgoulias, K., Arkouli, Z., and Makris, S. (2019). Methodology for enabling digital twin using advanced physics-based modelling in predictive maintenance. *Procedia CIRP*, 81:417–422.

[Akroyd et al., 2021] Akroyd, J., Mosbach, S., Bhave, A., and Kraft, M. (2021). Universal digital twin-a dynamic knowledge graph. *Data-Centric Engineering*, 2:14.

[Al Faruque et al., 2021] Al Faruque, M. A., Muthirayan, D., Yu, S. Y., and Khargonekar, P. P. (2021). Cognitive digital twin for manufacturing systems. *Proceedings -Design, Automation and Test in Europe, DATE*, 2021-February:440–445.

[Ali et al., 2021] Ali, M. I., Patel, P., Breslin, J. G., Harik, R., and Sheth, A. (2021). Cognitive digital twins for smart manufacturing. *IEEE Intelligent Systems*, 36:96–100.

[Almeida and Barbosa, 2009] Almeida, M. B. and Barbosa, R. R. (2009). Ontologies in knowledge management support: A case study. *Journal of the American Society for Information Science and Technology*, 60:2032–2047.

[Ansari, 2019] Ansari, F. (2019). Knowledge management 4.0: Theoretical and practical considerations in cyber physical production systems. volume 52, pages 1597–1602. Elsevier B.V.

[Ansari et al., 2020] Ansari, F., Hold, P., and Khobreh, M. (2020). A knowledge-based approach for representing jobholder profile toward optimal human–machine collaboration in cyber physical production systems. *CIRP Journal of Manufacturing Science and Technology*, 28:87–106.

[Ansari et al., 2018] Ansari, F., Khobreh, M., Seidenberg, U., and Sihn, W. (2018). A problem-solving ontology for human-centered cyber physical production systems. *CIRP Journal of Manufacturing Science and Technology*, 22:91–106.

[Arista et al., 2023] Arista, R., Zheng, X., Lu, J., and Mas, F. (2023). An ontology-based engineering system to support aircraft manufacturing system design. *Journal of Manufacturing Systems*, 68:270–288.

[Aroyo et al., 2006] Aroyo, L., Denaux, R., Dimitrova, V., and Pye, M. (2006). Interactive ontology-based user knowledge acquisition: A case study. volume 4011 LNCS, pages 560–574. Springer Verlag.

[Arp et al., 2015] Arp, R., Smith, B., and Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. The MIT Press.

[Ashraf et al., 2022] Ashraf, M., Eltawil, A., and Ali, I. (2022). Time-to-recovery prediction in a disrupted three-echelon supply chain using lstm. volume 55, pages 1319–1324. Elsevier B.V.

[Badenko et al., 2021] Badenko, V. L., Bolshakov, N. S., Tishchenko, E. B., Fedotov, A. A., Celani, A. C., and Yadykin, V. K. (2021). Integration of digital twin and bim technologies within factories of the future. *Magazine of Civil Engineering*, 101.

[Baek, 2022] Baek, M. S. (2022). Digital twin federation and data validation method. *APCC 2022 - 27th Asia-Pacific Conference on Communications: Creating Innovative Communication Technologies for Post-Pandemic Era*, pages 445–446.

[Bamunuarachchi et al., 2021] Bamunuarachchi, D., Georgakopoulos, D., Banerjee, A., and Jayaraman, P. P. (2021). Digital twins supporting efficient digital industrial transformation. *Sensors*, 21.

[Banerjee and Mittal, 2017] Banerjee, A. and Mittal, S. (2017). Generating digital twin models using knowledge graphs for industrial production lines.

[Barricelli et al., 2019] Barricelli, B. R., Casiraghi, E., and Fogli, D. (2019). A survey on digital twin: Definitions, characteristics, applications, and design implications.

[Bechberger and Perryman, 2020] Bechberger, D. and Perryman, J. (2020). *Graph databases in action*. Manning Publications.

[Beetz, 2021] Beetz, J. (2021). Semantic digital twins for the built environment-a key facilitator for the european green deal? (keynote).

[Bergs et al., 2021] Bergs, T., Gierlings, S., Auerbach, T., Klink, A., Schraknepper, D., and Augspurger, T. (2021). The concept of digital twin and digital shadow in manufacturing. *Procedia CIRP*, 101:81–84.

[Bevilacqua et al., 2020] Bevilacqua, M., Bottani, E., Ciarapica, F. E., Costantino, F., Donato, L. D., Ferraro, A., Mazzuto, G., Monteriù, A., Nardini, G., Ortenzi, M., Paroncini, M., Pirozzi, M., Prist, M., Quatrini, E., Tronci, M., and Vignali, G. (2020). Digital twin reference model development to prevent operators' risk in process plants. *Sustainability (Switzerland)*, 12.

[Bickford et al., 2020] Bickford, J., Bossuyt, D. L. V., Beery, P., and Pollman, A. (2020). Operationalizing digital twins through model-based systems engineering methods. *Systems Engineering*, 23:724–750.

[Biesinger et al., 2019] Biesinger, F., Kraß, B., and Weyrich, M. (2019). A survey on the necessity for a digital twin of production in the automotive industry.

[Bojārs et al., 2008] Bojārs, U., Breslin, J. G., Peristeras, V., Tummarello, G., and Decker, S. (2008). Interlinking the social web with semantics. *IEEE Intelligent Systems*, 23(3):29–40.

[Bonnaud et al., 2019] Bonnaud, S., Didier, C., and Kohler, A. (2019). Industry 4.0 and cognitive manufacturing architecture patterns, use cases and ibm solutions.

[Booth et al., 2012] Booth, A., Papaioannou, D., and Sutton, A. (2012). *Systematic approaches to a successful literature review*. SAGE Publications Ltd.

[Borgo et al., 2021] Borgo, S., Sanfilippo, E. M., and Terkaj, W. (2021). Capabilities, capacities, and functionalities of resources in industrial engineering.

[Boschert et al., 2018] Boschert, S., Heinrich, C., and Rosen, R. (2018). Next generation digital twin. University of Technology.

[Boyes and Watson, 2022] Boyes, H. and Watson, T. (2022). Digital twins: An analysis framework and open issues. *Computers in Industry*, 143.

[Braun and Clarke, 2006] Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3:77–101.

[Braun and Clarke, 2021] Braun, V. and Clarke, V. (2021). One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18:328–352.

[Brecher et al., 2021] Brecher, C., Dalibor, M., Rumpe, B., Schilling, K., and Wortmann, A. (2021). An ecosystem for digital shadows in manufacturing.

[Broo et al., 2021] Broo, D. G., Schooling, J., and ur Broo (2021). Digital twins in infrastructure: definitions, current practices, challenges and strategies. *International Journal of Construction Management*.

[Brosset et al., 2022] Brosset, P., Martins, T., Williams, S., and Watson, M. (2022). Next generation manufacturing systems architecture.

[Brut et al., 2008] Brut, M., Sedes, F., Jucan, T., Grigoras, R., and Charvillat, V. (2008). An ontology-based modeling approach for developing a competencies-oriented collective intelligence. Springer.

[Bécue et al., 2020] Bécue, A., Maia, E., Feeken, L., Borchers, P., and Praça, I. (2020). A new concept of digital twin supporting optimization and resilience of factories of the future. *Applied Sciences 2020, Vol. 10, Page 4482*, 10:4482.

[Camposano et al., 2021] Camposano, J. C., Smolander, K., and Ruippo, T. (2021). Seven metaphors to understand digital twins of built assets. *IEEE Access*, 9:27167–27181.

[Casati and Varzi, 2003] Casati, R. and Varzi, A. C. (2003). *Parts and Places: The Structures of Spatial Representation*, volume MIT Press. The MIT Press.

[Cho et al., 2019] Cho, S., May, G., and Kiritsis, D. (2019). A semantic-driven approach for industry 4.0. pages 347–354. Institute of Electrical and Electronics Engineers Inc.

[Coraddu et al., 2019] Coraddu, A., Oneto, L., Baldi, F., Cipollini, F., Atlar, M., and Savio, S. (2019). Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. *Ocean Engineering*, 186.

[Correia et al., 2023] Correia, J. B., Abel, M., and Becker, K. (2023). Data management in digital twins: a systematic literature review. *Knowledge and Information Systems*, 65:3165–3196.

[Dahmen and Rossmann, 2021] Dahmen, U. and Rossmann, J. (2021). What is a digital twin - a mediation approach. *IEEE International Conference on Electro Information Technology*, 2021-May:165–172.

[D'Amico et al., 2020] D'Amico, D., Ekoyuncu, J., Addepalli, S., Smith, C., Keedwell, E., Sibson, J., and Penver, S. (2020). Conceptual framework of a digital twin to evaluate the degradation status of complex engineering systems. volume 86, pages 61–67. Elsevier B.V.

[D'Amico et al., 2022] D'Amico, R., Sarkar, A., Karray, H., Addepalli, S., and Erkoyuncu, J. (2022). Detecting failure of a material handling system through a cognitive twin. *IFAC-PapersOnLine*, 55:2725–2730.

[D'Amico et al., 2021] D'Amico, R. D., Addepalli, P., and Erkoyuncu, J. A. (2021). Is a top level ontology based digital twin the solution to human-machine interoperability? *SSRN Electronic Journal*.

[D'Amico et al., 2023] D'Amico, R. D., Addepalli, S., and Erkoyuncu, J. A. (2023). Industrial insights on digital twins in manufacturing: Application landscape, current practices, and future needs. *Big Data and Cognitive Computing 2023, Vol. 7, Page 126*, 7:126.

[Dayananda et al., 2002] Dayananda, J., Irons, R., Harrison, S., Herbohn, J. L., and Rowland, P. (2002). *Capital budgeting: Financial appraisal of investment projects*. Cambridge University Press.

[Declerck et al., 2008] Declerck, T., Krieger, H.-U., Saggion, H., and Spies, M. (2008). Ontology-driven human language technology for semantic-based business intelligence. *ECAI 2008*, pages 841–842.

[Defèr et al., 2020] Defèr, F., Schuh, G., and Stich, V. (2020). Towards a unified reliability-centered information logistics model for production assets. volume 591 IFIP, pages 11–18. Springer, Cham.

[Demoly et al., 2012] Demoly, F., Matsokis, A., and Kiritsis, D. (2012). A mereotopological product relationship description approach for assembly oriented design. *Robotics and Computer-Integrated Manufacturing*, 28:681–693.

[Deng et al., 2021] Deng, Q., Franke, M., Lejardi, E. S., Rial, R. M., and Thoben, K. D. (2021). Development of a digital thread tool for extending the useful life of capital items in manufacturing companies - an example applied for the refurbishment protocol. volume 2021-September. Institute of Electrical and Electronics Engineers Inc.

[Durão et al., 2018] Durão, L. F. C., Haag, S., Anderl, R., Schützer, K., and Zancul, E. (2018). Digital twin requirements in the context of industry 4.0. *IFIP Advances in Information and Communication Technology*, 540:204–214.

[D'Amico et al., 2022] D'Amico, R. D., Erkoyuncu, J. A., Addepalli, S., and Penver, S. (2022). Cognitive digital twin: An approach to improve the maintenance management. *CIRP Journal of Manufacturing Science and Technology*, 38:613–630.

[Eirinakis et al., 2020] Eirinakis, P., Kalaboukas, K., Lounis, S., Mourtos, I., Rozanec, J. M., Stojanovic, N., and Zois, G. (2020). Enhancing cognition for digital twins.

[ElMaraghy and ElMaraghy, 2022] ElMaraghy, H. and ElMaraghy, W. (2022). Adaptive cognitive manufacturing system (acms) – a new paradigm. *International Journal of Production Research*, 60:7436–7449.

[Erkoyuncu et al., 2020] Erkoyuncu, J. A., del Amo, I. F., Ariansyah, D., Bulka, D., Vrabič, R., and Roy, R. (2020). A design framework for adaptive digital twins. *CIRP Annals*, 69:145–148.

[Eyre et al., 2020] Eyre, J., Hyde, S., Walker, D., Ojo, S., Hayes, O., Hartley, R., Scott, R., and Bray, J. (2020). Untangling the requirements of a digital twin.

[Farsi et al., 2021] Farsi, M., Ariansyah, D., Erkoyuncu, J. A., and Harrison, A. (2021). A digital twin architecture for effective product lifecycle cost estimation. volume 100, pages 506–511. Elsevier B.V.

[Fattahi et al., 2021] Fattahi, S., Okamoto, T., and Ura, S. (2021). Preparing datasets of surface roughness for constructing big data from the context of smart manufacturing and cognitive computing. *Big Data and Cognitive Computing 2021, Vol. 5, Page 58*, 5:58.

[Fellows and Liu, 2021] Fellows, R. and Liu, A. (2021). *Research methods for construction*. John Wiley Sons, fifth edition edition.

[Ferko et al., 2023] Ferko, E., Bucaioni, A., Pelliccione, P., and Behnam, M. (2023). Standardisation in digital twin architectures in manufacturing. pages 70–81. Institute of Electrical and Electronics Engineers Inc.

[Fernández et al., 2019] Fernández, F., Ángel Sánchez, Vélez, J. F., and Moreno, A. B. (2019). Symbiotic autonomous systems with consciousness using digital twins. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11487 LNCS:23–32.

[Fernández del Amo et al., 2018] Fernández del Amo, I., Erkoyuncu, J. A., Roy, R., Palmarini, R., and Onoufriou, D. (2018). A systematic review of augmented reality content-related techniques for knowledge transfer in maintenance applications.

[Forrest, 2010] Forrest, P. (2010). Mereotopology without mereology. *Journal of Philosophical Logic*, 39:229–254.

[Frolov et al., 2009] Frolov, V., Mengel, D., Bandara, W., Sun, Y., and Ma, L. (2009). Building an ontology and process architecture for engineering asset management. pages 86–97.

[Fuller et al., 2020] Fuller, A., Fan, Z., Day, C., and Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. *IEEE Access*, 8:108952–108971.

[Galton, 2014] Galton, A. (2014). Discrete mereotopology. *Mereology and the Sciences*, 371:293–321.

[Gasser, 2015] Gasser, U. (2015). Interoperability in the digital ecosystem. *Berkman Center Research Publication Series*.

[Gleim and Decker, 2020] Gleim, L. C. and Decker, S. J. (2020). Open challenges for the management and preservation of evolving data on the web. RWTH Aachen.

[Goddard and Wierzbicka, 2011] Goddard, C. and Wierzbicka, A. (2011). Semantics and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2:125–135.

[Gottschalk, 2009] Gottschalk, P. (2009). Maturity levels for interoperability in digital government. *Government Information Quarterly*, 26:75–81.

[Government, 2012] Government, H. (2012). Open standards principles: For software interoperability, data and document formats in government it specifications.

[Graessler and Poehler, 2017] Graessler, I. and Poehler, A. (2017). Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. *IEEE International Conference on Industrial Engineering and Engineering Management*, 2017-December:289–293.

[Grebenyuk et al., 2021] Grebenyuk, G. G., Kalyanov, G. N., Kovalyov, S. P., Krygin, A. A., Lukinova, O. V., and Nikishov, S. M. (2021). Technological infrastructure management models and methods based on digital twins. Institute of Electrical and Electronics Engineers Inc.

[Grieves and Vickers, 2016] Grieves, M. and Vickers, J. (2016). *Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems*, pages 85–113. Trans-Disciplinary Perspectives on System Complexity.

[Grieves, 2012] Grieves, M. W. (2012). Virtually indistinguishable systems engineering and plm.

[Grieves, 2015] Grieves, M. W. (2015). Digital twin : Manufacturing excellence through virtual factory replication. *White Paper*.

[Grieves, 2019] Grieves, M. W. (2019). *Virtually Intelligent Product Systems: Digital and Physical Twins*, pages 175–200. American Institute of Aeronautics and Astronautics, Inc.

[Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

[Gruber, 1995] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies*, 43:907–928.

[Grüninger and Fox, 1995] Grüninger, M. and Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.

[Guarino and Welty, 2009] Guarino, N. and Welty, C. A. (2009). *An Overview of OntoClean*, pages 201–220. Springer Berlin Heidelberg.

[Gómez-Berbís and de Amescua-Seco, 2019] Gómez-Berbís, J. M. and de Amescua-Seco, A. (2019). Sedit: Semantic digital twin based on industrial iot data management and knowledge graphs. volume 1124 CCIS, pages 178–188. Springer, Cham.

[Götz et al., 2020] Götz, C. S., Karlsson, P., and Yitmen, I. (2020). Exploring applicability, interoperability and integrability of blockchain-based digital twins for asset life cycle management. *Smart and Sustainable Built Environment*.

[Haag and Anderl, 2018] Haag, S. and Anderl, R. (2018). Digital twin – proof of concept. *Manufacturing Letters*, 15:64–66.

[Hahmann and Grüninger, 2012] Hahmann, T. and Grüninger, M. (2012). Region-based theories of space: Mereotopology and beyond. *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions*, pages 1–62.

[Hart, 2018] Hart, C. (2018). *Doing a Literature Review: Releasing the Research Imagination*. SAGE Study Skills Series, second edition edition.

[Hodapp and Hanelt, 2022] Hodapp, D. and Hanelt, A. (2022). Interoperability in the era of digital innovation: An information systems research agenda. *Journal of Information Technology*, 37:407–427.

[Hosamo et al., 2022] Hosamo, H. H., Svennevig, P. R., Svidt, K., Han, D., and Nielsen, H. K. (2022). A digital twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics. *Energy and Buildings*, 261:111988.

[Hsiao, 2003] Hsiao, Y. E. (2003). Semantics and cognition: An introduction.

[Hu et al., 2022] Hu, X., Arista, R., Lentes, J., Lu, J., Zheng, X., Sorvari, J., Ubis, F., and Kiritsis, D. (2022). Ontology-centric industrial requirements validation for aircraft assembly system design. volume 55, pages 3016–3021. Elsevier B.V.

[Huang et al., 2021a] Huang, Y., Dhouib, S., and Malenfant, J. (2021a). Aas capability-based operation and engineering of flexible production lines. volume 2021-September. Institute of Electrical and Electronics Engineers Inc.

[Huang et al., 2021b] Huang, Y., Dhouib, S., and Malenfant, J. (2021b). An aas modeling tool for capability-based engineering of flexible production lines. volume 2021-October. IEEE Computer Society.

[Ihsan et al., 2021] Ihsan, A. Z., Dessì, D., Alam, M., Sack, H., and Sandfeld, S. (2021). Steps towards a dislocation ontology for crystalline materials. In *Second International Workshop on Semantic Digital Twins*.

[Jiang et al., 2021] Jiang, Z., Lv, H., Li, Y., and Guo, Y. (2021). A novel application architecture of digital twin in smart grid. *Journal of Ambient Intelligence and Humanized Computing*.

[Jones et al., 2020] Jones, D., Snider, C., Nassehi, A., Yon, J., and Hicks, B. (2020). Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29:36–52.

[Kainzner et al., 2021] Kainzner, M., Klösch, C., Filipiak, D., Chhetri, T. R., and Fensel, A. (2021). Towards reusable ontology alignment for manufacturing maintenance.

[Kalaboukas et al., 2021] Kalaboukas, K., Rožanec, J., Košmerlj, A., Kiritsis, D., and Arampatzis, G. (2021). Implementation of cognitive digital twins in connected and agile supply networks-an operational model. *Applied Sciences (Switzerland)*, 11.

[Kapteyn et al., 2020] Kapteyn, M. G., Knezevic, D. J., Huynh, D. B., Tran, M., and Willcox, K. E. (2020). Data-driven physics-based digital twins via a library of component-based reduced-order models. *International Journal for Numerical Methods in Engineering*.

[Karray et al., 2021] Karray, M. H., Otte, N., Kiritsis, D., Ameri, F., Kulvatunyou, B., Will, C., Arista, R., Rai, R., and Smith, B. (2021). The industrial ontologies foundry (iof) perspectives.

[Keet et al., 2012] Keet, C. M., Fernández-Reyes, F. C., and Morales-González, A. (2012). Representing mereotopological relations in owl ontologies with ontoparts. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7295 LNCS:240–254.

[Khajavi et al., 2019] Khajavi, S. H., Motlagh, N. H., Jaribion, A., Werner, L. C., and Holmstrom, J. (2019). Digital twin: Vision, benefits, boundaries, and creation for buildings. *IEEE Access*, 7:147406–147419.

[Khan et al., 2022] Khan, S., Arslan, T., and Ratnarajah, T. (2022). Digital twin perspective of fourth industrial and healthcare revolution. *IEEE Access*, 10:25732–25754.

[Kim et al., 2008] Kim, H. L., Passant, A., Breslin, J. G., Scerri, S., and Decker, S. (2008). Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces. In *2008 IEEE International Conference on Semantic Computing*, pages 315–322.

[Kiritsis, 2013] Kiritsis, D. (2013). Semantic technologies for engineering asset life cycle management.

[Koschnick et al., 2015] Koschnick, G., Hankel, M., and Rexroth, B. (2015). Rami 4.0-structure the reference architectural model industrie 4.0 (rami 4.0) contact: Reference architectural model industrie 4.0.

[Koukias et al., 2013] Koukias, A., Nadoveza, D., and Kiritsis, D. (2013). Semantic data model for operation and maintenance of the engineering asset. volume 398, pages 49–55.

[Krishnamenon et al., 2021] Krishnamenon, M., Tuladhar, R., Azghadi, M. R., Loughran, J. G., and Pandey, G. (2021). Digital twins and their significance in engineering asset management. *2021 International Conference on Maintenance and Intelligent Asset Management, ICMIAM 2021*.

[Kritzinger et al., 2018] Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51:1016–1022.

[Kumar et al., 2020] Kumar, J., Bansal, S. A., and Mehta, M. (2020). Reliability analysis - a critical review. *Recent Trends in Industrial and Production Engineering*.

[Lee et al., 2021] Lee, M. J. L., Ho, H. Y., Hsu, L. T., and Au, S. L. M. (2021). Bips: Building information positioning system. Institute of Electrical and Electronics Engineers Inc.

[Lehner and Dorffner, 2020] Lehner, H. and Dorffner, L. (2020). Digital geotwin vienna: Towards a digital twin city as geodata hub. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88:63–75.

[Li et al., 2017] Li, C., Mahadevan, S., Ling, Y., Choze, S., and Wang, L. (2017). Dynamic bayesian network for aircraft wing health monitoring digital twin. *AIAA Journal*, 55:930–941.

[Li et al., 2022] Li, X., Lu, W., Xue, F., Wu, L., Zhao, R., Lou, J., and Xu, J. (2022). Blockchain-enabled iot-bim platform for supply chain management in modular construction. *Journal of Construction Engineering and Management*, 148.

[Li et al., 2021] Li, X., Wang, L., Zhu, C., and Liu, Z. (2021). Framework for manufacturing-tasks semantic modelling and manufacturing-resource recommendation for digital twin shop-floor. *Journal of Manufacturing Systems*, 58:281–292.

[Lim et al., 2020] Lim, K. Y. H., Zheng, P., and Chen, C. H. (2020). A state-of-the-art survey of digital twin: techniques, engineering product lifecycle management and business innovation perspectives. *Journal of Intelligent Manufacturing*, 31:1313–1337.

[Liu et al., 2022] Liu, Y. K., Ong, S. K., and Nee, A. Y. C. (2022). State-of-the-art survey on digital twin implementations. *Advances in Manufacturing*, 10.

[Liu et al., 2019] Liu, Z., Chen, W., Zhang, C., Yang, C., and Chu, H. (2019). Data super-network fault prediction model and maintenance strategy for mechanical product based on digital twin. *IEEE Access*, 7:177284–177296.

[Longo et al., 2019] Longo, F., Nicoletti, L., and Padovano, A. (2019). Ubiquitous knowledge empowers the smart factory: The impacts of a service-oriented digital twin on enterprises' performance. *Annual Reviews in Control*, 47:221–236.

[Lu et al., 2021] Lu, J., Ma, J., Zheng, X., Wang, G., Li, H., and Kiritsis, D. (2021). Design ontology supporting model-based systems engineering formalisms. *IEEE Systems Journal*.

[Lu et al., 2022] Lu, J., Yang, Z., Zheng, X., Wang, J., and Kiritsis, D. (2022). Exploring the concept of cognitive digital twin from model-based systems engineering perspective. *International Journal of Advanced Manufacturing Technology*, 121:5835–5854.

[Lu et al., 2020a] Lu, J., Zheng, X., Gharaei, A., Kalaboukas, K., and Kiritsis, D. (2020a). Cognitive twins for supporting decision-makings of internet of things systems. pages 105–115.

[Lu et al., 2011] Lu, Q., Xie, X., Heaton, J., Parlikad, A. K., and Schooling, J. (2011). From bim towards digital twin: Strategy and future development for smart asset management.

[Lu et al., 2020b] Lu, Q., Xie, X., Parlikad, A. K., and Schooling, J. M. (2020b). Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance. *Automation in Construction*, 118.

[Luo et al., 2019] Luo, W., Hu, T., Zhang, C., and Wei, Y. (2019). Digital twin for cnc machine tool: modeling and using strategy. *Journal of Ambient Intelligence and Humanized Computing*, 10:1129–1140.

[Macchi et al., 2018] Macchi, M., Roda, I., Negri, E., and Fumagalli, L. (2018). Exploring the role of digital twin for asset lifecycle management. volume 51, pages 790–795. Elsevier B.V.

[Madni et al., 2019] Madni, A. M., Madni, C. C., and Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. *Systems*, 7.

[Martín-Recuerda et al., 2020] Martín-Recuerda, F., Walther, D., Eisinger, S., Moore, G., Andersen, P., Opdahl, P. O., and Hella, L. (2020). Revisiting ontologies of units of measure for harmonising quantity values – a use case. volume 12507 LNCS, pages 551–567. Springer Science and Business Media Deutschland GmbH.

[May et al., 2022] May, M. C., Kiefer, L., Kuhnle, A., and Lanza, G. (2022). Ontology-based production simulation with ontologysim. *Applied Sciences (Switzerland)*, 12.

[Merrell et al., 2022] Merrell, E. C., Limbaugh, D. G., Ac, P. M. K., and Smith, B. (2022). Capabilities.

[Meyers et al., 2022] Meyers, B., Noten, J. V., Lietaert, P., Tielemans, B., Hristov, H., Maes, D., and Gadeyne, K. (2022). Knowledge graphs in digital twins for manufacturing - lessons learned from an industrial case at atlas copco airpower. *IFAC-PapersOnLine*, 55:13–18.

[Mitchell et al., 2021] Mitchell, D., Blanche, J., Zaki, O., Roe, J., Kong, L., Harper, S., Robu, V., Lim, T., and Flynn, D. (2021). Symbiotic system of systems design for safe and resilient autonomous robotics in offshore wind farms. *IEEE Access*, 9:141421–141452.

[Mohammadi et al., 2020] Mohammadi, N., Vimal, A., and Taylor, J. E. (2020). Knowledge discovery in smart city digital twins. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2020-January:1656–1664.

[Montavon et al., 2020] Montavon, B., Dahlem, P., Peterek, M., and Schmitt, R. H. (2020). A digital perspective on machine tool calibration. *International Journal of Automation Technology*, 14:360–368.

[Moretti et al., 2020] Moretti, N., Xie, X., Merino, J., Brazauskas, J., and Parlikad, A. K. (2020). An openbim approach to iot integration with incomplete as-built data. *Applied Sciences (Switzerland)*, 10:1–17.

[Mortlock et al., 2022] Mortlock, T., Muthirayan, D., Yu, S. Y., Khargonekar, P. P., and Faruque, M. A. A. (2022). Graph learning for cognitive digital twins in manufacturing systems. *IEEE Transactions on Emerging Topics in Computing*, 10:34–45.

[Musiolik, 2021] Musiolik, T. (2021). Technology vision 2021 - leaders wanted - masters of change at a moment of truth.

[Najdataei et al., 2019] Najdataei, H., Subramaniyan, M., Gulisano, V., Skoogh, A., and Papatriantafilou, M. (2019). Stream-it: Continuous and dynamic processing of production systems data - throughput bottlenecks as a case-study. volume 2019-June, pages 1328–1333. Institute of Electrical and Electronics Engineers Inc.

[National Infrastructure Commission, 2018] National Infrastructure Commission, N. (2018). Data for the public good.

[Noy and McGuinness, 2001] Noy, N. F. and McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. *Stanford Knowledge Systems Laboratory*, page 25.

[Ocker et al., 2022] Ocker, F., Vogel-Heuser, B., Schon, H., and Mieth, R. (2022). Leveraging digital twins for compatibility checks in production systems engineering. pages 103–107. Institute of Electrical and Electronics Engineers (IEEE).

[Oliveira, 2020] Oliveira, P. P. (2020). Digital twin development for airport management. *Journal of Airport Management*, 14:246–259.

[Palmarini et al., 2018] Palmarini, R., Erkoyuncu, J. A., Roy, R., and Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49:215–228.

[Partridge et al., 2020] Partridge, C., Mitchell, A., Cook, A., Leal, D., Sullivan, J., and West, M. (2020). A survey of top-level ontologies.

[Passant et al., 2009] Passant, A., Samwald, M., Breslin, J. G., and Decker, S. (2009). Federating distributed social data to build an interlinked online information society. *IEEE Intelligent Systems*, 24(6):44–48.

[Phillips et al., 2011] Phillips, J. J., Brantley, W., and Phillips, P. P. (2011). *Project Management ROI: A Step-by-Step Guide for Measuring the Impact and ROI for Projects*. John Wiley Sons, Inc.

[Poveda-Villalón et al., 2022] Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., and García-Castro, R. (2022). Lot: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, 111:104755.

[Prenzel and Provost, 2019] Prenzel, L. and Provost, J. (2019). Fbbeam: An erlang-based iec 61499 implementation.

[Price and Bodington, 2004] Price, D. and Bodington, R. (2004). Applying semantic web technology to the life cycle support of complex engineering assets. *The SemanticWeb – ISWC 2004*.

[Rahm et al., 2017] Rahm, J., Graube, M., and Urbas, L. (2017). A proposal for an interactive roundtrip engineering system.

[Randall, 2011] Randall, R. B. (2011). *Vibration-based condition monitoring industrial, aerospace and automotive applications*. Wiley.

[Rasheed et al., 2020] Rasheed, A., San, O., and Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 8:21980–22012.

[Raubal and Adams, 2010] Raubal, M. and Adams, B. (2010). The semantic web needs more cognition. *Semantic Web*, 1:69–74.

[Redeker et al., 2021] Redeker, M., Weskamp, J. N., Rossl, B., and Pethig, F. (2021). Towards a digital twin platform for industrie 4.0. pages 39–46. Institute of Electrical and Electronics Engineers Inc.

[Riemer, 2019] Riemer, D. (2019). Feeding the digital twin: Basics, models and lessons learned from building an iot analytics toolbox (invited talk). page 4212. Institute of Electrical and Electronics Engineers Inc.

[Robinson et al., 2015] Robinson, I., Webber, J., and Eifrem, E. (2015). *Graph databases: New Opportunities for Connected Data*. O'Reilly.

[Rosen et al., 2019] Rosen, R., Fischer, J., and Boschert, S. (2019). Next generation digital twin: an ecosystem for mechatronic systems? *IFAC-PapersOnLine*, 52:265–270.

[Rožanec et al., 2020] Rožanec, J. M., Lu, J., Kosmerlj, A., Kenda, K., Kiritsis, D., Jovanoski, V., Rupnik, J., Karloveec, M., and Fortuna, B. (2020). Towards actionable cognitive digital twins for manufacturing.

[Rožanec et al., 2021] Rožanec, J. M., Lu, J., Rupnik, J., Škrjanc, M., Mladenić, D., Fortuna, B., Zheng, X., and Kiritsis, D. (2021). Actionable cognitive twins for decision making in manufacturing. *International Journal of Production Research*, 60:452–478.

[Sahal et al., 2021] Sahal, R., Alsamhi, S. H., Breslin, J. G., Brown, K. N., and Ali, M. I. (2021). Digital twins collaboration for automatic erratic operational data detection in industry 4.0. *Applied Sciences (Switzerland)*, 11.

[Sahlab et al., 2021] Sahlab, N., Kamm, S., Muller, T., Jazdi, N., and Weyrich, M. (2021). Knowledge graphs as enhancers of intelligent digital twins. pages 19–24. Institute of Electrical and Electronics Engineers Inc.

[Salama et al., 2021] Salama, F., Korkan, E., Käbisch, S., and Steinhorst, S. (2021). Towards a behavioral description of cyber-physical systems using the thing description. pages 6–9. Association for Computing Machinery, Inc.

[Salustri, 2002] Salustri, F. A. (2002). Mereotopology for product modelling. a new framework for product modelling based on logic. *J. of Design Research*, 2:0.

[Santos et al., 2019] Santos, S. M., Icyuz, M., Pound, I., William, D., Guo, J., McKinney, B. A., Niederweis, M., Rodgers, J., and Hartman, J. L. (2019). A humanized yeast phenomic model of deoxycytidine kinase to predict genetic buffering of nucleoside analog cytotoxicity. *Genes*, 10.

[Sarkar et al., 2020] Sarkar, A., Šormaz, D., Koonce, D., and Farah, S. (2020). Developing a resource-based manufacturing process capability ontology. *Communications in Computer and Information Science*, 1407 CCIS:293–306.

[Semeraro et al., 2021] Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. *Computers in Industry*, 130:103469.

[Sicilia, 2005] Sicilia, M.-A. (2005). Ontology-based competency management: Infrastructures for the knowledge intensive learning organization. *Intelligent Learning Infrastructure for Knowledge Intensive Organizations: A Semantic Web Perspective*.

[Sider, 1997] Sider, T. (1997). Four dimensionalism. *Philosophical Review*, 106:197–231.

[Singh et al., 2021] Singh, S., Shehab, E., Higgins, N., Fowler, K., Erkoyuncu, J. A., and Gadd, P. (2021). Towards information management framework for digital twin in aircraft manufacturing. *Procedia CIRP*, 96:163–168.

[Smith, 1996] Smith, B. (1996). Mereotopology: A theory of parts and boundaries.

[Smith et al., 2007] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S. A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S. (2007). The obo foundry: Coordinated evolution of ontologies to support biomedical data integration.

[Smith et al., 2008] Smith, B., Kusnierczyk, W., Schober, D., and Ceusters, W. (2008). Towards a reference terminology for ontology research and development in the biomedical domain.

[Solso et al., 2008] Solso, R. L., MacLin, O. H., and MacLin, M. K. (2008). *Cognitive psychology*. Pearson/Allyn and Bacon.

[Somers et al., 2020] Somers, S., Oltramari, A., and Lebiere, C. (2020). Cognitive twin: A cognitive approach to personalized assistants.

[Stark et al., 2019] Stark, R., Fresemann, C., and Lindow, K. (2019). Development and operation of digital twins for technical systems and services. *CIRP Annals*, 68:129–132.

[Steiner and Albert, 2017] Steiner, C. M. and Albert, D. (2017). Validating domain ontologies: A methodology exemplified for concept maps. *Cogent Education*, 4.

[Suh, 1998] Suh, N. P. (1998). Axiomatic design theory for systems. *Research in Engineering Design*, pages 189–209.

[Tao et al., 2019] Tao, F., Zhang, H., Liu, A., and Nee, A. Y. (2019). Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, 15:2405–2415.

[Tao and Zhang, 2017] Tao, F. and Zhang, M. (2017). Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing. *IEEE Access*, 5:20418–20427.

[Tao et al., 2020] Tao, X., Conzon, D., Ferrera, E., Maillet-Contoz, L., Michel, E., Diaz-Nava, M., Li, S., Baouya, A., Chehida, S., and Goetz, J. (2020). Model based methodology and framework for design and management of next-gen iot systems.

[Timperi et al., 2023] Timperi, M., Kokkonen, K., Hannola, L., and Elfvengren, K. (2023). Impacts of digital twins on new business creation: insights from manufacturing industry. *Measuring Business Excellence*, ahead-of-print.

[Trauer et al., 2020] Trauer, J., Schweigert-Recksiek, S., Okamoto, L. O., Spreitzer, K., Mörtl, M., and Zimmermann, M. (2020). Data-driven engineering definitions and insights from an industrial case study for a new approach in technical product development. The Design Society.

[Treleaven et al., 2022] Treleaven, P., Smietanka, M., and Pithadia, H. (2022). Federated learning: The pioneering distributed machine learning and privacy-preserving data technology. *Computer*, 55:20–29.

[Tuli et al., 2021] Tuli, T. B., Kohl, L., Chala, S. A., Manns, M., and Ansari, F. (2021). Knowledge-based digital twin for predicting interactions in human-robot collaboration. volume 2021-September. Institute of Electrical and Electronics Engineers Inc.

[Törmä et al., 2019] Törmä, S., Toivola, P., Kiviniemi, M., Puntila, P., Lampi, M., and Mätäsniemi, T. (2019). Ontology-based sharing of structural health monitoring data. pages 2214–2221. International Association for Bridge and Structural Engineering (IABSE).

[Vakaj and Cheung, 2021] Vakaj, E. K. and Cheung, F. K. T. (2021). A semantic offsite construction digital twin-offsite manufacturing production workflow (opw) ontology semantic technologies for linked building data view project semcon: Semantic and contextual objective metric view project.

[Varga, 2010] Varga, M. (2010). Challenges of data management in always-on enterprise information systems. *Always-On Enterprise Information Systems for Business Continuance: Technologies for Reliable and Scalable Operations*, pages 109–128.

[Varzi, 1996] Varzi, A. C. (1996). Parts, wholes, and part-whole relations: The prospects of mereotopology. *Data Knowledge Engineering*, 20:259–286.

[Varzi, 1998] Varzi, A. C. (1998). Basic problems of mereotopology. pages 29–38. IOS Press.

[Velosa and Kutnick, 2017] Velosa, A. and Kutnick, D. (2017). Digital twins will impact economic and business models.

[Verzelen et al., 2021] Verzelen, F., Lacy, P., and Stacey, N. (2021). The critical role of virtual twins in accelerating sustainability.

[Vogel-Heuser et al., 2021] Vogel-Heuser, B., Ocker, F., Weiß, I., Mieth, R., and Mann, F. (2021). Potential for combining semantics and data analysis in the context of digital twins. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379.

[Vrana, 2021] Vrana, J. (2021). The core of the fourth revolutions: Industrial internet of things, digital twin, and cyber-physical loops. *Journal of Nondestructive Evaluation*, 40.

[Wanasinghe et al., 2020] Wanasinghe, T. R., Wroblewski, L., Petersen, B. K., Gosine, R. G., James, L. A., Silva, O. D., Mann, G. K., and Warrian, P. J. (2020). Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. *IEEE Access*, 8:104175–104197.

[Wang et al., 2019] Wang, W., Tian, G., Cao, L., Liu, J., Zhou, H., Liu, X., and Wu, M. F. (2019). Dynamic evaluation method of machining process planning based on the digital twin-based process model. *IEEE Access*, 7:1–1.

[Wilkinson et al., 2016] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Lei, J. V. D., Mulligen, E. V., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). Comment: The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3.

[Wright and Davidson, 2020] Wright, L. and Davidson, S. (2020). How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences*, 7.

[Xiao and Li, 2021] Xiao, B. and Li, F. (2021). Knowledge-based formal modeling for cpps in personalized intelligent manufacturing. pages 371–378. IEEE.

[Xie and Wan, 2023] Xie, J. and Wan, J. (2023). Digital twin four-dimension fusion modeling method design and application to the discrete manufacturing line. *Big Data and Cognitive Computing 2023, Vol. 7, Page 89*, 7:89.

[Xie et al., 2020] Xie, X., Lu, Q., Rodenas-Herraiz, D., Parlikad, A. K., and Schooling, J. M. (2020). Visualised inspection system for monitoring environmental anomalies during daily operation and maintenance. *Engineering, Construction and Architectural Management*, 27:1835–1852.

[Xu et al., 2019] Xu, M., Chen, X., and Kou, G. (2019). A systematic review of blockchain. *Financial Innovation*, 5.

[Yitmen et al., 2021] Yitmen, I., Alizadehsalehi, S., İlknur Akıner, and Akıner, M. E. (2021). An adapted model of cognitive digital twins for building lifecycle management. *Applied Sciences (Switzerland)*, 11.

[Yu et al., 2021] Yu, G., Wang, Y., Mao, Z., Hu, M., Sugumaran, V., and Wang, Y. K. (2021). A digital twin-based decision analysis framework for operation and maintenance of tunnels. *Tunnelling and Underground Space Technology*, 116.

[Yun et al., 2017] Yun, S., Park, J. H., and Kim, W. T. (2017). Data-centric middleware based digital twin platform for dependable cyber-physical systems. *International Conference on Ubiquitous and Future Networks, ICUFN*, pages 922–926.

[Zaki et al., 2020] Zaki, O. F., Flynn, D., Blanche, J. R. D., Roe, J. K., Kong, L., Mitchell, D., Lim, T., Harper, S. T., and Robu, V. (2020). Self-certification and safety compliance for robotics platforms. volume 2020-May. Offshore Technology Conference.

[Zehnder and Riemer, 2019] Zehnder, P. and Riemer, D. (2019). Representing industrial data streams in digital twins using semantic labeling. pages 4223–4226. Institute of Electrical and Electronics Engineers Inc.

[Zhang et al., 2020] Zhang, C., Zhou, G., Hu, J., and Li, J. (2020). Deep learning-enabled intelligent process planning for digital twin manufacturing cell. *Knowledge-Based Systems*, 191.

[Zheng et al., 2022a] Zheng, X., Hu, X., Arista, R., Lu, J., Sorvari, J., Lentes, J., Ubis, F., and Kiritsis, D. (2022a). A semantic-driven tradespace framework to accelerate aircraft manufacturing system design. *Journal of Intelligent Manufacturing*.

[Zheng et al., 2022b] Zheng, X., Hu, X., Arista, R., Lu, J., Sorvari, J., Lentes, J., Ubis, F., and Kiritsis, D. (2022b). A semantic-driven tradespace framework to accelerate aircraft manufacturing system design. *Journal of Intelligent Manufacturing*, pages 1–24.

[Zheng et al., 2022c] Zheng, X., Lu, J., and Kiritsis, D. (2022c). The emergence of cognitive digital twin: vision, challenges and opportunities. *International Journal of Production Research*, 60:7610–7632.

[Zhiqiang et al., 2019] Zhiqiang, Z., Quan, C. W., Xiaoming, D., and Ashok, P. V. (2019). Smart factory of industry 4.0: Connection infrastructure, data acquisition, data processing, and applications.

[Zikmund, 2013] Zikmund, W. G. (2013). Business research methods. *South-Western College Pub*, page 63.

[Zimmermann, 1980] Zimmermann, H. (1980). Osi reference model-the iso model of architecture for open systems interconnection. *IEEE Transactions on Communications*, 28:425–432.

[Zinnikus et al., 2017] Zinnikus, I., Antakli, A., Kapahnke, P., Klusch, M., Krauss, C., Nonnengart, A., and Slusallek, P. (2017). Integrated semantic fault analysis and worker support for cyber-physical production systems. volume 1, pages 207–216. Institute of Electrical and Electronics Engineers Inc.

[Ágota Bányai et al., 2019] Ágota Bányai, Illés, B., Glistau, E., Machado, N. I. C., Tamás, P., Manzoor, F., and Bányai, T. (2019). Smart cyber-physical manufacturing: Extended and real-time optimization of logistics resources in matrix production. *Applied Sciences*, 9:1287.