Joint Research Programme
BTO 2023.080 | October 2023

A Roadmap to Creating Digital Twins for Drinking Water Treatment

Joint Research Programme

Report

A Roadmap to Creating Digital Twins for Drinking Water Treatment

BTO 2023.080 | October 2023

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045/310

Project manager
Ina Vertommen

Client
BTO - Thematical research - Hydroinformatics

Authors
Siddharth Seshan
Mollie Torello
Dr Martin Korevaar
Dr Mark Morley

Quality Assurance
Dr Peter van Thienen

Sent to
This report is distributed to BTO-participants.

A year after publication it is public.

Keywords
Digital Twin

Year of publishing
2023

More information
Siddharth Seshan, M.Sc.
T +31 6 15883975
E siddharth.seshan@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

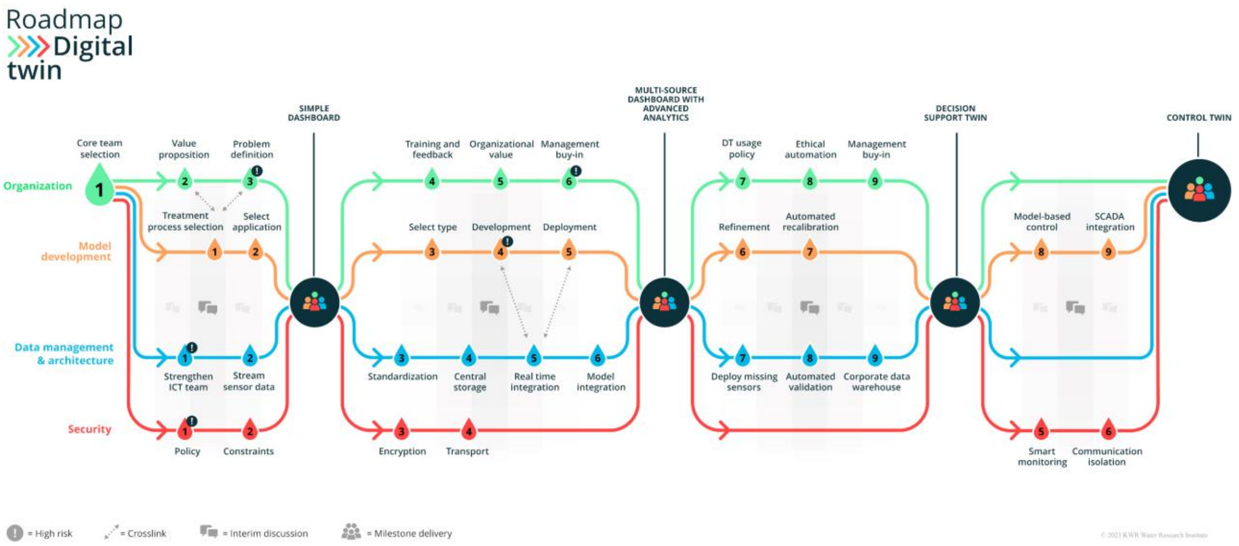
T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

Management Summary

A Roadmap to Creating Digital Twins for Drinking Water Treatment

Authors: Siddharth Seshan, Mollie Torello, Dr Martin Korevaar, Dr Mark Morley

The concept of Digital Twins, as evolving representations of physical systems, has gained substantial traction across diverse fields including drinking water treatment. However, realizing the full potential of Digital Twins faces significant challenges, due to a lack of understanding and the absence of standardized terminologies and practices. To address this, we provide a comprehensive review, defining Digital Twins and exploring various data architectural frameworks for their implementation alongside modelling approaches for drinking water treatment processes. Subsequently, we introduce a roadmap developed for Digital Twin implementation within drinking water treatment, elucidating various organizational and technological paths conceptualised based on the literature, interviews with leading organisations and domain expertise. Finally, a functional design is presented, which serves as the blueprint for putting the roadmap into action for a real-world application. It focuses on a case study at De Watergroep, specifically a water softening treatment process. This design provides a comprehensive framework and analytical methodology to understand stakeholder motivations, identify the Digital Twin's goals formulate the requirements and application services that can be included in the design and implementation of a Digital Twin for the process.



The developed Digital Twin Roadmap

Importance: Demystifying and Clarifying Digital Twins

The increasing popularity of Digital Twins has resulted in inconsistent terminology and a lack of standardized methods for implementation. Water companies need clarity on the necessary organizational and technical prerequisites for implementing Digital Twins and specific guidance for to embark on developing and integrating Digital Twins to enhance their drinking water treatment processes.

Method: Literature Review, Interviews and Functional Design Thinking

A comprehensive literature review covered aspects such as Digital Twin definitions, types of Digital Twins, industrial examples, data architectural frameworks and modelling of drinking water treatment processes.

In addition, interviews were conducted with leading organisations where key information is gathered on the experiences gained from real-world applications. We explored various data architectural frameworks for their implementation alongside modelling approaches for drinking water treatment processes. Finally, we used the functional design methodology and design thinking based on stakeholder engagement and Enterprise Architectural Framework principles on a case study from De Watergroep. Specifically a softening treatment process.

Results: Digital Twin Roadmap, Functional Design
A Digital Twin is defined as a digital copy (virtual model) of a physical system. This model is continually fed with data and provides a glimpse into the asset's past, present and future behaviour. We developed a robust and comprehensive roadmap for creating and integrating Digital Twins for drinking water treatment processes. The roadmap was applied for the functional design of a pilot case treatment unit, which includes the motivations, requirements and identified Digital Twin applications.

Application: Use the roadmap to guide Digital Twin implementation
Water companies are currently undergoing major organisational and technical changes as part of their digital transformation journey and digitalisation strategy. Water companies should focus on interdisciplinary collaboration, data infrastructure and technology assessment and organizational readiness to effectively implement Digital Twins. This report and the developed roadmap serve as guiding principles to facilitate this transition.

Report
This research is reported in *A Roadmap to Creating Digital Twins for Drinking Water Treatment* (BTO-2023.080).

More information
Siddharth Seshan, M.Sc.
T +31 6 15883975
E siddharth.seshan@kwrwater.nl

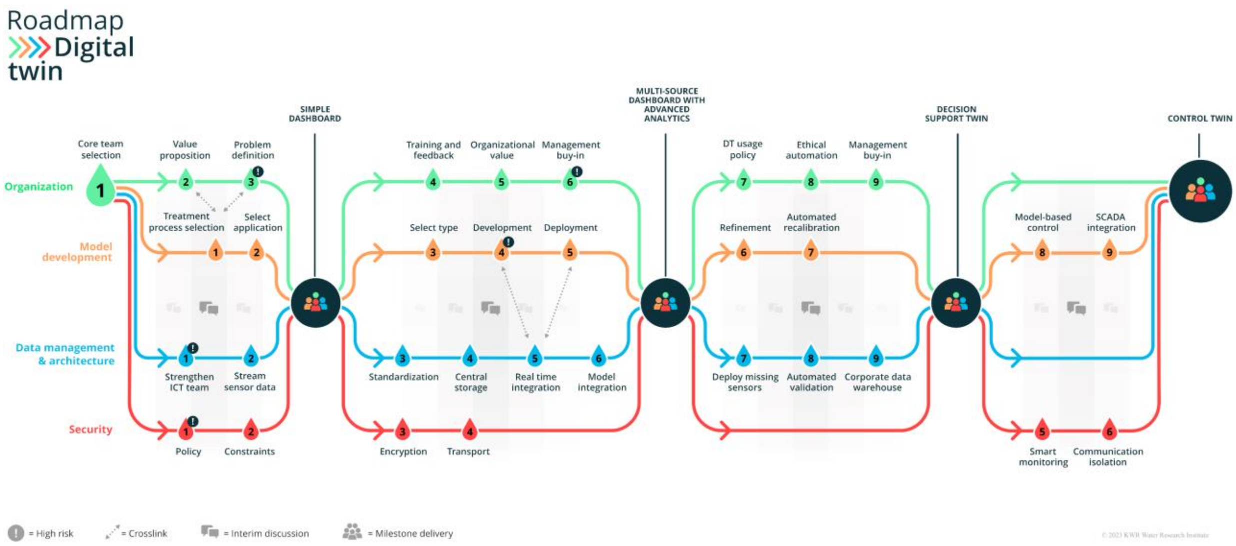
PO Box 1072
3430 BB Nieuwegein
The Netherlands

Managementsamenvatting

Een roadmap voor het creëren van Digital Twins voor drinkwaterbehandeling

Auteur(s): Siddharth Seshan, Mollie Torello, Dr Martin Korevaar, Dr Mark Morley

Het concept van Digital Twins, als zich ontwikkelende representaties van fysieke systemen, krijgt op verschillende gebieden veel aandacht, inclusief de drinkwaterbehandeling. Er bestaan echter grote uitdagingen rond het realiseren van het volledige potentieel van Digital Twins, zoals een gebrek aan begrip en het ontbreken van gestandaardiseerde terminologie en praktijken. Om dit aan te pakken, bieden we een uitgebreide review, waarin Digital Twins worden gedefinieerd en verschillende gegevensarchitectuurkaders voor hun implementatie worden onderzocht, naast modelleringsbenaderingen voor drinkwaterbehandelingsprocessen. Vervolgens introduceren we een stappenplan dat is ontwikkeld voor de implementatie van Digital Twins binnen de drinkwaterbehandeling, waarbij verschillende organisatorische en technologische paden worden belicht die zijn geconceptualiseerd op basis van de literatuur, interviews met toonaangevende organisaties en domeinkennis. Tot slot wordt een functioneel ontwerp gepresenteerd, dat dient als blauwdruk voor het in praktijk brengen van het stappenplan voor een toepassing in de echte wereld. Het richt zich op een case study van De Watergroep, specifiek een wateronthardingsproces. Dit ontwerp biedt een uitgebreid raamwerk en analytische methodologie om de motivaties van belanghebbenden te begrijpen, de doelen van de Digital Twin te identificeren en de vereisten en applicatieservices te formuleren die kunnen worden opgenomen in het ontwerp en de implementatie van een Digital Twin voor drinkwaterbehandelingsprocessen.



De roadmap voor een Digital Twin

Belang: Digital Twins verduidelijken en verhelderen
De toenemende populariteit van Digital Twins heeft geleid tot inconsistente terminologie en een gebrek aan gestandaardiseerde methoden voor implementatie. Waterbedrijven hebben duidelijkheid nodig over de noodzakelijke organisatorische en technische voorwaarden voor het implementeren van Digital Twins en specifieke begeleiding om te beginnen met het ontwikkelen en integreren van

Digital Twins om hun drinkwaterbehandelings processen te verbeteren.

Aanpak: Literatuuroverzicht, interviews en functioneel ontwerpdenken
Een uitgebreide literatuurstudie behandelde aspecten zoals definities van Digital Twin, soorten Digital Twins, industriële voorbeelden, raamwerken voor gegevensarchitectuur en modellering van drinkwaterbehandelingsprocessen. Daarnaast werden interviews gehouden met toonaangevende organisaties om belangrijke informatie te verzamelen over de ervaringen die zijn opgedaan met toepassingen in de praktijk. We onderzochten verschillende raamwerken voor gegevensarchitectuur voor hun implementatie, naast modelleringsbenaderingen voor drinkwaterbehandelingsprocessen. Tot slot gebruikten we de functional design-methodologie en design thinking op basis van betrokkenheid van belanghebbenden en Enterprise Architectural Framework-principes op een casestudy van De Watergroep, specifiek een wateronthardingsproces.

Resultaten: Digital Twin Roadmap en functional design
Een Digital Twin wordt gedefinieerd als een digitale kopie (virtueel model) van een fysiek systeem. Dit model wordt voortdurend gevoed met gegevens en biedt een kijkje in het verleden, het heden en het toekomstige gedrag van het bedrijfsmiddel. We ontwikkelden een robuust en uitgebreid stappenplan voor het creëren en integreren van Digital Twins voor drinkwaterbehandelingsprocessen. Het stappenplan werd toegepast voor het functionele ontwerp van een pilot case behandelingseenheid, met daarin de motivaties en vereisten en geïdentificeerde Digital Twin toepassingen.

Toepassing: gebruik de routekaart om implementatie van een Digital Twin te begeleiden
Waterbedrijven ondergaan momenteel grote organisatorische en technische veranderingen als onderdeel van hun digitale transformatietraject en digitaliseringsstrategie. Waterbedrijven moeten zich richten op interdisciplinaire samenwerking, beoordeling van data-infrastructuur en -technologie en organisatorische gereedheid om Digital Twins effectief te implementeren. Dit rapport en de ontwikkelde routekaart kunnen als leidraad dienen om deze overgang te vergemakkelijken.

Rapport
Dit onderzoek is beschreven in het rapport *A Roadmap to Creating Digital Twins for Drinking Water Treatment* (BTO-2023.080).

More information
Siddharth Seshan, M.Sc.
T +31 6 15883975
E siddharth.seshan@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

KWR

Contents

List of Figures

List of Tables

Acknowledgements

We would like to extend our heartfelt gratitude to the esteemed members of the Steering Committee, comprising Arend te Boveldt (Brabant Water), Rob van Ewijk (Waternet), Nele Philips (De Watergroep), Elina Krapels (Dunea) and Nicael Jooste (Dunea). Their invaluable insights, unwavering support and commitment to this endeavour have been instrumental in the success of the Digital Twin Zuivering BTO project.

A special note of appreciation goes to Nele Philips and her dedicated team at De Watergroep for their exceptional contributions to the functional design. Their expertise and diligent efforts have significantly enriched our project, making it possible to craft a tailored and effective solution for digitalisation challenges.

We would also like to thank TNO for their generosity in sharing their profound knowledge and experience as digital twin experts. Their willingness to collaborate and provide insights has been invaluable.

We are also deeply thankful to Herbert ter Maat for stepping in and assisting with the functional design interviews. His expertise and contributions have been invaluable in this critical phase of the project.

Last, but not least, our sincere appreciation goes to Vitens and Waternet for their active participation in the interviews. As trailblazers in digitalization within the water sector, their willingness to share their experiences and expertise has been a cornerstone of this project's success.

We extend our gratitude to all individuals and organizations who have played a part in shaping this report and helping us move forward in our mission to revolutionize the water industry through digital twins.

1 Introduction

1.1 Description

This report describes the development of a roadmap for implementing a Digital Twin for drinking water treatment processes, enabling almost real-time water quality monitoring. Building upon existing knowledge within and outside the sector, the roadmap guides the development of Digital Twins, which represent virtual copies or models of real systems. These Digital Twins mimic physical processes and rely on the same data inputs as the actual processes. For drinking water treatment, a Digital Twin comprises process models fed by data from a data management or process information system. Depending on its functionality, a Digital Twin can provide insights into water quality, enhance monitoring and enable more efficient management of the process assets and operation. It forms a crucial foundation for information-driven control and automation of treatment processes, offering real-time views of various water quality parameters through physical-chemical models and sensor data fusion. This project seeks to advance water purification processes by integrating Digital Twin technology, offering a systematic roadmap that offers both a comprehensive approach and guidelines for implementation. The outcomes are expected to empower water companies with real-time data, leading to optimized treatment processes and more sustainable production of high-quality drinking water.

1.2 Purpose and Significance

The project responds to the growing need for organizations to reduce reliance on employee experience and shift towards more data-driven decision-making whilst seeking to retain operator knowledge about the systems involved. Digital Twins can play a key role in this transition, offering a virtual representation of reality. The roadmap outlines the steps involved in developing a Digital Twin specifically for water treatment process, ensuring not only technical integration but also addressing organizational aspects such as data governance, skill development and structural changes.

1.3 Outputs and Applications

The outcomes of this study include a literature review of practical and scientific knowledge related to Digital Twins, a shared understanding of Digital Twin concepts within the Dutch and Flemish drinking water sector and a detailed roadmap for developing Digital Twins for water treatment processes. The roadmap covers aspects such as the definition of the purpose of the Digital Twin, technical integration with water company data systems, organizational recommendations and guidelines for adopting standards and development platforms.

1.4 Target Audience

This report seeks to inform innovators, management, ICT professionals and practitioners who are interested in or tasked with the early stages of the development of a Digital Twin for water treatment processes. It is also relevant to the wider audience of stakeholders in this domain. Those involved in this project have included process technologists, ICT professionals, data engineers, data architects, management and innovation professionals within water companies. The relevance of the Digital Twin to these stakeholders is readily apparent: Process technologists benefit from real-time water quality insights, while data engineers are able to use Digital Twins to simplify their analyses and collaborate with technologists. ICT professionals provide technical support and data scientists and researchers generate insights and apply them. Collaboration between different domain experts, including process technologists, asset managers, ICT professionals and data managers, is crucial for the successful implementation, adoption and maintenance of a Digital Twin.

1.5 Activities

The project activities involved gathering scientific knowledge, exploring existing implementations of Digital Twins, identifying relevant Key Performance Indicators (KPIs), assessing current data infrastructure at water companies and developing a comprehensive roadmap. A functional design of a case study water treatment process, water softening at De Watergroep, has been developed with intensive stakeholder-engagement to both inform the design of the roadmap and to serve as an example of its application. The functional design encompasses an overview of the treatment process selected as the case study, the procedure of selecting a core team, identifying an appropriate value proposition and a set of appropriate Application Services to implement a successful Digital Twin. The design further considers the necessary steps in elucidating the drivers and concerns of the process stakeholders, deriving goals for the Digital Twin from them and clarifying the Design Principles and Requirements that lead to the identification of the Application Services and Outcomes required.

2 Literature Review

2.1 Introduction

The concept of Digital Twins, as an evolving representation of physical systems, has much interest across diverse fields. Digital Twins offer real-time insights, facilitate simulations and promise to revolutionize industries by enhancing decision-making and system optimization. However, the full realization of Digital Twin technology's potential faces significant challenges, primarily related to the lack of standardized terminologies and practices. This impedes the identification of commonalities in frameworks and makes it challenging for users to navigate the field.

The architectural implementation of Digital Twins is tackled in detail in the literature. The Reference Architecture Model establishes a common framework for Industry 4.0 solutions, aiding organizations and stakeholders in integrating technologies and defining Digital Twin applications. It includes components such as physical devices, edge devices for data processing, network infrastructure for data exchange, cloud infrastructure for storage and analysis, software applications for data analysis and decision-making and user interfaces for interaction with other systems.

In addition to addressing issues related to standardization in Digital Twin architectures, there is growing attention to the development of models that underpin these digital representations. Model development is recognized as an integral component for the successful implementation of Digital Twins. Models serve as alternative representations of real-world systems, encoding knowledge and information acquired from prior experience and processed data. Particularly in the context of water treatment systems, these models often rely on mathematical equations to define temporal and spatial relationships among variables, explaining various processes (Therrien *et al*., 2020). With the growing popularity of artificial intelligence (AI), data-driven methods and the upcoming field of hybrid modelling, the choice of model types and technologies plays a key role for Digital Twins and can vary for different treatment processes.

In the following sections, we delve into a comprehensive review starting by defining what a Digital Twin is, discuss the various stages and types of Digital Twins and present industrial example. We then examine different data architectural frameworks used in Digital Twin implementations. Additionally, we explore various modelling approaches for DWT processes, including process, data-driven and hybrid modelling. Finally, we introduce a decision framework to guide the selection of the model type for a Digital Twin, considering the specific characteristics of the DWT process and its operations.

2.2 General

2.2.1 What is a Digital Twin?

A Digital Twin is defined as a digital copy (virtual model) of a physical system. This model is continually fed with data and provides a glimpse into the past, present and future behaviour of the asset (Alzamora *et al*., 2021; Armstrong, 2020). The concept of a Digital Twin was first developed by Dr. Michael Grieves from Michigan University (USA) in 2003 (Grieves, 2018). Since then, many sectors have benefited from taking sensor data and linking them to Digital Twins such as aerospace, automotive, infrastructure, energy, medical, logistics, river management, flood control and manufacturing (Castro-Gama *et al*., 2020).

For technical systems, Digital Twins are used for mirroring physical systems so that decision makers can make better informed decisions to improve the system management and life-cycle (Alzamora, Conejos *et al*. 2021). Some examples of Digital Twin usages are (Alzamora *et al*., 2021; Armstrong, 2020; Castro-Gama *et al*., 2020; Pesantez *et al*., 2022):

- development of Master Plans to make long-term projections and evaluating new scenarios;
- estimate unprecedented scenarios and long-term impacts on a system;
- sustainable reengineering of systems aimed at reducing environmental, health and safety impacts;
- provide a better understanding of the system performance to optimize operations and maintenance;
- assist operators on decision making in real time by simulating issues in the Digital Twin prior.

Digital Twins should not be confused with models. Conventionally, models are usually offline static representations that do not allow for real-time data integration. Digital Twins are dynamic and constantly changing with the introduction of real-time data from the Physical Twin. Digital Twins can provide insight into the interactions between multiple data sources and physical assets unlike a traditional model. Therefore, a Digital Twin is able to show in real-time what is happening within the Physical Twin unlike a traditional model which only provides "what if" scenarios (Castro-Gama *et al*., 2020; Jones *et al*., 2020). However, the operationalising of a traditional model to perform "what if" scenarios, along with its integration with historical data, can in fact be considered a feature of a Digital Twin. Therefore, a Digital Twin can very much include a model within its functionality, but they also include the integration with multiple data and/or real-time data. Another distinction is the potential capability of Digital Twins to provide specific information back to the Physical Twin. This is typically in the lines of performing real-time control by steering the processes using real-time data and analysis.

2.2.2 Stages of Digital Twinning

According to the study of Csaba Ruzsa in 2021, there are four types (development stages) of a Digital Twin. As a Digital Twin progresses through the different developmental stage the level of interaction, control and real-time simulation increases.

The four developmental stages are:

- *Testing model:* where a virtual copy of the physical asset exists but no real-time data is fed into the models.
- *Surveillance model:* where real-time data from the physical asset is fed into the Digital Twin; however, the digital version only provides a supervisory role and does not interact with the physical asset.
- *Control Twin:* an interactive connection where the physical asset provides data to the Digital Twin and the Digital Twin has a controlling function within the physical asset.
- *Simulation Twin*: where data is provided by the physical asset and simulations of changing circumstances are projected within the Digital Twin.

2.2.3 Types of Digital Twins

There are a number of different varieties of Digital Twins. The type required depends on the sector and the expected applied usage of the Digital Twin (Castro-Gama *et al*., 2020; Jones *et al*., 2020; Liu *et al*., 2021).

- *Product Twin:* in a Product Twin the physical asset is something that has been manufactured, such as an automobile, aircraft turbine or a water pump.
- *Process Twin:* a Process Twin is where the Digital Twin emulates a process and monitors the behaviour of the system such as a water distribution network or manufacturing assembly.
- *Performance Twin:* a Performance Twin is a combination of a Product Twin and a Process Twin where the Digital Twin monitors the individual asset components and the overall system health. In some situations, this monitoring will allow the application of predictive maintenance.

2.2.4 Benefits of Digital Twins

The benefits of using Digital Twins have been acknowledged by both scholars and practitioners (Kockmann, 2019; Qi *et al*., 2021; Lim *et al*., 2020; Negri *et al*.,2020). The benefits include, but are not limited to, streamlining processes, downtime reduction, decreasing lead times (Perno *et al*., 2022). In the study of Lim *et al*. (2020), a reference framework for Digital Twin development was applied to a case study of a tower crane model. Here, the Digital Twin was able to demonstrate simulations which resulted in a reduction in operator workload, run risk simulations to assess damage to equipment and to provide data to be used in future projects.

2.2.5 Barriers to Digital Twins

When looking to use Digital Twins for a process system, such as water treatment, there is little literature on the implementation for Digital Twins of this nature. The research previously published is fragmented across different industries and topics. Due to the immaturity of research in combination with the intrinsic complexity of various production processes, it is difficult for companies to make decisions upon the appropriate approach for Digital Twin implementation (Kockmann, 2019; Perno *et al*., 2022).

According to the literature review of Perno, Hvam and Haug (2022), 79 different articles pertaining to Digital Twin development were analysed and a description of the barriers facing the implantation of Digital Twins is presented in Table 1. Overall, the most common barriers cited are those caused by issues in system integration, security and system performance.

Table 1 List and categorization of Digital Twin barriers (from Perno et al. 2022)

List and categorization of DT barriers.

Category	Barrier	Number of papers mentioning barrier
System integration issues	Lack of system integration	12
	Difficulty in ensuring interoperability	8
	Compatibility between new and legacy systems	4
	Difficulty of integrating legacy manufacturing systems with modern IoT/IIoT service ecosystems	2
Security issues	Security and privacy	17
	Difficulty in ensuring data transparency	2
	Difficulty in ensuring protection of intellectual property	2
	Need to share the DT among multiple application systems involving multiple stakeholders	1
Performance issues	Difficulty in ensuring a high level of performance in real-time communication	11
	Difficulty in ensuring efficient storage, processing, and analysis of large volumes of data	8
	Difficulty in ensuring a proficient interaction between the digital and physical assets	6
	Difficulty in ensuring reliability and robustness	6
	Difficulty in ensuring low latency communication, tracking, and reporting	4
	Difficulty in predicting complex systems	3
	Difficulty in ensuring the availability of relevant data for DT when needed	2
	Difficulty in ensuring a satisfactory level of accuracy	1
	Ability to track machine status and usage, also when no Internet connection is available	1
	Discovering and retrieving data from the IoT	1
	Difficulty in ensuring proper scalability	1
	Difficulty in ensuring flexibility and modularity	1
	High expected timeliness for DT development and implementation	1
	Risk of overfilling bandwidth	1
	Difficulty of finding a balance between enough (and relevant) information and overwhelming (irrelevant) information	1
Organizational issues	Bureaucracy, cultural inertia, and knowledge assessment	4
	Lack of specialists and expertize	3
	Difficulty in ensuring centralization, simplification, and standardization	2
	Difficulty in setting realistic expectations and trust	1
	Difficulty in identifying clear value propositions associated with DT solutions	1
	Lack of investments	1
	Difficulty in making suitable decisions and investments regarding the enabling technologies	1
	Isolated, fragmented, and stagnant data management	1
	Difficulty in combining product lifecycle management, manufacturing execution system, and operation management systems	1
Data quality issues	Uncertainties in the quality and reliability of data	3
	Data unavailability	2
	Difficulty in ensuring data validity	2
	Difficulty in ensuring data governance, ownership, and management	1
	Impossibility of directly measuring all data relevant for the DT	1
Environmental issues	Lack of methodologies and tools	5
	Low maturity of literature and practical industrial implementations of DTs	2
	Global advancements	1
	Difficulty in choosing the right software among a vast selection of open-source software	1
	Global connectivity	1
	Lack of high-fidelity models for simulation and virtual testing at multiple scales	1
	Multi-disciplinarity of environments for design and development	1
	Lack of education on the topic at universities	1
	High degree of heterogeneity in equipment manufacturers and their software solutions	1

2.2.6 Digital Twins in industry sectors

The first practical implementation of Digital Twins was completed in the aerospace industry (Negri *et al*., 2017), but rapidly spread to other industries including construction, healthcare, automotive, education, meteorology and building science (Perno *et al*., 2022). Digital Twins have also historically been used in the oil and gas industries. These industries have built upon traditional simulation techniques and added the applications for real-time simulation (Scheifele *et al*., 2019).

2.2.7 Social constraints of Digital Twins

For the implementation and utilisation of Digital Twins organisations need to be aware of the social applications, constraints and barriers. Digital Twins are powerful tools that can digitally reproduce the relationships and connections between the people, the process and the economics of the system. This can be seen in the 2020 study of Castro-Gama *et al.* where cellular phone usage data was employed to look at customer mobility within a water network distribution system area. This type of Digital Twin incorporated real time social data showing the connection between society, economy and the water distribution process (Castro-Gama *et al*., 2020).

Moreover, Digital Twins are powerful tools, but they require constant input from the system. In the early stages of implementation this means data validation is vital from experts within the company. The practical knowledge of system operations on a day-to-day basis is embodied in the employee. Additionally, for successful implementation the management of the organisation needs to be able to trust that the digital reproduction has an adequate level of accuracy – particularly when implementing more highly developed types of Digital Twins (i.e. Control or Simulation Twins).

2.2.8 Legal constraints of Digital Twins

Due to the immaturity of research surrounding Digital Twins, little has been published about the legal restrictions and environment in which Digital Twins need to operate. Recently, research has been published within the medical field around the use of personal information as a type of Digital Twin of a person that is created by a doctor (Teller, 2021).

For process systems, one of the principal advantages to be gained from implementation is the monitoring and analysis of data for regulatory compliance in a standardized fashion. The Digital Twin can be used to simulate impacts on the environment to estimate irregularities in the system as well and expansion of the system in the future. Standardization between similar processes further allows for the easier sharing of information between organisations (Castro-Gama *et al*., 2020; Perno *et al*., 2022; Teller, 2021).

Notwithstanding this, there are some major legal barriers and concerns that also must be addressed. One major fear surrounds cyber-security. Control and Simulations Twins continually feed information between the Twins. Therefore, monitoring data use rights to protect data privacy needs to be considered in addition to ensuring the robustness and integrity of the Digital Twin itself and to ensure that there is no opportunity for the physical twin's operation to be compromised by external actors (Perno *et al*., 2022; Teller, 2021).

2.3 Architecture

This part of the literature review is intended to examine the broader architectural context of Digital Twins as a virtual representation of a physical object or system that enables real-time monitoring, simulation and analysis of its performance. The Digital Twin is created by gathering and integrating data from various sources such as sensors, simulations and other data-producing systems. This data is then used to generate a dynamic, digital model of the physical object or system that can be analysed and optimized. The Digital Twin can be used in various industries such as manufacturing, transportation and healthcare to improve decision-making, reduce downtime and enhance customer experiences.

Industry 4.0 is an umbrella term frequently used to set the scene for the development of Digital Twins. This describes the "fourth industrial revolution" spurred by digitalization to incorporate communication and information technologies and refers to the integration of physical and digital technologies (such as IoT, robotics, artificial intelligence and cloud computing) into industrial processes and the manufacturing sector. The principal goal of Industry 4.0 is to create a smart and connected industrial environment that enables increased efficiency, flexibility and customization in outputs: resulting in enhanced productivity and improved customer experiences.

2.3.1 Nomenclature

A lack of standard terminologies and practices in the implementation of Digital Twin architectures are seen as a major challenge to the realization of the technology's full potential. Partly this is down to the literature on Digital Twin architectures being dominated by application- and/or technology-specific architectures that use different terminologies for components, hindering the identification of commonality in frameworks and making it difficult for practitioners in the field to find appropriate guidance.

To this end, Steindl *et al.* (2020) propose a generic architecture for implementing Digital Twins that conform with the information technology layers of the Reference Architecture Model for Industry 4.0. They report that the design and implementation of Digital Twins has attracted considerable attention in recent years, especially in the industrial energy systems domain, as it can facilitate flexible and optimized operation and help the industry transition to renewable energy sources. However, as in many fields, implementations of Digital Twins are often application-specific, lacking general architectural concepts and their structures and concepts vary. To address this issue, they propose a Generic Digital Twin Architecture which is aligned with the information technology layers of the Reference Architecture Model Industry 4.0 (RAMI4.0) to ensure a common naming and understanding of the proposed architectural structure and how it relates to the underlying concepts. Their Digital Twin is technology-independent and based on the 5D-Digital Twin concept, which has been evaluated based on a prototypical proof-of-concept implementation. Ontologies are used to build the foundation for the Shared Knowledge base of the Smart Data Service, facilitating interoperability which is demonstrated with an example of the generic architecture being applied to a Digital Twin representation of a Packed-Bed Thermal Energy Storage system.

The underlying Industry 4.0 Reference Architecture Model for their approach is a framework that defines the key components, their relationships and the flow of data and information between them in an Industry 4.0 environment. The reference architecture model provides a common language and understanding for organizations and stakeholders to build and integrate Industry 4.0 solutions and technologies which are useful in defining an appropriate nomenclature for Digital Twin-based applications.

The reference architecture model typically includes the following components:

- Physical devices and machines: These include sensors, actuators and other components that make up the physical assets in an Industry 4.0 environment.
- Edge devices: These are responsible for decentralized processing and filtering the data generated by physical devices.
- Network infrastructure: This includes the communication and data transport systems that enable the exchange of data between physical devices, edge devices and the cloud.
- Cloud infrastructure: This includes the centralized/distributed data storage, processing and analysis systems in the cloud.
- Applications: These are the software applications that enable data analysis, process control and decision-making in an Industry 4.0 environment.
- User interfaces: These provide a way for users to interact with Industry 4.0 systems and technologies.

Nwogu *et al.* (2022) also identify the lack of standardization in the nomenclature as an issue. To address this problem, they propose a requirement-driven, technology-agnostic Digital Twin architecture consisting of standard components traceable to the definitions, requirements and mandatory functionalities of Digital Twins captured in existing literature. The proposed architecture can be applied to various fields and use cases based on their respective needs and seeks the standardization of Digital Twin architectures by matching the components of the architecture to the core Digital Twin requirements. Their proposed architecture is affected by limitations such as the relationship between the Digital Twin requirements and their implementation within an existing information system and additional requirements that may arise from specific manufacturing or service systems.

2.3.2 Composition

A review paper by Adamenko, *et al.* (2020) investigates the methods used to design Digital Twins in the literature, noting that despite the widespread interest and emerging importance of Digital Twins in industry the technology has not yet fully established itself, partially due to the lack of clarity on how these should be built up. Digital Twins can offer benefits such as consistent documentation, better starting points for simulations and optimizations and the possibility of improving preventive maintenance. The authors propose two different approaches that can be followed when creating a Digital Twin, either data-based or system-based and conclude that a combination of both approaches is most beneficial.

This approach is mirrored by Meierhofer, *et al.* (2021) who describe a novel conceptual model for integrating Digital Twins in industrial service ecosystems to facilitate the use of Decision Support Systems. Their model uses a semantic ontology approach to interlink Digital Twins of equipment and processes in the ecosystem, enabling the development of a blueprint for implementors to create their own Digital Twin-based services. The hierarchical-modelling approach breaks down complex decision-making problems into sub-questions and individual Digital Twin components, systematically interlinking them using ontologies. The KARMA language is used to describe Digital Twins and generate ontology models automatically. The paper concludes by presenting a case study of a manufacturing SME to illustrate the implementation of the model, showing how different operational states in the ecosystem can be simulated as needed to support decision-making. This presents a theoretical concept for implementing Digital Twins on the level of service ecosystems and integrating Digital Twins based on a unified ontology, which can be used to create value for decision support.

A similar ontological approach to decomposing the constituents of a Digital Twin is proposed by Fujii, *et al.* (2022) who have proposed a new approach based on a Digital Twin model that incorporates real-time data collected by IoT sensors to improve machine learning methods. The Digital Twin household ontology model includes topological and behavioural aspects of accommodation, as well as metadata that can be used with energy consumption forecasts by other systems. Household metadata is modelled in an ontology to facilitate the integration of real-time monitoring and prediction information with new interfaces, such as personalized conversational agents and dashboards for improving personalized demand response suggestions and engaging consumers in the transition to renewable energy.

A predominant theme in Digital Twin architecture design is the concept of structuring the system into tiers. At its most basic this tiering is split into three levels, representing the Physical space, the Virtual space and a tier between them to exchange information and control messages.

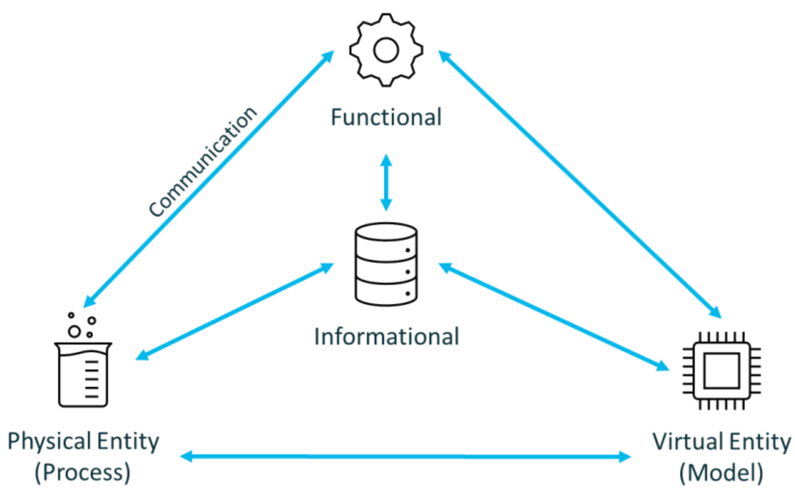


Figure 1 Functional elements of a Digital Twin architecture (after Steindl, et al. 2020)

The lack of a comprehensive architecture covering the necessary components of a Digital Twin to realize various use cases is further noted by Ashtari Talkhestani, *et al.* (2019). This paper proposes an architecture for a Digital Twin and an Intelligent Digital Twin and their required components to enable use cases such as plug and produce and predictive maintenance. This approach adopts a tiered architecture along the lines of that used by Steindl, *et al.* (2020) in Figure 1 above – although lacking the standardized nomenclature. The authors propose that a Digital Twin requires three main characteristics: synchronization with the real asset, active data acquisition from the real environment and simulation ability. Their proposed architecture for a Digital Twin, as part of a Cyber Physical Production System, includes several methods such as the Anchor-Point-Method, a method for heterogeneous data acquisition and data integration and an agent-based method for the development of a co-simulation between Digital Twins. The Anchor Point Method synchronizes multi-disciplinary models of a Digital Twin, while the cloud-based approach for data acquisition and data integration using semantic technologies acquires operational data to enable the use of machine-learning to detect anomalies and predict failures.

Chaux, *et al.* (2021) look at the need to achieve food security and increase production of agricultural systems while reducing resource usage. To do this they propose Digital Twin and Controlled Environment Agriculture (CEA) systems that can offer the potential to optimize productivity and improve food security. A Digital Twin architecture for CEA systems that utilizes simulation software to optimize climate control strategies and crop management is presented which demonstrates a specific application of the tiered-style architecture similar to Figure 2. The architecture was applied to a prototype greenhouse and validated through the assessment of communication latency. The proposed architecture can be used by companies to retrofit their CEA systems with Digital Twin functionality and by universities to investigate automation in agricultural laboratories - including the application of optimization through heuristics and algorithms.

Combining the concepts of the tiered architectures considered here permits us to formulate a general tiered architecture for a Decision-Support aware Digital Twin as seen in Figure 2. As can be seen, the architecture retains the discrete tiers for the Physical and Virtual entities along with those for the Informational and Functional components. It formalizes the Communication tier as a separate entity which represents the boundary between the IT (Information Technology) and OT (Operational Technology) constituents of the system. This, ordinarily, will include some form of demilitarized zone (DMZ) to keep the OT environment of the Physical Entity at arm's length from the data warehousing and processing of the Informational tier. In addition, the generalized architecture proposes a separate Presentation tier for applications facing the wider business including dashboards.

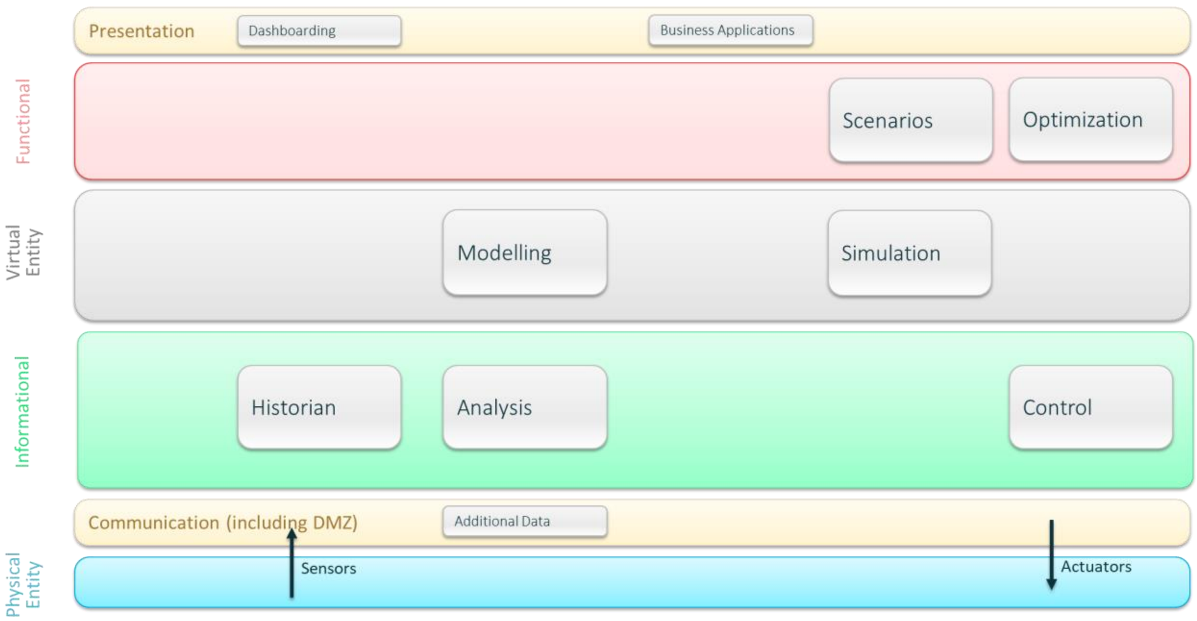


Figure 2 Generalized Architecture tiers for a Digital Twin facilitating Decision Support

2.4 Model requirements

An integral component for the implementation of a Digital Twin is the development of models. Models are alternative representations of a given a real-world system or installation. Through modelling, knowledge and information of a given task can be encoded, where the knowledge is acquired from prior experience and information gained from the collection of processed data of a given system (Therrien *et al*., 2020). Typically for water treatment systems, the representation of a system is achieved through mathematical equations which include the temporal and spatial relationships between different variables that explain the processes. The modelling of water treatment systems has over the years been considered beneficial for water professionals in aiding in the design of new plants and to optimise the operations and maintenance of existing plants - supported by predictions to estimate the future behaviour or state of a given treatment system (Therrien *et al*., 2020).

In practice, with respect to drinking water treatment (DWT), the development of models that are either incorporated into a Digital Twin or are themselves considered to be a Digital Twin, have been seldomly reported and this application can be considered to be still in its infancy. However, it must be considered that Digital Twins, while gaining popularity as a concept, have not widely adopted a standardised terminology. As a result, there have been many models or modelling studies conducted for various DWT processes which can be considered to be a form or a component of a Digital Twin, without explicitly mentioning the term. In the handful of studies reported that do mention Digital Twins, either a discussion has been undertaken at a conceptual level (e.g. Curl *et al*., 2019) or Digital Twins have been developed to investigate specific challenges such as anomaly detection and resilience to cyber-attacks (Patriarca *et al*., 2022; Wei *et al*., 2022). In the discussion of Curl *et al*. (2019), Digital Twins have been stated to be the “next big” technological advancement for water utilities since the introduction of Supervisory Control And Data Acquisition (SCADA) systems. They rationalise that Digital Twins essentially are a central repository for information and provide a basis to analyse the treatment facility operations and performance. In such a concept, two principal forms of Digital Twins were identified, one being termed as a ‘facility Digital Twin’ and the second being the ‘flight simulator’. The facility Digital Twin was explained as a means to conduct, for example, dynamic simulations of the hydraulics, water quality and controls of a DWT plant prior to the facility being built - thereby aiding in the designing and commissioning process. The flight simulator model was considered as a mechanism for operators to simulate failures or test optimisation strategies of their facilities to be considered for implementation or to be of assistance in operator training. However, a current gap that can be identified from such a discussion is whether such models are capable of providing simulations and predictions in (near) real-time. This leads to the question as to whether such models should be considered a Digital Twin to begin with.

A clear requirement to implement a Digital Twin for DWT is the availability of a model, be it for a specific asset, a facility Digital Twin, or a flight simulator. The model should be capable of accurately representing and mimicking the treatment processes when input with a certain initial state (i.e. controls, boundary conditions, environmental and operating conditions). The modelling of different DWT processes can be largely categorised into three approaches, based on the technology utilised, data availability and the complexity of the processes modelled (Matheri *et al*., 2022; Therrien *et al*., 2020):

- *mechanistic modelling;*
- *data-driven modelling;*
- *hybrid modelling.*

In sub-sections 2.4.1, 2.4.2 and 2.4.3 below, individual reviews of the different schools of modelling are discussed while restricted the subject to source water quality and some examples of key DWT processes that are of interest and importance to the Dutch and Flemish water sector such as coagulation/flocculation, softening and membrane processes (such as Reversed Osmosis and Ultra Filtration). In 2.4.4, the implications and decision-making processes that water companies can use in deciding which model type to consider for a given DWT process are highlighted.

2.4.1 Mechanistic Models for DWT

Mechanistic models for drinking water treatment are a mathematical representation of the underlying physical, chemical and biological processes that occur during drinking water treatment. Mechanistic models provide a framework or a system of equations, that are not data dependent but are based on implied knowledge or hypothesis about a specific system (Aliashrafi *et al*., 2021). Mechanistic models provide predictions of certain physical and/or chemical behaviour of the system under different conditions. Typically, the calibration of mechanistic models require the estimation of various model parameters, such as rate constants, hydraulic properties, etc. Predominantly, mechanistic models utilise mass-balance based systems of equations that predict the fate of certain water quality parameters, based on the input operating and environmental conditions. The mechanistic modelling of DWT processes is specific to the various treatment units that are seen within DWT installations. The coupling of various mechanistic models together can result in an integrated model, where the outputs of a specific unit, which could be the concentration of certain parameters and the flowrate of the output water would serve as the input to a subsequent unit. This has been highlighted as an example in Figure 3.

The modelling of individual treatment units is governed by the process equations that are applicable to the given system. For coagulation and flocculation, it is a three-stage process, consisting of rapid mixing, coagulation of the colloidal particles and flocculation. The key process in this treatment is destabilisation and hence that is a phenomenon that requires accuracy in a predictive model (Akinmolayan, 2017). Destabilisation is the process in which particles that are in a stable suspension state are aggravated or modified to increase the probability of their inclination to attract to one another. The mechanistic modelling of this process comes down to accurately representing the hydrodynamic forces between particles, particle size distribution and the coagulant dosing.

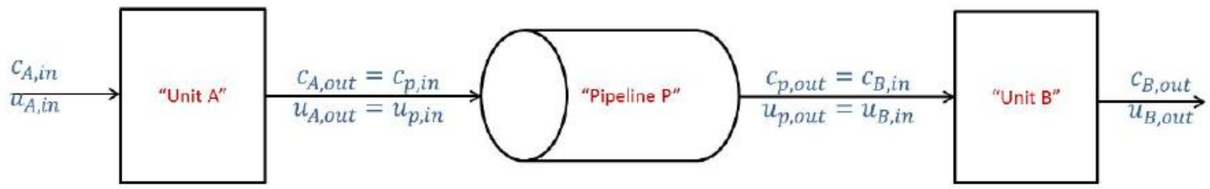


Figure 3 Individual mechanistic model connections to form an integrated model for DWT (after Akinmolayan, 2017)

These phenomena can be described based on physical chemistry-related equations that have been researched extensively. The particle size distribution and aggregation process can be modelled as a population balance model (Jeldres *et al*., 2018). A soft sensor for the coagulation-flocculation process is currently being developed as part of a BTO Bedrijfsonderzoek project called Softsensor flocculatie. However, the interacting chemical and physical processes in coagulation and flocculation make it a challenging process to accurately model. Turbidity, which is the primary indicator of process performance, is not only influenced by the content of suspended solids, but also the composition, particle size, shape and impurities in water, which makes it difficult to accurately predict (Li *et al*., 2021).

Another key process is softening, which involves the removal of calcium concentration ($CaCO_3$) through the dosing of chemicals such as sodium hydroxide (NaOH), which promotes the precipitation of the calcium carbonate due to the rise in pH. The modelling of the softening process involves the use of chemical and mass balance equations that can model the crystallisation of $CaCO_3$, based on operating conditions, chemical dosage and by solving the carbonic acid equilibrium set of equations (van Schagen *et al*., 2006). For membrane filtration, a wide variety of theoretical mechanistic models have been developed. Primarily, the models have been developed to predict and diagnose membrane fouling and can be valuable as they can assist in optimising the fouling removal through backwashes and other preventive methods. Furthermore, they can also help establish interactions and potential relationships between different filtration variables (AlSwaftah *et al*., 2021). For Reverse Osmosis (RO) systems, predictive models have been developed to simulate the development of membrane fouling over time. Such models use fouling potential, i.e., the increase in membrane resistance due to a unit volume of permeate passing through the membrane, as a concept for predictions (Chen *et al*., 2004). In this study a strong interaction between permeate flux and membrane fouling was seen. Duclos-Orsello *et al*. (2006) developed a model that accounted for three classical fouling stages – 1) pore constriction; 2) pore blockage; 3) cake formation. Other versions and theoretical formulations of membrane fouling have also been developed and further for other membrane-based processes (Chang *et al*. 2011 ; Mondal and De, 2010).

Mechanistic models can be challenging to implement as they can rely on design parameters that are difficult to measure (Aliashrafi *et al*., 2021). Furthermore, many mechanistic models for DWT are typically theorical models and mathematical formulations that have been validated under simple conditions and laboratory-based measurements. The complexity of feed water to full-scale drinking water treatment plants prevents the operationalising of such theoretical models that struggle to make good predictions (Li *et al*., 2021). Therefore, mechanistic models have various disadvantages and can potentially be difficult to utilise in Digital Twins due to the need for solving complex equations and a lack of accuracy. However, these white-box models can still be highly beneficial and useful to perform scenario analysis and aid in decision-making that requires a high level of understanding of the processes.

2.4.2 Data-driven Models for DWT

Data-driven models map inputs onto specific outputs without considering the real processes underlying the relationship. As a result, with no prior knowledge being utilised, such models are fully reliant on the data being used for their development (Therrien *et al*., 2020). Like many other domains such as computer vision and health care, Artificial Intelligence (AI) and the application of data science have demonstrated their critical value to its application in the drinking water treatment sector (Aliashrafi *et al*., 2021). This has been made possible with high resolution online sensor data of key process parameters now being available. Such data-driven models have been considered to be effective in embedding complex non-linear relationships that can be found solely within the data without needing to explicitly define the relevant features, relationships or variables, for the prediction of certain outputs (Aliashrafi *et al*., 2021; Li *et al*., 2021; Therrien *et al*., 2020).

For the data-driven modelling of DWT processes, supervised and unsupervised learning models have been utilised. In supervised learning, data-driven models conduct predictions using water quality data. Such predictive models utilise historical data and learn the underlying relationship to make accurate predictions for a given label or target value (Aliashrafi *et al*., 2021). With supervised predictive models, the tasks performed could be of the type of regression or classification. Regression models provide predictions of continuous numerical values based on a numerical input. Classification models provide discrete predictions, by assigning a category to each sample considered (Aliashrafi *et al*., 2021). An example of a classification task using water quality data could be the use of different water quality measures for the prediction of a water quality index (Abba *et al*., 2020). For supervised learning models, be it for regression or classification, labelled data is required. However, generating large amounts of high-quality labelled datasets can be costly and challenging (Aliashrafi *et al*., 2021). As a result, unsupervised learning modelling methods, such as clustering, are also considered. Clustering algorithms allow one to identify patterns or group (cluster) subsets of samples that tend to have similar behaviour (Aliashrafi *et al*., 2021). For example, by inputting water quality data, techniques such as K-means clustering can be used to identify and cluster different states of a water treatment process, thereby allowing one to evaluate the reaction of the treatment process to different water quality conditions (Juntunen *et al*., 2013). Additionally, data-driven models are widely used for the purpose of dimensionality reduction, which transforms high-dimensional inputs to a lower-dimensional representation. Common methods used include Principal Component Analysis (PCA) and Self-Organised Maps (SOM). Such techniques can also help improve the predictive performance of data-driven models such as artificial neural networks (ANNs) by reducing the number of inputs (Aliashrafi *et al*., 2021).

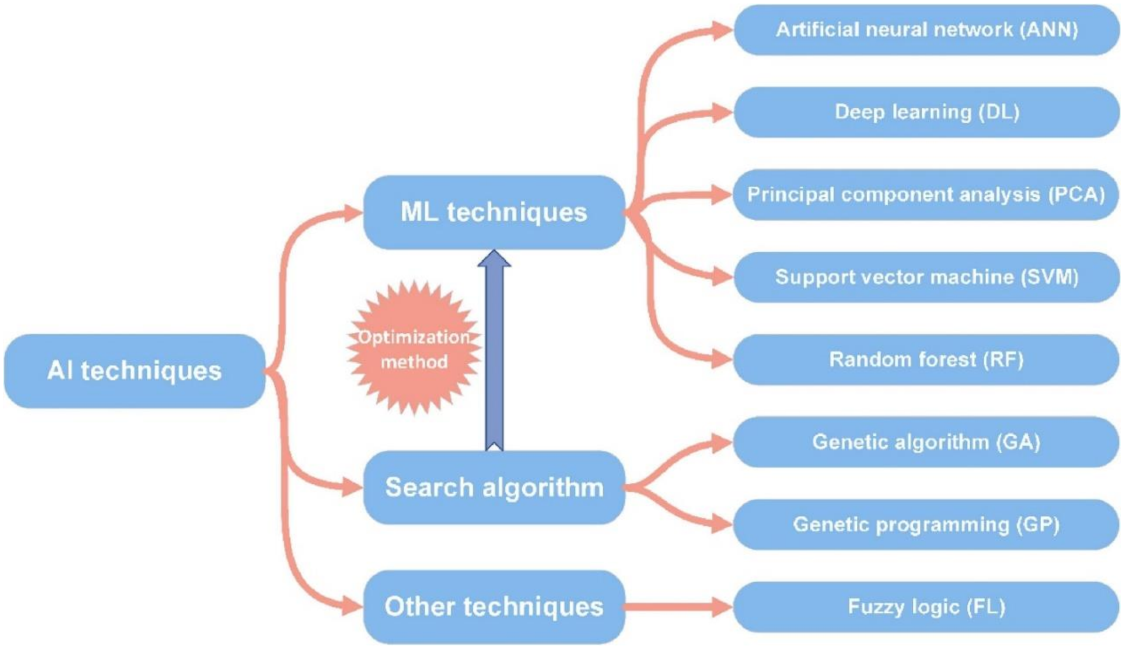


Figure 4 AI techniques utilised for the modelling of drinking water treatment (from Li, et al. 2021)

In Figure 4, a summary of the AI techniques used in the modelling of DWT processes is illustrated (Li *et al*., 2021). As can be seen, different Machine Learning (ML) techniques such as ANNs, Deep learning (DL) models, Support Vector machines (SVMs) and Random Forest (RF) models are commonly used, with ANNs being the most utilised. The choice of techniques published have been reported based on the stage of the DWT that was modelled.

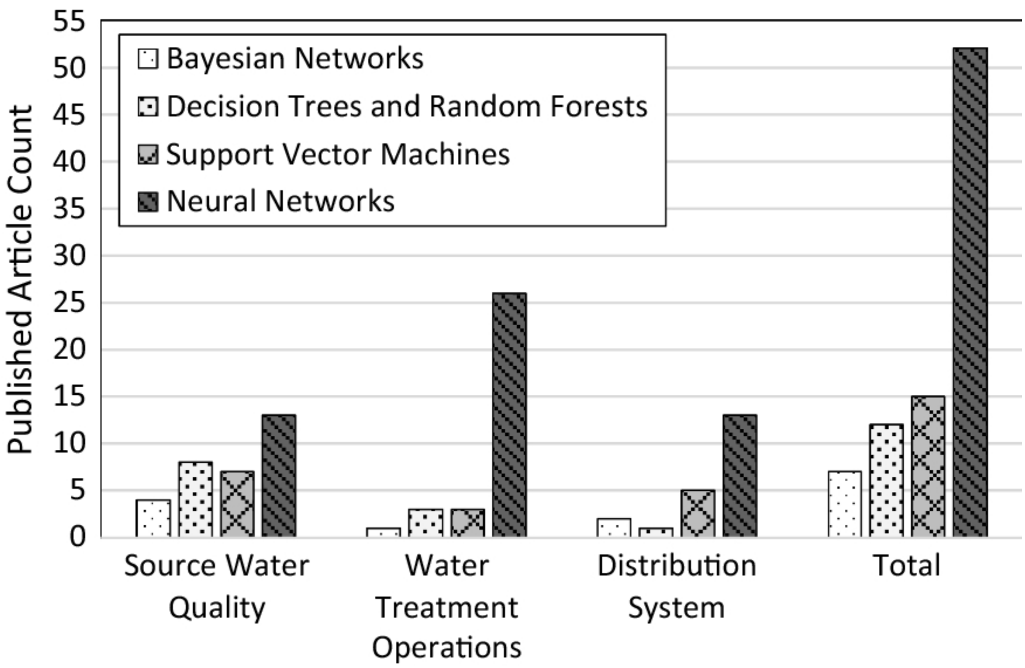


Figure 5 Articles that published ML techniques used for modelling water quality for different stages in DWT (after Aliasharfi, et al. 2021)

In Figure 5, as reported by Aliashrafi *et al*. (2021), neural networks are most common technique used for predicting water quality for various stages in DWT, followed by SVMs and decision tree-based models such as RFs. Based on the data-driven modelling of DWT processes reported in the literature coupled with processes that are of importance to the Dutch and Flemish water sector, some examples of using the aforementioned techniques are highlighted below.

For source water, data-driven modelling is primarily used for predicting the water quality, from a physical (e.g. odour), chemical (toxic contaminants) and microbial (such as algae blooms) perspective. Interestingly, AI technologies have an advantage in being able to predict complex water quality parameters using water quality indicators that are easily measured (Li *et al*., 2021). An example of this is the prediction of Arsenic in groundwater with principal component ANN (PC-ANN) using water quality indicators such as pH and electrical conductivity (Cho *et al*., 2011). In their investigation, Chen *et al*. (2020) demonstrated that the use of big data significantly enhanced the performance of their water quality prediction models, which included various machine learning approaches. Their study involved training models, such as Random Forest (RF) and Deep Cascade Forest (DCF), using parameters like pH, dissolved oxygen (DO) and NH3-N to predict compliance with governmental regulations regarding water quality standards. Interestingly, as they expanded their datasets, they were able to streamline the model inputs by focusing on the most critical parameters and simplifying the model structures. Consequently, the availability of a substantial dataset is a crucial consideration when exploring different machine learning techniques for water quality predictions.

For coagulation and flocculation processes, AI techniques are primarily used to predict the turbidity of the effluent or the amount of coagulant dosing required to achieve a given effluent quality (Li *et al*., 2021). To this end, the use ANNs have been widely used. For example, Kennedy *et al*. (2015) trained and tested four variations of ANNs, such as a Multi-layer Perceptron (MLP), Radial Basis Function (RBF) network and a Generic Regression Neural Network (GRNN), to predict the turbidity and dissolved organic matter (DOM) removal due to coagulation in a full-scale WTP. The input to the ANN models included the alum and other chemical dosing concentrations, turbidity, alkalinity, pH, hardness and temperature or the raw water and the settled turbidity levels itself. Similarly, Kim & Parnichkun (2017) trained an MLP, adaptive neuro fuzzy inference system (ANFIS) and a GRNN to determine the coagulant dosing using 8,760 historical datasets with hourly resolution from a full-scale WTP. The MLP and ANFIS models met the proposed validation conditions, with the MLP performing well during high turbidity zones over 20 NTU and the ANFIS providing consistent and better results at lower turbidity zones where there is higher disorder of coagulant dosage data. The GRNN model was concluded to fail in making accurate predictions during validation. Griffiths & Andrews (2011) developed two ANN models, one to predict the settled water turbidity and the second to predict the optimal alum dosage for a full-scale coagulation installation in a WTP. In both models, physical properties and quality of the raw water were used as input (such a turbidity, pH, conductivity and temperature) as well as the chemical dosing levels. For both models, average R^2 values were achieved, ranging between 0.63 to 0.79 and 0.78 to 0.89, respectively. In contrast, better results were also achieved using ANNs by Maier *et al*. (2004), where multiple ANN models were developed to predict multiple outputs that are controlled by the addition of alum, such as the treated water turbidity, colour and ultraviolet absorbance at a wavelength of 254 nm.

For membrane filtration processes, data-driven models have primarily been employed to analyse and predict membrane fouling behaviour, as well as to assist in membrane preparation and cost optimization. Many studies have been conducted to predict membrane fouling. For instance, Piron *et al*. (1997) utilized artificial neural networks (ANNs) and a semi-physical model that combined prior knowledge with a neural network to compute unknown parameters. Predictions achieved from the semi-physical model were more accurate, which was hypothesized to be due to the inclusion of mass balance equations. Additionally, the prediction of transmembrane pressure (TMP) and backwash efficiency using various data-driven techniques has been attempted. Delgrange *et al*. (1998) developed an ANN to predict TMP for an ultrafiltration pilot using input data on operating conditions and water quality parameters. This pilot produced drinking water by ultrafiltration (UF) of natural water. The neural network model accurately fit experimental data. In a study by Delgrange-Vincent *et al*. (2000), a back-propagation ANN (BPANN) model successfully predicted TMP for a UF pilot for both short-term and long-term processes. The study found that input parameters such as permeate flow rate, filtration time, turbidity, dissolved oxygen (DO), pH, ultraviolet (UV) intensity, backwash pressure and chlorine concentrations were the most significant. The model's good performance over the long term suggests that such models can be suitable for control, even in cases of varying water quality changes. Another interesting case was the study conducted by Shetty *et al*. (2003), where ANNs achieved high-accuracy predictions for permeate water quality during municipal nanofiltration for various source waters, membrane types and operating conditions, demonstrating their generalization capabilities. More recently, Corbatón-Báguena *et al*. (2016) investigated the importance of various data preprocessing steps to improve the fit of ANNs in predicting permeate flux decay in UFs. They also compared the data-driven model to Hermia pore-blocking models and concluded that the performance of both models was comparable. In the field of deep learning, convolutional neural networks (CNNs) were used to predict the increase in membrane fouling in nanofiltration and reverse osmosis (Park *et al*., 2019). The model was trained for image recognition using high-resolution dirt layer images and it provided promising results.

The use of artificial neural networks (ANNs) has found significant application in drinking water treatment (DWT) processes. This popularity can be attributed to their relatively easy training and their ability to rapidly capture the complex, non-linear relationships between input parameters related to water quality and operational factors, ultimately leading to target variable predictions. Additionally, neural networks excel in forecasting tasks, making them valuable for predictive decision support and control. In this regard, the use of deep learning models namely, recurrent neural network (RNN) type of models such as gated recurrent unit (GRUs) and long shot-term memory (LSTM) has shown great promise in the water sector as well. These models can be instrumental for making informed decisions and implementing predictive control strategies. However, a caveat of such models is the vast amount of data required. This explains the limited number of studies utilising such models for modelling DWT processes, but the field is rapidly evolving. Another caveat that such models face is overfitting to the training data. These models can then face challenges in performing sufficiently under conditions are not prevalent in the training dataset. This can be a crucial issue in rapidly changing environments. Furthermore, decision tree-based models such as RF and XGBoost offer valuable insights through their inherent ability to assess feature importance. They help in understanding the relevance of input parameters when making predictions, making them more interpretable compared to neural networks. They serve as robust tools for gaining a deeper understanding of complex processes by converting raw data into actionable process information. In summary, there remains considerable potential in leveraging data-driven methods for modelling DWT processes, particularly given the continuous growth in data collection. These models play a pivotal role in the development and application of Digital Twins for water treatment systems.

2.4.3 Hybrid Models for Water Treatment

Mechanistic models are usually mathematical in nature, based on physical or chemical laws that govern the intricate relationships between input and output variables. While these models leverage extensive domain knowledge accumulated over years, they remain approximations of reality. This approximation is due to the inherent limitations of incomplete knowledge and the inherent complexity of certain processes, leading to the need to make certain assumptions which introduces a degree of bias (Jia *et al*., 2022). Mechanistic models can also be empirical in nature, where relations between variables are described based on elementary physical understanding of the process. This results in less complex mathematical formulations. In either case, these models frequently incorporate multiple fine-tuning parameters, which could be the physically meaningful parameters in the case of purely mechanistic models or ad-hoc parameters in empirical models. Calibrating these parameters often entails an extensive search through a vast parameter space to pinpoint the combination that yields the best-performing model on the training data. However, this can lead to high computational costs during calibration and also when making predictions on unseen data, which hinders its usage in real-time and dynamic system modelling, typically required for control purposes.

On the other hand, the use of data-driven or ML models has gained rapid momentum, owing to their proven success in various commercial applications including computer vision, natural language processing. ML models are widely being considered a potential alternative to mechanistic models, particularly in the scientific community. Therefore, a substantial amount of research and innovative full-scale applications have focussed on utilising data-driven models to address specifical challenges related to DWT processes, particularly when there is sufficient data available. However, such models do possess certain limitations and caveats as well. The “black-box” nature of data-driven models results in them lacking transparency and making it difficult to understand the inner workings of the model responsible for generating predictions. Furthermore, DWT processes are often characterised by data sparsity.

Therefore, ML models trained on DWT process data can underperform, particularly when used to solve a supervised learning problem, which require a great deal of labelled data. Furthermore, the effectiveness of ML models is greatly dependent on its capabilities of learning complex patterns in data. This is simply conducted by identifying statistical relationships between the input and the target variables of interest, without regarding any form of physical and chemical laws. Such a training process quite often leads data-driven models to overfit on the training data provided. The identification of ML models that can successfully generalize across different scenarios remains an ongoing challenge.

To address the shortcomings that are prevalent in both schools of modelling, hybrid modelling, that combine the mechanistic or first-principles models with data-driven models, is a promising concept being considered at a scientific and commercial level. A common and sufficiently investigated approach is through residual AI modelling, where the data-driven models are trained to predict the errors made by the mechanistic model when comparing its predictions with observed data (Hvala & Kocijan , 2020; Keskitalo & Leiviskä, 2014; Wan *et al*., 2018). However, a key limitation of such an approach is still the inability of the data-driven models to make predictions that are consistent with the physical/chemical laws governing the process (Jia *et al*., 2022). Therefore, a multitude of other methodologies are now being considered in this field. Aspen Tech, in their development of such technology, has made a convenient demarcation of the types of hybrid models possible. These are, AI-driven hybrid models, reduced order hybrid models and first principles-driven hybrid models (Aspen Tech, 2020).

AI-driven hybrid models involves using ML models trained on observed data from full-scale plants or experiments, while incorporating the first principles and domain knowledge to achieve a more accurate model. This concept can involve incorporating a constraint layer as part of the model architecture which represents certain elements of the physics (Beucler *et al*., 2019) or the inclusion of a physical or chemical based loss parameter in the loss function that is being optimised during the model training process (Donnelly *et al*., 2023). In the reduced order hybrid modelling approach, an empirical model based on ML uses data provided from simulation runs of mechanistic model for training. This is primarily done to develop a surrogate model that can compute predictions more quickly than the mechanistic models, while retaining information on the physics. Finally, the first principles-driven hybrid modelling is an interesting approach where ML is integrated within an existing mechanistic model to increase the model's accuracy and predictability. This can be done by introducing physics-based equations within the training procedure of the ML models, thereby enforcing the data-driven models to also learn the governing principles. An example of this method has been showcased by Jia *et al*. (2022), where equations based on energy conservation and mass conservation were incorporated to the training of an LSTM-based AI model. The framework was termed Physics-Guided Recurrent Neural Network (PGRNN). Another interesting method is to use the technique known as Differentiable Parameter Learning (dPL), where the parameterisation of a mechanistic model can be dynamically identified by using a neural network, which then feeds the mechanistic model. This eradicates a limitation, mechanistic models utilising static values for the fine-tuned parameters, which are now dynamically provided based on the system conditions as represented in the input data provided.

In summary, hybrid modelling can offer numerous advantages. It can enhance the accuracy and interpretability of data-driven models, especially when dealing with sparse data. Moreover, it can reduce computational time, making mechanistic models suitable for real-time control. Additionally, hybrid modelling can ensure that data-driven models adhere to physical and chemical laws, improving their generalizability. It is anticipated that hybrid modelling holds great promise for modelling drinking water treatment (DWT) processes, especially in addressing the complexity and dynamics inherent in various treatment processes like coagulation and membrane filtration fouling.

2.4.4 Implications of Model Development for Digital Twin Implementation

As previously discussed, the modelling of drinking water treatment (DWT) processes can be approached using mechanistic models, data-driven models, or a combination of both through hybrid models. The decision regarding the type of modelling to employ for a specific DWT process can be complex, as it depends on various factors and the ambitions of the water company with respect to the model usage. In Figure 6, a decision tree has been illustrated to inform the decision-making process to choose an appropriate model type to develop for a given DWT process. It is anticipated that the fundamental starting point relies on how much high quality data is available. Limited data, particularly only from laboratory measurements should directly lead to a choice of using mechanistic models, as such models can provide adequate results during low data availability and data-driven models cannot be considered.

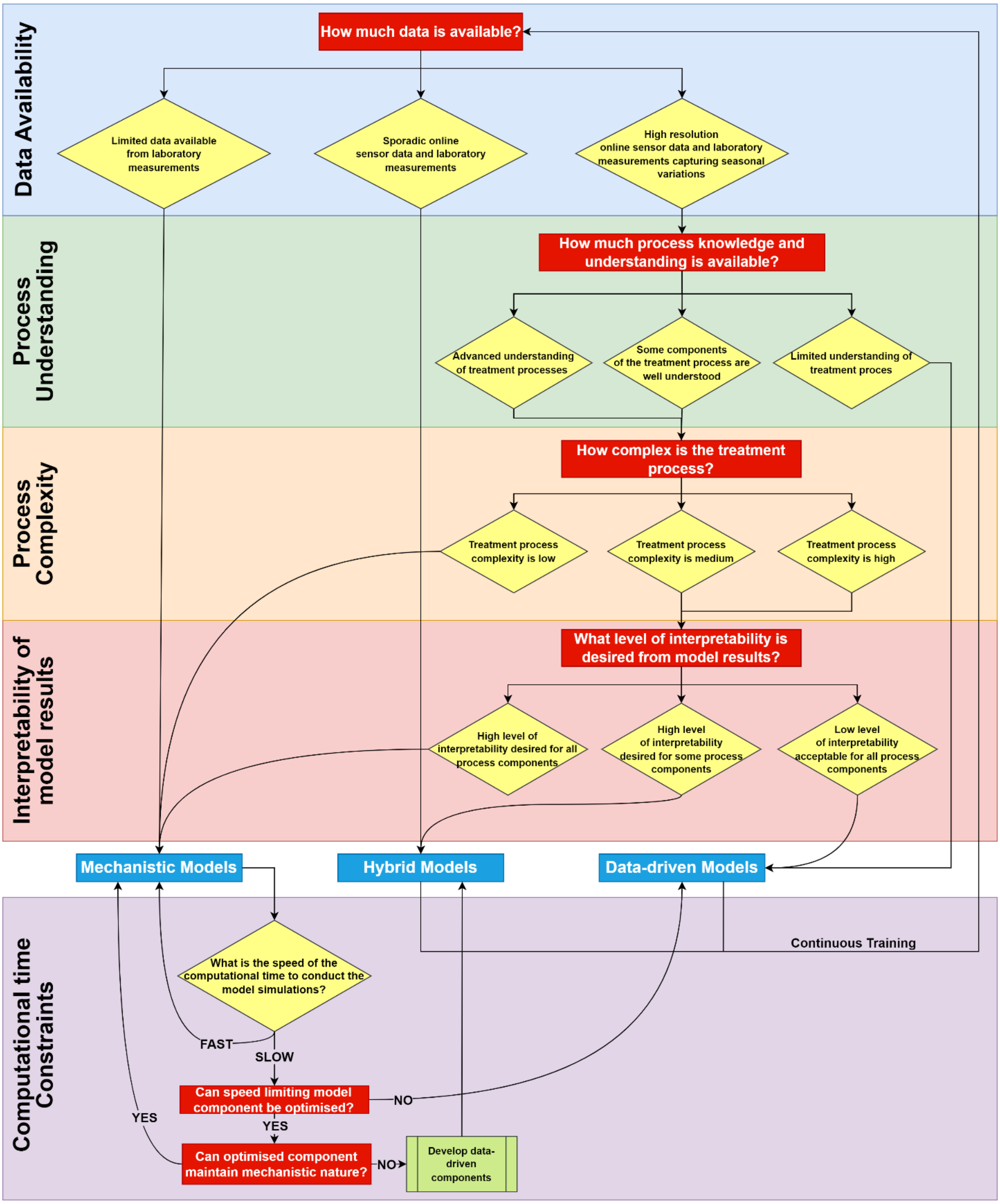


Figure 6 Decision tree considering various factors that influence the choice of modelling type for a given DWT process

In the event of sparse data being available, it is envisioned that hybrid modelling approaches should be considered. Hybrid models provide a promising direction where adding the process-based knowledge to the data-driven modelling could lead to achieving better performing models even when sparse data is utilised. For cases of high-resolution data capturing seasonal variations being available, the amount of process knowledge and understanding available on the given DWT process, will act as the next 'gate'. In the case that limited understanding of the treatment process is available, it is recommended that a data-driven modelling approach is chosen, considering ML models such as neural networks have been seen to work well in identifying complex non-linear relationships between variables, even when those processes are not fully known. For the cases of advanced understanding of treatment processes or some components of the treatment process is understood, the process complexity must be considered. In the event that the treatment process is not complex, mechanistic models can be incorporated. For medium to high process complexity, the ambitions of the water company with respect to the level of interpretability from the model results are desired will be considered. When high level of interpretability is desired, then mechanistic models must be considered. This can be the case when the model would be used for what-if? scenarios, capacity building and training, informing operational decisions etc. When low interpretability is desired, data-driven modelling can be used. This can be the case when data-driven models are trained on historical operational data and deployed as a control twin, providing setpoints for control variables. In this regard, the most important aspect is having a highly accurate model. For cases of high interpretability desired for some process components, it is recommended to consider the hybrid modelling route, where mechanistic models could be considered for those specific processes. Finally, the amount of computational time a model simulation takes can influence the final decision, particularly if the model is intended to be used for control purposes.

Additionally, the decision-tree for model selection follows an iterative and continuous approach. This was largely incorporated considering the rapidly growing volumes of data available and advancement in data-driven technologies. For data-driven and hybrid models, continuous training is recommended to keep the models up-to-date with the latest data. This will ensure that the models are (re)trained on process conditions that were not previously available in the former training set. This intervention promotes model generalization, reduces overfitting challenges, and maximizes data utilization. Moreover, recent data availability can influence updates to model selection decisions. Increased data availability could potentially increase the level of understanding of a given process and so on, as one traverses through the decision-tree. Furthermore, enhanced data availability may prompt a reconsideration of advanced modelling techniques previously dismissed due to data scarcity.

3 Roadmap Design

3.1 Interviews with Leading Organization

Within the scope of this project, a series of comprehensive interviews were conducted with two prominent water companies in the Netherlands, to explore their experiences, strategies and challenges in the development and deployment of Digital Twins (Digital Twins). The primary aim of these interviews was to accumulate valuable insights for informing the development of the project's roadmap. These insights serve as a foundation for both the project team and the broader drinking water sector.

The interviews encompassed seven key topics:

1. Definition of a Digital Twin
2. IT Infrastructure/Architecture
3. Data Collection and Sensor Deployment
4. Models Trained/Calibrated within the Digital Twin
5. Dashboarding and Advanced Visualization
6. Perception of the Digital Twin within the Organization
7. Legality and Regulations

It's important to note that not all questions of the survey were covered during the interviews, owing to variations in knowledge and the extent of Digital Twin implementation across organizations. The full list of questions is available in Appendix I.

3.1.1 Key Findings and Insights

1. Definition of a Digital Twin

Two water utilities view Digital Twins as a transformative concept and define them as representations of real installations and processes, emphasizing real-time data input and predictive capabilities.

2. Organizational Drivers

Drivers for Digital Twin implementation include the impending retirement of the workforce, technology advancements and environmental considerations. The need to preserve institutional knowledge before retirements, improving operational efficiency through digitalization and meeting sustainability goals were identified as primary drivers.

3. Organizational Changes

Two water utilities have undergone significant organizational changes to adopt digitalization strategies and support Digital Twin implementation. Effective strategies involve cross-departmental collaboration and seeking input from end-users.

4. Barriers to Digital Twin Deployment

Barriers to Digital Twin deployment include a lack of Digital Twin knowledge, data security concerns, IT limitations and resistance to change within the workforce. Recruitment of a specialized team for Digital Twin deployment and maintenance was suggested as a solution to IT limitations.

5. IT Infrastructure/Architecture

No responses were provided in this category, highlighting the importance of the ongoing project's assessment of IT infrastructure and architecture.

6. Data Collection and Sensor Deployment
One water utility highlighted their sensor data collection methods, emphasizing standardized sensors and potential future additions, such as energy consumption meters.

7. Models Trained/Calibrated with the Digital Twin
One water utility discussed the integration of multiple models, covering physical, chemical and data-driven models. Integration of models is crucial for comprehensive results and the exploration of AI models for complex processes is underway.

8. Dashboarding and Advanced Visualization
No responses were provided in this category, emphasizing the need for further evaluation.

9. Perception of the Digital Twin within the Organization
One water utility underscored the importance of trust-building and communication for Digital Twin acceptance within the organization. Training and selecting the right operational processes were identified as essential for fostering trust and achieving positive results.

10. Legality and Regulations
No legal issues were reported for Digital Twin implementation in the Netherlands.

11. Digital Twin Perception
Both organizations view Digital Twin implementation positively. While one water utility predominantly uses static Digital Twins, newer employees are more inclined to embrace them. The other water utility, regards Digital Twins as the only way to achieve their goals, with the control agent showing promise in accurate control actions.

Notably, the preliminary results suggest that Digital Twins offer significant benefits, particularly in compliance reporting, enabling constraint quantification and emissions calculations.

These interviews have provided crucial insights into the potential, challenges and best practices for Digital Twin implementation within the water sector. The knowledge gained is integral to shaping the project's roadmap and fostering broader adoption of Digital Twins within the drinking water sector.

3.1.2 Interview with external Digital Twin Expert

Background
The interviewee, an IT professional with expertise in computer science, is actively involved in translating Digital Twin concepts into practical strategies for businesses. Their role primarily revolves around IT-related tasks, serving as a bridge between management, developers, domain experts and modelers. Their responsibilities include explaining the necessary steps for Digital Twin implementation, coordinating various teams and facilitating communication between different organizational units. The interviewee maintains a practical perspective that prioritizes the successful execution of Digital Twins within a business context.

The interviewee has prior experience in the water sector, having worked with stakeholders in this field. Although they had interactions with water-related projects, it was observed that the industry may not have been fully prepared for Digital Twin adoption at that time.

Organizational Issues

The interviewee emphasized that their perspective primarily focuses on the actions required once a Digital Twin model is established within an organization. They are knowledgeable about the organizational challenges associated with Digital Twin implementation and referenced a 7-step model (Figure 7) designed to guide companies through the process. This model assesses the digital readiness of companies and provides guidance for a successful Digital Twin integration.

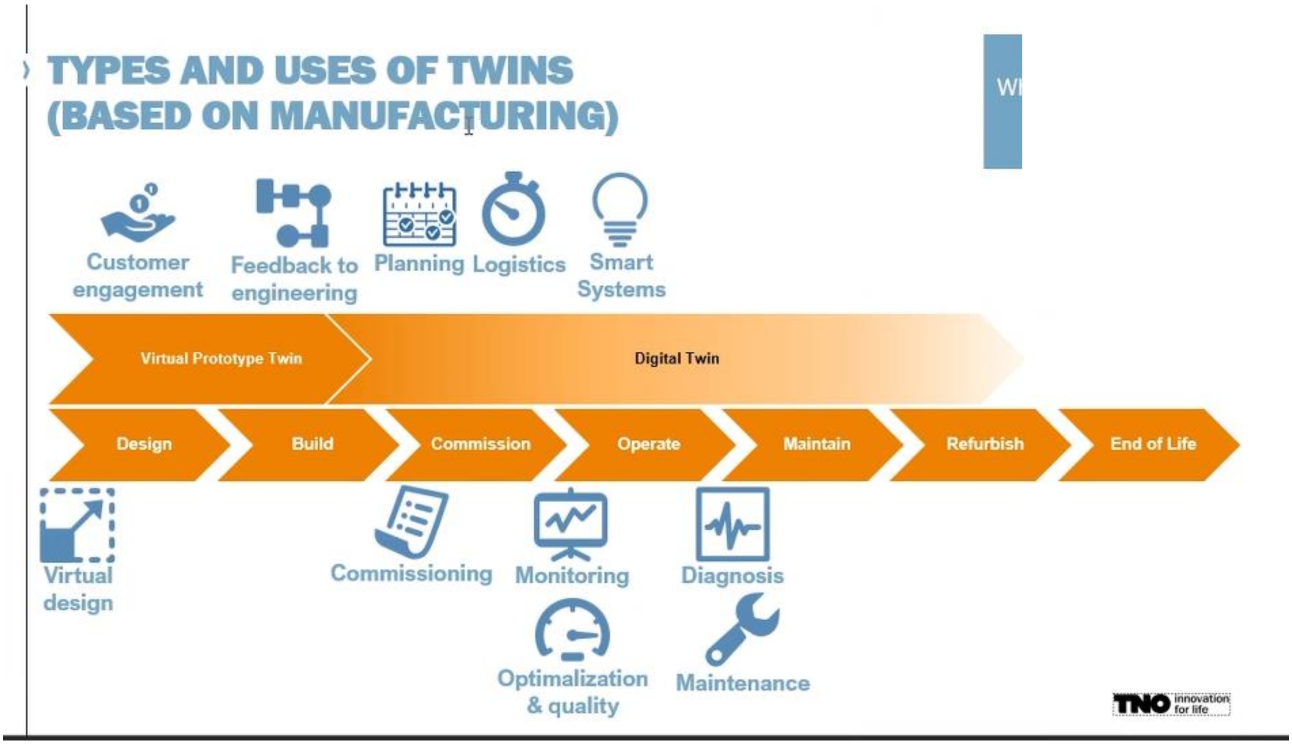


Figure 7 Digital Twin types and usages (Credit: TNO)

The interviewee also highlighted the challenges of creating a generic roadmap for all water companies. While theoretically possible, it often encounters management-related obstacles. They proposed that the development of such a roadmap should address three fundamental questions:

- the purpose and objectives of Digital Twins;
- the current digital maturity of water companies;
- identification and resolution of non-technical issues.

Non-technical issues were noted to be challenging to generalize and are not confined to the water sector. They represent a significant portion (3/4) of the digitization process.

Round Table Discussion/Community of Practice (CoP)
The interviewee discussed the potential value of organizing round table discussions or Communities of Practice. These forums could help bridge the gap between different domains and sectors. An existing cross-sector round table was recognized as beneficial for fostering cross-sector learning and addressing problems not necessarily tied to a specific domain.

Non-technical challenges were noted to be prevalent, comprising a significant proportion of the digitalization process. The interviewee emphasized the importance of involving management in these discussions and suggested that engaging management can be facilitated by providing answers to specific questions, possibly through a series of workshops.

Creating a Business Case
A crucial aspect emphasized by the interviewee was the need to establish a clear "business case" for Digital Twin implementation. They suggested that quantifying the problem to be addressed, such as understaffing or excessive chemical dosing, is essential. The ability to collect and analyse data efficiently plays a vital role in assessing the business case. It was pointed out that understanding the problems in the domain processes should take precedence before considering Digital Twins as solutions.

Contextual Data Management
The interviewee discussed the relevance of data management in the context of Digital Twin creation. Depending on the Digital Twin's requirements, data integration, particularly for business context data, may be necessary. For many treatment steps, where data integration is minimal, data management may not be a prerequisite. Most of the groundwork for operationalizing a Digital Twin is already completed in such cases. Integration of the model with sensor data is often the primary focus.

The interviewee expressed their willingness to contribute to workshops, provided they align with their schedule. They emphasized the need for engaging and valuable workshops, as the Digital Twin in question is more of a system-wide nature than a product-specific Digital Twin. They view this as an opportunity to provide practical examples and share the message with the wider community, which currently lacks robust Digital Twin implementation examples.

Stakeholders Within Companies
It was highlighted that discussions within companies should involve a range of stakeholders, including process operators, business developers, IT professionals and data specialists.

The interviewee expressed their interest in participating and supporting the development of the initial roadmap, leveraging their experience and insights.

3.2 Roadmapping Workshop

After completing the interview, a better understanding of best practices, barriers and risks which can occur in the development of Digital Twins was obtained. However, the ways to make a generic roadmap that works for all water companies remained a prominent question. Therefore, a workshop was held at KWR with the steering committee of this project to better understand the questions of: What would the digital twins be used for? How digital are the water companies to begin with? And what are the non-technical issues that are needing to be solved in the digitalisation process? The answers from the steering committee were used to inform the assessment of what the water companies need in order to develop digital twins for water treatment.

3.3 Roadmap

In this era characterized by rapid technological advancements and the ever-increasing complexities of water treatment, the concept of a Digital Twin has emerged as a transformative method for water utilities to understand, monitor and optimize their systems and processes. A Digital Twin is more than a digital replica; it is a dynamic, data-driven embodiment of the physical world, capable of offering real-time insights, predictive analytics and enhanced decision-making capabilities.

As organisations embark on their digitalisation journey, the need for a comprehensive and structured development roadmap has become apparent. The roadmap developed through this study not only informs the creation of Digital Twins but also ensures alignment with organizational objectives, modelling accuracy, robust data management and stringent security measures.

This Digital Twin Development roadmap consists of four distinct, yet interrelated paths, each vital to the success of the process of digitalisation.

1. ***Organisation:*** The Organisational path encompasses the selection of core teams, the identification of stakeholders and the establishment of clear policies and procedures. This path is crucial for ensuring that the right people are involved, the project stays on track, brings value and ensures ethical and regulatory considerations are met.

2. ***Model development:*** Within the Model Development path, the focus shifts to the creation and refinement of the Digital Twin's mathematical models, algorithms and simulations. With appropriate development the Digital Twin will be able to mirror the physical system it represents. Model development is the foundation for scenarios, predictions and automation.

3. ***Data management & architecture:*** Clean, accurate and validated data is the heart of a Digital Twin. This path involves the integration of diverse data sources real-time data collection and the design of a scalable architecture to store, organize and retrieve data. A well-structured data management system ensures the accuracy and relevance of the Digital Twin.

4. ***Security:*** The Security path is paramount as there are constant concerns of emerging cyber threats and the risk of data leaks. Implementing robust security measures to safeguard both the Digital Twin and the associated data is essential. The policy, encryption, transport and secure isolated communication are integral aspects of this path.

Throughout this roadmap, the aim is to foster synergy among these four paths, with each contributing to the development of a Digital Twin. By following this structured approach, organizations can create powerful Digital Twins and leverage their full potential in enhancing operational efficiency, decision-making and long-term competitiveness.

In the following sections, we will delve deeper into each of these four paths, providing a comprehensive understanding of each of the steps and considerations required along the path. The roadmap has been developed based on the integration of all knowledge and insights gathered and described in the previous sections, including literature, interviews, workshops, and the associated information analysis

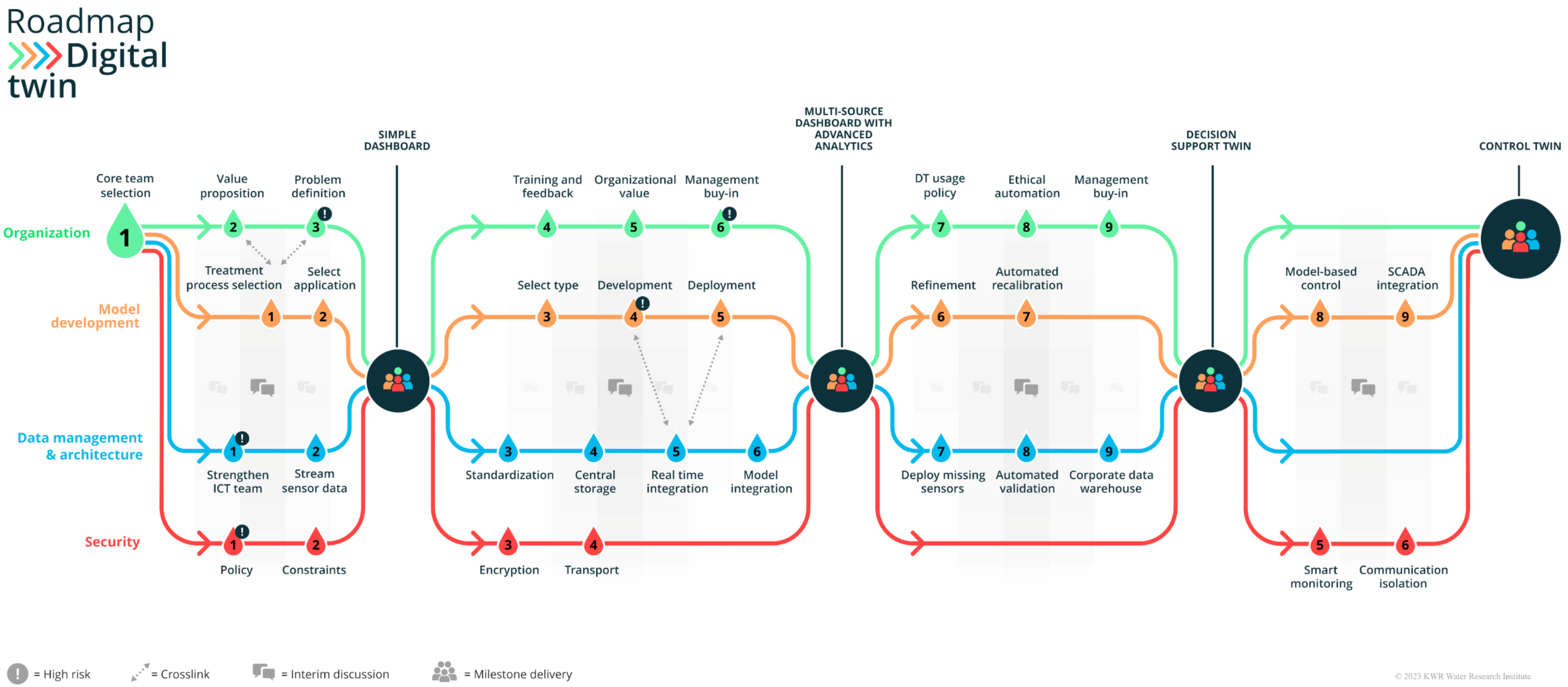


Figure 8 Roadmap to a Digital Twin: Overview

3.3.1 How to use the roadmap

Through extensive analysis of the needs, desires and challenges of the drinking water utilities this roadmap details the necessary actions needed to be taken for the development of Digital Twins.

The roadmap is read left to right along the four different lanes: Organisational, Model development, Data management and architecture and Security. The first step users should take when adopting this roadmap is to decide where their organization currently sits on each of the four paths. Users may already be further advanced in one (or more) roadmap paths then others. This step will help users realize where the focus should start in order to reach the next milestone. The coordination team can then establish their starting point.

The roadmap is broken down into four milestone with each milestone providing a tangible result that is beneficial to the organisation. The fours milestones are: 1. Simple dashboard, 2. Multi-source dashboard with advanced analytics, 3. Decision supporting Digital Twin and 4. Control Twin. Each milestone also acts as a Go/No-go stopping point where the coordination team should assess the needs, desires and challenges for the organization. Additionally, this encourages management involvement by showing value to the organization.

Between each roadmap milestone, it is necessary to facilitate regular interim discussions and assessments. These discussions serve as crucial touchpoints where the core team can engage with management to evaluate progress, adapt strategies if necessary and ensure alignment with organizational objectives. These interim discussions not only enhance transparency, but also enable timely adjustments fostering a dynamic responsive digitalization process that increases the likelihood of success.

Significant emphasis should be placed on the collaboration between the core team and the management team while following the roadmap on the route through digitalisation. Each organisation must also select the end point they wish to achieve on this roadmap that aligns with their goals and objectives. Although, the roadmap ends with a Control Twin, that level of control maybe not be necessary for each organisation. This collaborative effort will ensure that the initiative will not only reach its intended destination but also be effectively integrated and utilised by the organisation.

A full risk assessment of the roadmap was developed and can be found in Appendix II. The probability and impact levels of the assessment for the identified risks associated with a roadmap step were determined based on expert judgements and insights gained from the workshop conducted, as described in Section 3.2. The high and extreme risks are highlighted below for each milestone.

3.3.2 Milestone 1: Development of a Simple Dashboard

Initiation of Milestone 1

The inception of Milestone 1 marks the initiation of the Digital Twin development journey, with the primary aim of crafting a straightforward yet powerful dashboard. This dashboard is envisioned to offer an intuitive user interface, rendering key data in a comprehensible and accessible format. The fundamental goal is to expedite data-driven decision-making by providing stakeholders with a clear understanding of process health and performance. Achieving user-friendliness, clarity and ease of use is paramount.

Commencing Milestone 1 necessitates the assembly of a cross-functional Core Team. Each team member should possess a digital mindset and be equipped to tackle the challenges ahead. Notably, the inclusion of an Information and Communication Technology (ICT) specialist is pivotal to ensure the dashboard's functionality and its potential transformation into a full-fledged Digital Twin.

Objectives for Milestone 1

The objectives for Milestone 1 aim for the realization of a user-centric dashboard which encompass:

- Creating an intuitive interface that grants stakeholders real-time access to critical data.
- Enhancing decision-making by presenting data in a concise, user-friendly manner.
- Facilitating the rapid analysis of trends, patterns and key performance metrics.
- Prioritizing ease of use to enable both technical and non-technical users to navigate the dashboard effortlessly.
- Empowering stakeholders with actionable insights, enabling well-informed decision-making, progress tracking and continuous improvement.

Cross-Referencing in Milestone 1

In the roadmap's organizational path, Step 2, 'Value Proposition,' and Step 3, 'Problem Definition,' closely intertwine with the model development line's 'Treatment Process Selection.' The value proposition inherently guides the selection of the treatment process that the Digital Twin will focus on, aligning the technology's application with the organization's core objectives and the specific challenges it aims to address. Simultaneously, the problem definition, as detailed in Step 3, aids in the precise delineation of the issues that the Digital Twin will tackle within the chosen treatment process. This ensures a synergy between the organizational vision and the model development, setting the foundation for a Digital Twin that is not only technically robust but also impeccably aligned with the organization's strategic goals.

Deliverables of Milestone 1

Upon the successful completion of Milestone 1, efforts will culminate in a comprehensive inventory of sensors and pertinent data sources, specifically tailored to address the defined problem. This amalgamation of data will be organized into a centralized repository, thereby establishing the foundation for a functional dashboard.

Risk Mitigation in Milestone 1

Table 2 highlights the High and Extreme risks associated with the components of Milestone 1. These risks primarily revolve around resource constraints, knowledge gaps and the precise definition of the problem domain. To mitigate these risks, a judicious approach to team member selection, with an emphasis on expertise, is essential.

Of particular significance is the ICT policy governing data control and management. Its impact on the project's success is undeniable. Therefore, careful scrutiny of existing ICT policies and proactive revision, if needed, in consultation with management, is paramount to circumvent potential hindrances to the deployment of specific technologies.

Table 2 High and Extreme Risks of Milestone 1

Roadmap Step	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Core team selection	Lack of resources	All	4 Likely	4 Major	16	Extreme *(15-25)*
	Lack of knowledge	All	3 Possible	4 Major	12	High *(8-12)*
Problem definition	Lack of knowledge	All	4 Possible	5 Major	20	Extreme *(15-25)*
Treatment process selection	Lack of problem definition	All	3 Possible	4 Major	12	High *(8-12)*
Strengthen ICT team	Lack of resources	All	4 Likely	4 Major	16	Extreme *(15-25)*

Roadmap Step	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Policy	Lack of management buy-in	All	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Constraints	Policy creating issues on allowability of specific technology	All	**3** Possible	**4** Major	**12**	High *(8-12)*

3.3.3 Milestone 2: Multi-source Dashboard with Advanced Analytics

Goal for Milestone 2

The central objective in the development of a multi-source dashboard with advanced analytics lies in the creation of a dynamic platform capable of aggregating data from diverse sources and harnessing the power of advanced analytical tools, including soft sensors. This milestone aspires to construct a sophisticated interface that facilitates the monitoring, analysis and visualization of complex datasets. By doing so, it empowers users to make data-driven decisions. The emphasis is placed on the seamless integration of varied data streams, advanced analytical processes and the presentation of meaningful information in an accessible manner. This strategic approach aims to endow users with profound insights and reinforce their capacity for strategic decision-making.

A pivotal component of Milestone 2 involves the development of the chosen model for the Digital Twin. This encompasses data standardization within the data lake. Notably, a significant focus is directed toward Information and Communication Technology (ICT) planning, encompassing secure data transport and encryption methodologies.

Cross referencing in Milestone 2

In the Model Development path, Step 4, Development and Step 5, Deployment, form an integral partnership with Data Management & Architecture Step 6 Real-time Integration. The Development phase is where the chosen model for the Digital Twin takes concrete shape, incorporating advanced data analytics and algorithms. Simultaneously, in the Deployment phase, the model is operationalized and integrated into the system, ushering in real-time data collection and feedback. This dynamic integration corresponds with the Data Management & Architecture line's focus on the organization, storage and accessibility of data. Furthermore, the Real-time Integration aspect ensures that the real-time data feeds from the model enhance the Digital Twin's capacity to provide continuous, accurate insights, thereby reinforcing the synergy between model development and data architecture.

Output for Milestone 2

Upon the successful completion of Milestone 2, the desired outcome is the realization of the chosen model for the Digital Twin. This entails the integration of live model sensor fusion. A pivotal aspect within this milestone revolves around the secure transport and encryption of data. All data, essential for the Digital Twin, should conform to a standardized data format to facilitate seamless integration into the data lake. From the data lake's modelling outputs, soft sensors can be automatically integrated into the models in real-time.

The ultimate deliverable, showcasing value to the organization and persuading management to proceed to Milestone 3, will be a fully-fledged dashboard drawing from multiple data sources, equipped with (user-friendly) soft sensors and statiYustical analytics.

Risks Associated with Milestone 2

Table 2 provides a comprehensive overview of the High and Extreme risks associated with Milestone 2. These risks encompass a spectrum of challenges, from data security to model development complexities. Mitigating these risks necessitates a vigilant approach, encompassing meticulous planning, rigorous testing and the involvement of experts in data transport and encryption protocols. The successful management of these risks is essential for the seamless progression of the project.

Table 3 High and Extreme Risks of Milestone 2

Roadmap Step	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Select type	Unclear understanding of necessary models	Developers, Modellers	4 Likely	4 Major	16	Extreme *(15-25)*
Development	Incorrect/insufficient data	Developers, Modellers	5 Almost Certain	4 Major	20	Extreme *(15-25)*
	Lack of sensors	Developers, Modellers, Operators	4 Likely	3 Moderate	12	High *(8-12)*
Deployment	Run time is too long	Developers, Modellers, IT	3 Possible	3 Moderate	9	High *(8-12)*
Standardization	Lack of investment in data management plan	Developers, Modellers, Operators, Management	4 Likely	4 Major	16	Extreme *(15-25)*
Model Integration	Data errors creating incorrect soft sensor data	Developers, Modellers, Operators, Management	4 Likely	3 Moderate	12	High *(8-12)*
Encryption	Cyber-security of data	All	3 Possible	3 Moderate	9	High *(8-12)*
Transport	Cyber-security of data	All	3 Possible	3 Moderate	9	High *(8-12)*

3.3.4 Milestone 3: Decision Support Twins

Goal for Milestone 3

The objective in the development of a Decision Support Digital Twin is to engineer an accurate, real-time and versatile Twin. This tool is designed to comprehensively model physical systems, seamlessly integrate real-time data, deliver predictive insights, facilitate scenario analysis and optimization, provide an intuitive user interface, ensure robust security and scalability, establish a feedback loop for continuous learning and, above all, enhance decision-making while optimizing cost-effectiveness and operational efficiency.

This milestone's achievement hinges on the meticulous refinement and recalibration of the dashboard developed in Milestone 2 to transform it into a Decision Support Twin. This transformation involves the deployment of any missing hard and soft sensors, alongside the implementation of automated validation processes. The data gathered during this process must be securely stored in the corporate data warehouse.

Output for Milestone 3

The culmination of Milestone 3 yields a product that aligns with the definition detailed in Section 2.2.3 of the Literature Review, thereby realizing a Decision Support Twin. At this advanced level of Digital Twinning, operators gain the capability to run system scenarios within the process, subsequently enabling more informed and data-driven decision-making.

Risks Associated with Milestone 3

Table 4 provides an exhaustive analysis of the High and Extreme risks linked with Milestone 3. These risks span various challenges, encompassing data security, model recalibration complexities and the potential for system disruptions. Attentive risk management and meticulous execution are paramount to the successful advancement of this pivotal stage in the project.

Table 4 High and Extreme Risks of Milestone 3

Roadmap Step	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Digital Twin usage policy	Lack of acceptance and incapability to switch to a new way of working	Operators, Supervisors	3 Possible	3 Moderate	9	High *(8-12)*
Ethical automation	Ethics of Digital Twin making decisions. Liability	All	4 Likely	4 Major	12	High *(8-12)*
Management buy-in	Results do not prove necessity of management	All	3 Possible	5 Catastrophic	15	Extreme *(15-25)*
Refinement	There has been no proper development pipeline implemented which makes deploying of new versions hard	Developers, Modellers Data Expert, IT	3 Possible	3 Moderate	9	High *(8-12)*
Automated validation	Lack of investment in proper infrastructure (context broker)	All	2 Unlikely	4 Major	8	High *(8-12)*
	Inaccurate models for reconciliation	Data Engineer	2 Unlikely	4 Major	8	High *(8-12)*
Corporate data warehouse	Lack of investment in infrastructure (databases and servers/cloud)	IT Data Manager, Management	4 Likely	5 Catastrophic	20	Extreme *(15-25)*

3.3.5 Milestone 4: Control Twin

Goal for Milestone 4

The central aim in the development of a control Digital Twin is the creation of a resilient and adaptable tool that replicates physical systems and processes, while enabling seamless real-time data integration and control. The overarching goal is to facilitate predictive control, optimize system performance, offer a user-friendly interface for monitoring and interventions, ensure data security, scalability, compliance with standards and foster continuous learning and adaptation. This holistic approach is geared towards enhancing system control and efficiency, while simultaneously optimizing cost-effectiveness.

The attainment of this milestone's goal hinges on advancing model development and introducing model-based controls. Furthermore, real-time integration of the Supervisory Control and Data Acquisition (SCADA) system is pivotal. Model-based controls encompass human-defined protocols for decision-making under specific operating parameters. The introduction of these elements heightens the importance of security measures. 'Smart monitoring' is imperative to mitigate security risks and the establishment of communication isolation, or air-gapping, provides an additional layer of robust cybersecurity.

Output for Milestone 4

The culmination of Milestone 4 yields a product that aligns with the definition outlined in Section 2.2.3 of the Literature Review, thereby realizing a Control Twin. At this advanced level of Digital Twinning, the Control Twin is endowed with the capability to autonomously make decisions for optimizing the operating conditions of the treatment process. This entails real-time adjustments such as regulating chemical concentrations within the system, resulting in agile, data-driven control.

Risks Associated with Milestone 4

The risks associated with Milestone 4 consist of security concerns, data integrity and safeguard against breaches. Additionally, the establishment of communication isolation, or air-gapping, introduces a layer of cybersecurity. Mitigating these risks requires rigorous planning, vigilant monitoring and expert-level coordination to achieve a secure and successful milestone.

There are high or extreme risks associated with this phase which will have already been addressed through the progress towards the first three milestones.

3.3.6 Organization

In Table 5, definitions of the different steps in the Organization path of the Roadmap (Figure 9) are provided.

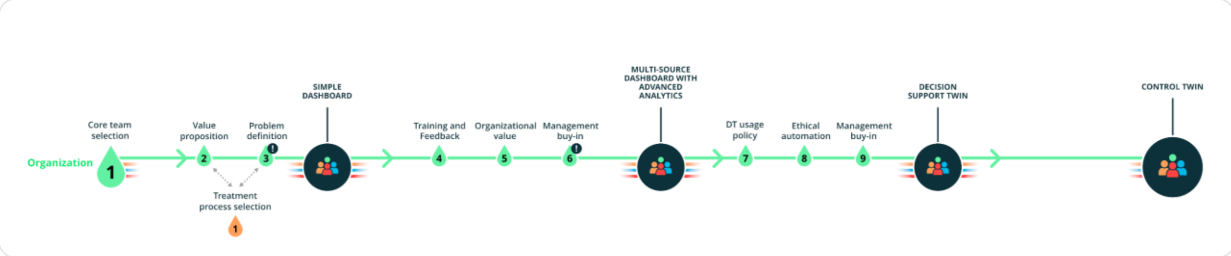


Figure 9 Roadmap: Organization path

Table 5 Organization path explanation

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
1	Core team selection	The process of carefully choosing a dedicated team of experts and stakeholders within the organization responsible for leading and executing the Digital Twin development project.	-	-
2	Value Proposition	Articulating the specific benefits, advantages and strategic value that the Digital Twin will provide to the organization, including its potential to address key challenges and deliver positive outcomes.	Model Development 1	-
3	Problem definition	Clearly defining the specific problem or challenge the Digital Twin is intended to address, ensuring alignment with organizational goals and objectives.	Model Development 1	High Risk
4	Training and feedback	Providing training and continuous feedback mechanisms to ensure that employees and stakeholders understand and can effectively utilize the Digital Twin technology.	-	-
5	Organizational value	Assessing and articulating the potential value that the Digital Twin will bring to the organization, including improvements in efficiency, decision-making and cost savings.	-	-
6	Management buy-in	Securing support and commitment from top management and decision-makers to allocate resources.	-	High Risk
7	Digital Twin usage policy	Establishing clear guidelines and policies for the responsible and ethical use of Digital Twin, including data privacy, usage, retention and security considerations.	-	-
8	Ethical automation	Ensuring that the automation and decision-making processes within the Digital Twin align with ethical principles and do not compromise fairness or transparency.	-	-
9	Management buy-in	Securing support and commitment from top management and decision-makers to allocate resources.	-	-

3.3.7 Model Development

In Table 6, definitions of the different steps in the Model Development path of the Roadmap (Figure 10) are provided.

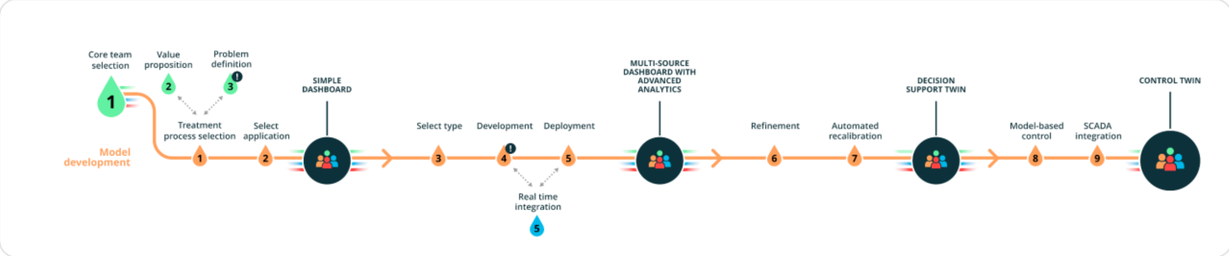


Figure 10 Roadmap: Model Development path

Table 6 Model Deployment path explanation

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
1	Treatment process selection	The initial step in the roadmap involves choosing the specific treatment process that the Digital Twin will focus on, considering factors like relevance and impact.	Organization 2 & 3	-
2	Select application	Identifying the practical applications and use cases where the Digital Twin will be applied within the chosen treatment process.	-	-
3	Select type	Determining the type of model Digital Twin (e.g., process-based, data-driven, hybrid)) that best suits the selected treatment process based on data availability, process complexity and computational costs. Further details can be found in Section 2.4.4.	-	-
4	Development	The phase where the actual Digital Twin model is built, involving creating the model, integrating data sources and developing the necessary algorithms and interfaces.	Data management & Architecture 5	High Risk
5	Deployment	Deploying the developed Digital Twin model into the operational environment, ensuring it can interact with the physical treatment process and collect real-time data.	Data management & Architecture 5	-
6	Refinement	Continuously improving and optimizing the Digital Twin's model performance based on real-world data and feedback, enhancing its accuracy and effectiveness.	-	-
7	Automated recalibration	Implementing automated recalibration mechanisms to ensure that the Digital Twin model remains accurate and up-to-date with changes in the treatment process.	-	-

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
8	Model-based controls	Integrating model-based controls to allow the Digital Twin to influence and optimize the treatment process through automated adjustments.	-	-
9	SCADA integration	Integrating the Digital Twin with Supervisory Control and Data Acquisition (SCADA) systems to enhance real-time monitoring, control and data exchange capabilities.	-	-

3.3.8 Data Management and Architecture

In Table 7, definitions of the different steps in the Data Management and Architecture path of the Roadmap (Figure 11) are provided.

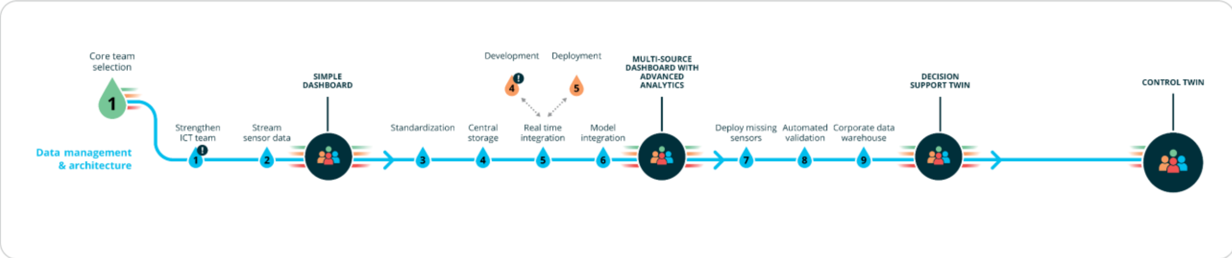


Figure 11 Roadmap: Data Management & Architecture path

Table 7 Data Management & Architecture path explanation

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
1	Strengthen ICT team	The initial step involves reinforcing the ICT team's skills, expertise and resources to effectively support the development and operation of Digital Twins	-	High Risk
2	Stream sensor data	Establishing efficient channels and protocols to collect and transmit sensor data from various sources to the Digital Twin in real-time, ensuring data availability for analysis.	-	-
3	Standardization	Implementing standardized data formats, communication protocols and best practices to ensure interoperability and consistency across Digital Twin-related processes and systems.	-	-
4	Central storage	Setting up a centralized and secure data storage infrastructure to store and manage the vast amount of data generated and utilized by Digital Twins for easy access and analysis.	-	-
5	Real time integration	Enabling real-time data integration and synchronization between the Digital Twins and other ICT systems to support timely decision-making and responsiveness.	Model Development 4 &5	-
6	Model integration	Integrating various Digital Twin models and simulations to create a holistic view of the organization's assets and processes, enhancing the Digital Twin's analytical capabilities.	-	-
7	Deploy missing sensors	Identifying and deploying additional sensors or data sources to fill data gaps and enhance the Digital Twin's ability to monitor and analyse the entire system effectively.	-	-

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
8	Automated validation	Implementing automated validation processes to ensure the accuracy and reliability of data, models and predictions.	-	-
9	Corporate data warehouse	Establishing a centralized corporate data warehouse to consolidate and manage data from multiple sources..	-	-

3.3.9 Security

In Table 8, definitions of the different steps in the Security path of the Roadmap (Figure 12) are provided.

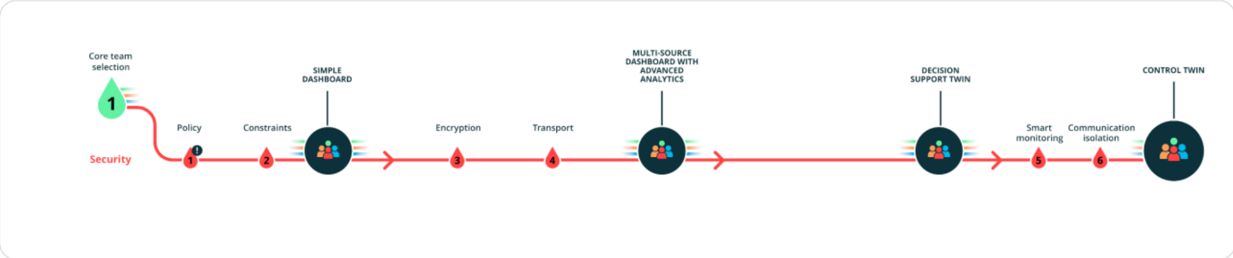


Figure 12 Roadmap: Security path

Table 8 Security path explanation

Roadmap Step no.	Roadmap Step name	Definition	Cross-referencing	Risk
1	Policy	Establishing comprehensive cybersecurity policies that outline guidelines, rules and procedures for securing Digital Twins and associated data, ensuring compliance and protection.	-	High Risk
2	Constraints	Identifying and implementing constraints and access controls to limit and regulate who can access and modify Digital Twins and their data and from what locations, minimizing security risks.	-	-
3	Encryption	Implementing robust encryption mechanisms to safeguard data both at rest and in transit within the Digital Twin ecosystem, protecting it from unauthorized access or interception.	-	-
4	Transport	Ensuring secure data transport and communication protocols to prevent data breaches, including secure connections between Digital Twins and external systems or networks.	-	-
5	Smart monitoring	Employing smart monitoring tools and techniques to continuously assess the cybersecurity posture of Digital Twins, detecting and responding to potential threats and suspected intrusions in real-time.	-	-
6	Communication isolation	Implementing communication isolation measures to segment and protect different parts of the Digital Twin ecosystem, preventing lateral movement of cyber-threats within the network.	-	-

4 Functional Design

4.1 Overview

Based on a comprehensive literature review and interviews with leading organisations, a roadmap was developed for the creation and implementation of a Digital Twin for water treatment. As a subsequent phase, this roadmap was put into action and tested in a real-world application. The practical testing took the form of a functional design, where a softening treatment process at De Watergroep's (DWG) Bilzen water treatment facility served as the pilot case. In the sections below, a concise introduction to the softening process is provided, followed by a detailed overview of the methodology adopted for the functional design. Subsequently, the outcomes of the functional design are discussed, including a comprehensive set of requirements and services that should be integrated into a Digital Twin for this specific process. Finally, a series of actions and target recommendations are presented.

4.2 Softening Treatment Process at WPC Bilzen

The softening treatment process at the drinking water production facility in Bilzen is an integral component of a series of treatment processes dedicated to supplying drinking water to approximately 50,000 residents in the regions of Bilzen and Riemst, situated within the province of Limburg. A schematic process-flow diagram of WPC Bilzen can be found in Figure 13. In summary, raw groundwater is sourced from five wells, pumped to the water treatment works and subsequently distributed to three pellet-softening reactors. The number of pellet reactors in operation at any given time is determined by the volume of raw groundwater extracted from the wells. This, in turn, is regulated by the water level of the reservoir that stores the drinking water. As drinking water is distributed to customers, leading, to a decrease in the water level of the reservoir, the extraction flow-rate from the wells (and the number of wells in use) is increased. This adjustment controls the number of softening reactors online. The distribution of the total flow-rate is a complex process due to variations in hydraulic resistance between the reactors. To address this complexity, a dedicated control-logic was designed and implemented within the Programmable Logic Controller (PLC). This ensures that the reactor with the highest hydraulic resistance has its inlet valve fully opened, while the inlet valves of the remaining reactors are adjusted to guarantee an equitable distribution of the flow between all of the available reactors.

The softening reactors are filled with fine-grained sand. To raise the pH for the softening process, sodium hydroxide (NaOH) is proportionally dosed with respect to the feed water flow-rate and mixed intensively. The chemical increase in the pH due to the addition of NaOH leads to the precipitation of calcium carbonate and lime from the water, which then crystallises on the fine sand, forming pellets. The dosing of NaOH is primarily determined by the output hardness requirement of the reactor. Over time, the pellets grow, reducing the surface area available for new pellet formation. Used pellets are regularly removed and new fine-grained sand is introduced into the reactor. The effluent from the softening reactor undergoes pH adjustment through the addition of carbon dioxide (CO_2). Within the softening reactor, online measurements are taken for the water level, pH, hardness and turbidity. The turbidity is used to control the suspension of particles and assess the cloudiness of the water. This data informs pellet-formation rate and guides the related operations for pellet removal and the addition of new sand.

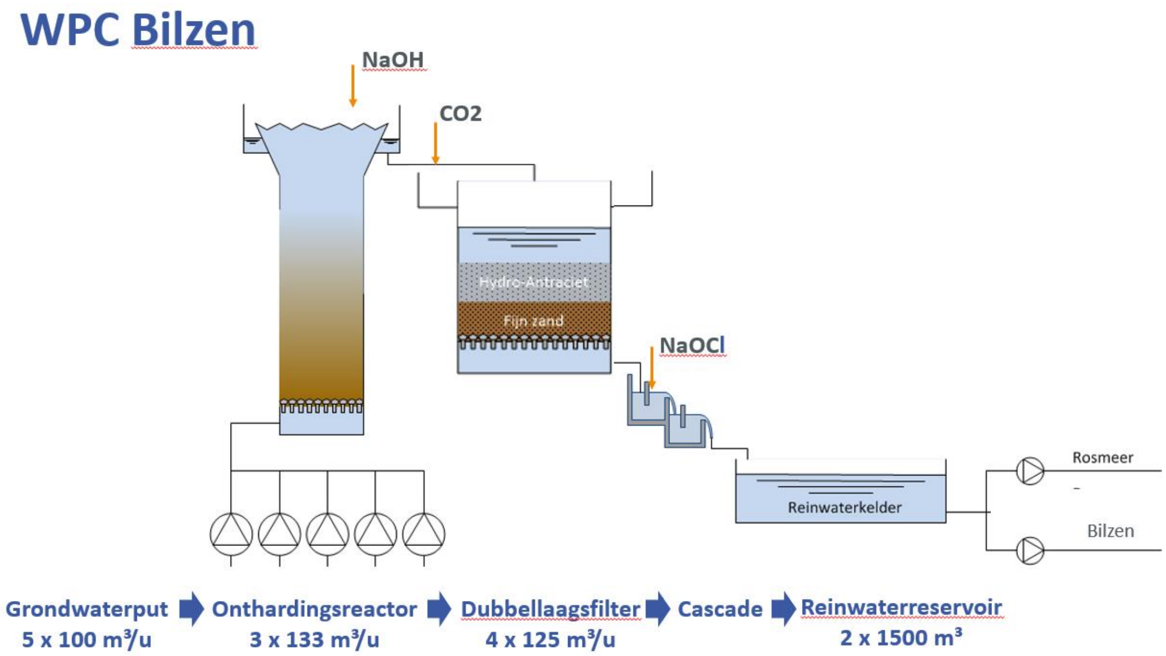


Figure 13 Schematic process flow diagram of WPC Bilzen. The Softening treatment process (onthardingsreactor) is fed with raw groundwater for hardness removal and then sent to the double filtration system.

The selection of the Bilzen softening process for this functional design was determined by several factors. Firstly, it is one of the simpler softening processes within DWG, making it an accessible starting point for investigating the potential benefits of a Digital Twin. Secondly, a process engineer has developed an offline model using PhreeqC for simulating scenarios, such as assessing pH under different chemical-dosing rates and raw water compositions. Connecting this model with real-time data can offer valuable insights for operational and strategic decisions. Lastly, despite the well-known nature of the softening process, its dynamic operations from hydraulic and water quality perspectives provide opportunities for a Digital Twin. This includes advanced monitoring, data analysis, cost savings and sustainability improvements, especially in chemical dosing and energy consumption, while upholding stringent drinking water quality standards.

4.3 Methodology & Functional Design Thinking

The functional design process and design thinking adopted for this pilot case incorporated elements from an enterprise architecture methodology and framework which were tailored to align with DWG's organisational structure and the stakeholders involved. Figure 11 illustrates the functional design framework, which integrates components from recognised open standards. Notably, it draws from The Open Group Architecture Framework (TOGAF) enterprise architecture (TOGAF 9.2, 2018), combined with the concepts elaborated in the Digital Twin roadmap. Specifically, the framework leverages terminology and concepts of motivational elements, as part of the TOGAF enterprise architecture design, such as identifying *Drivers & Concerns*, *Requirements*, *Goals* and *Outcomes*. Concurrently, it integrates elements such as core team selection, design and a value proposition, which are fundamental concepts of the Digital Twin roadmap. The core roadmap elements selected in this methodology represent the initial steps in the journey of creating and integrating a Digital Twin. These steps are essential precursors. As the project progresses, during which the identified requirements are detailed and the design of Digital Twin applications are considered, additional elements of the roadmap, such as model selection, model deployment, central storage for data and real-time integration, will come into focus. However, this level of detail is beyond the scope of this functional design exercise and should be considered as follow-up work to complement the current analysis.

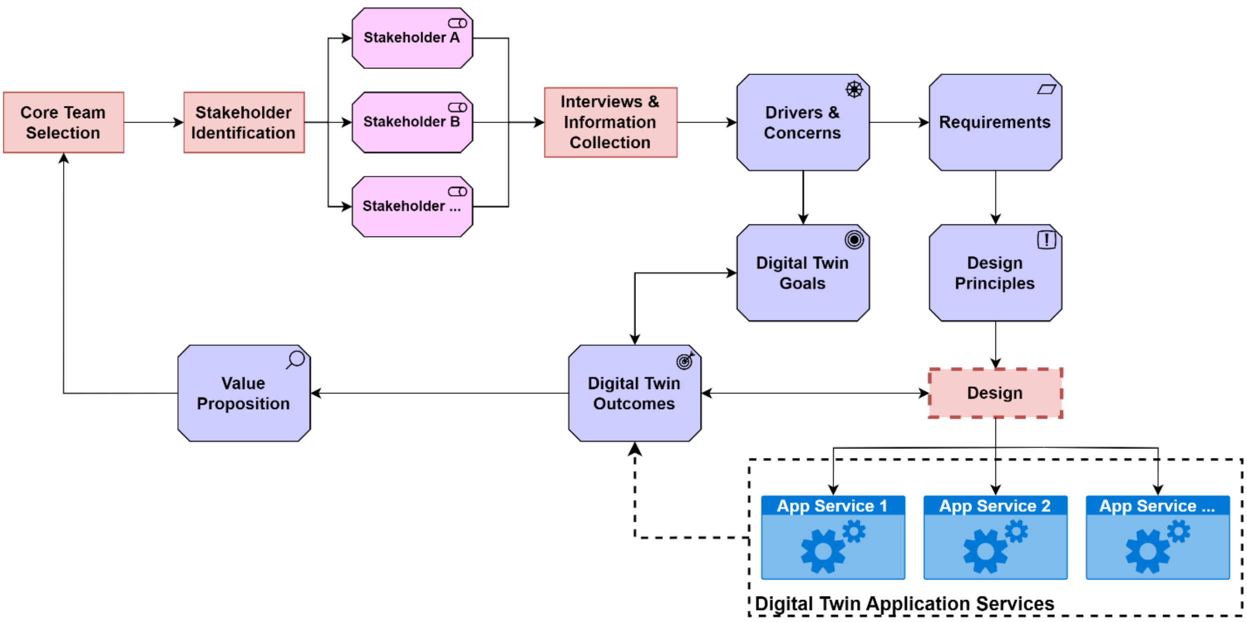


Figure 14 Customised functional design methodology framework containing Digital Twin roadmap concepts, motivation elements and identification of Digital Twin application services.

The flow of the framework and adopted methodology is as follows:

- Initially, a core team is assembled, consisting of individuals with essential skills and expertise within the organization. This forms the initial and primary step in the Digital Twin roadmap.
- Key stakeholders relevant to the softening process and its operations are identified based on their roles and responsibilities.
- Interviews are conducted with these key stakeholders to gather essential information and data.
- An analysis of the responses is performed to identify stakeholders' concerns regarding the softening process and the driving factors for implementing a Digital Twin. These drivers and concerns represent the motivations for stakeholders to implement changes in pursuit of specific goals.
- The identified drivers and concerns are used to compose a list of requirements essential to designing a Digital Twin. In addition, they provide valuable insights for defining the primary Digital Twin goals, representing the intended end state expected by stakeholders and the organization.
- The requirements defined serve as the basis for formulating the design principles, which offer governing guidelines and general attributes that any developed solution must adhere to.
- Building upon the requirements and adhering to specific design principles, the design phase begins, allowing the development of tangible solutions in the form of functional Digital Twin application services.
- During the design of the application services, it is important to make estimations and projections about the potential outcomes for the Digital Twin. These are then compared against the defined Digital Twin goals. This assessment of potential outcomes informs the core team about the Digital Twin's potential value proposition, aiding in deciding its economic feasibility based on the expected impacts. Conversely, the outcomes, as agreed upon based on the requirements and ambitions of stakeholders, can also help inform and guide the designing of the application services. Additionally, actual use of the deployed Digital Twin applications will allow for a reassessment of the achieved outcomes, which can be used to validate the original, projected value proposition.

As can be seen, the functional design methodology framework has been conceptualised to be both general and adaptable for implementation in various systems and processes within a water company. Certain components of the functional design, especially those related to the motivational elements, can serve as recurring activities that can be performed at different stages in the lifecycle of developing the Digital Twin. Specifically, such a framework can be followed prior to attaining the specific milestones outlined in the roadmap. In the context of this design framework, applied to the softening process case, it is essential to clarify that the actual design and development of Digital Twin applications falls outside the scope of this project. However, potential applications have been identified based on the collected information. It is important to note that, as these applications have not been developed and deployed, assessing the Digital Twin outcomes is currently not feasible. Nevertheless, an example of potential outcomes is provided based on the specific Digital Twin goals identified.

4.4 Functional Designing Digital Twin Creation for a Softening Treatment Process

4.4.1 Core-team Selection and identified stakeholders

The successful development and implementation of a Digital Twin for the softening treatment process at DWG requires the careful selection of a dedicated core team and identification of key stakeholders. This section provides insights into the critical aspects of this initial phase of the function design.

Core-Team selection

The core team is the backbone of the Digital Twin development project. It is vital to assemble a team of individuals with a diverse set of skills, expertise and experiences to ensure the success of the project. The core team for the project should include the following roles:

1. *Project manager:* A skilled project manager who oversees the entire project ensuring it stays on schedule and within budget. This individual will be responsible for managing the resources, coordinating tasks and monitoring progress across all stakeholder groups. This is particularly important due to the diversity of the stakeholder groups and the demands of different departments responsible for data across the data domain owners.
2. *Domain Experts:* Water treatment specialists with a deep understanding of the softening treatment process. These experts will provide valuable insights and guidance in modelling and simulating the softening process accurately.
3. *Data scientists, modellers and data domain owners:* Experts in these topics are crucial for providing the data for the Digital Twin. Additionally, they are responsible for creating and refining the Digital Twin's mathematical models, algorithms and predictive capabilities. Their expertise is vital to produce accurate and reliable results.
4. *ICT/OT specialists:* Information and Communication Technology (ICT)/Operational Technology (OT) specialists will focus on the technical aspects of data integration, real-time monitoring and system integration. They will ensure the Digital Twin can seamlessly communicate with sensors and other data sources.
5. *Security Experts:* Given the sensitivity of water treatment processes and the potential implications of a Digital Twin, experts in data privacy and cybersecurity should be included to guarantee security measures are integrated into the project from the outset.

Identification of Stakeholders

In addition to the core team, it is essential to identify and engage with relevant stakeholders. For the functional design a stakeholder analysis was conducted and the following stakeholders were identified:

1. *Softening treatment process plant operators:* These individuals play a critical role in the day-to-day operation of the softening process. Involving them from the beginning will help ensure that the Digital Twin aligns with their operational capabilities, needs and goals.

2. ***Process engineer:*** Process engineers are essential for understanding the technical intricacies of the softening treatment process. They can contribute their knowledge to ensure the Digital Twin accurately models and optimizes the treatment process.
3. ***Head of operational technology:*** This stakeholder is crucial for ensuring the integration of the Digital Twin with existing operational systems and technology. They can help align the project with current infrastructure and company goals.
4. ***Head of Research and Development (R&D):*** The Head of R&D can provide strategic guidance and insights into cutting- edge technology, ensuring the Digital Twin remains relevant, innovative and impactful.
5. ***Head of operations:*** The Head of operations oversees the entire operational process of the drinking water treatment overall. They can provide top-level support, ensuring the Digital Twin for softening aligns with the broader organisational goals.
6. ***ICT Manager/Department:*** The ICT department is critical for technical support and cyber security considerations. They play a pivotal role in data integration and system security.
7. ***Programmable Logic Controller (PLC) programmer:*** This stakeholder will ensure the integration of sensors, data transfer and automation of the process into the Digital Twin.
8. ***Energy and Environmental coordinators:*** This stakeholder has a high interest in the accuracy of the Digital Twin. DWG has internal goals for sustainability, resource efficiency and emission reduction.
9. ***Quality control manager:*** Quality control experts provide valuable input to maintain water quality and safety standards needed from the softening progress to program the Digital Twin project. The quality control department are one of the primary users of the data and predictive nature of the Digital Twin.
10. ***Data Architect:*** The data architect is responsible for designing the framework for organizing and managing data, within the company, needed for the Digital Twin.
11. ***Supervisory Control and Data Acquisition (SCADA) system manager:*** Stakeholder who will facilitate real-time monitoring of control data which is utilized in the Digital Twin.
12. ***Sensor company:*** Multiple companies which provide critical, high-quality sensors and data sources, ensuring that the Digital Twin receives accurate and reliable real-time data. A disruption within the supply chain of the sensors could lead to a critical failure of the Digital Twin.
13. ***Scientific researchers (KWR):*** As a member of the BTO, DWG works closely with KWR on water research across many different topics. KWR researchers provide valuable insights into the latest advancements and research into water-data ecosystems. Researchers' expertise can help ensure the Digital Twin incorporates state-of-the-art knowledge and techniques.

From the stakeholder identification process eight groups of stakeholders were selected to be interviewed for the functional design. The selection was based on their roles and responsibilities, and level of influence and interest specific to the softening treatment process. These stakeholders were: ***the operators***, ***process engineers***, ***R&D Department***, *PLC programmers*, ***energy and environmental coordinators***, ***water quality control department***, ***data architects*** and the *ICT department*.

4.4.2 Value Proposition and Roadmap Endpoint Selection

Developing and deploying a Digital Twin for any treatment process necessitates a significant capital investment, encompassing expenses from hiring personnel, deploying workforce for designing and deploying application services, investing in software and hardware and allocating organisational resources. However, particularly with digital technologies, the return on investment (ROI) is not always straightforward and clear. The uncertain nature of the benefits associated has posed a substantial obstacle to actively pursuing ventures in digital technology development, including Digital Twins. Therefore, it is very important to create a value proposition, as early as possible in the development lifecycle, which involves demonstrating the potential benefits and advantages the Digital Twin can offer to stakeholders and the organisation. For the softening process, this can include showcasing how it can improve operational efficiency, reducing costs, enhancing decision-making, ensure water quality and support sustainability goals. The value proposition can include quantitative and qualitative elements.

In this functional design process, the concept of defining Digital Twin Outcomes has been introduced to support the creation of the value proposition. It is recommended to define the outcomes during the design phase of the Digital Twin Application Services, which will be sufficiently informed, by the stakeholders needs and organisational ambitions, as discussed in the sections below. Identifying outcomes during the design phase offers insights that can inform the core team and management, providing valuable information before the complete design and development of the Digital Twin, which will require further investment and resource allocation. Although this functional design does not cover the detailed design and development of the application services, Section 4.4.8 includes a brief discussion on Digital Twin Outcomes along with an example of how these outcomes can be quantified, using a specific use-case that the Digital Twin addresses.

Another crucial element contributing to the streamlining of the entire Digital Twin design process is the selection of an endpoint in the Digital Twin Roadmap, akin to the concept of *Backcasting*. This endpoint serves as a clear and defined goal, providing direction for decision-making based on information gathered from the value proposition. Following discussions with stakeholders, particularly at the management level, a clear indication was given that DWG would like to consider the endpoint to be ***Milestone #3 – Decision Support Twin***.

4.4.3 Drivers and Concerns

The questions posed to the stakeholders regarding drivers and concerns were based on their beliefs concerning their roles and responsibilities within the organization, with a particular focus on the softening treatment process. Drivers, whether originating internally or externally, serve as motivating factors for stakeholders to advocate and initiate necessary changes. Concerns can encompass various aspects of a system, including its functioning, operations, development, performance, reliability and evolution. Table 9 below displays the identified Drivers & Concerns (D&C), which have been aggregated based on the insights gathered from the interviews and grouped according to their relevance to specific stakeholders.

Table 9 Drivers & concerns aggregated from responses by stakeholders.

S. No.	Drivers & Concerns	Description	Relevant to Stakeholder
D&C.1	Downtime of assets	Reducing the duration of physical assets being non-operational.	PLC Programmer, Process Engineer
D&C.2	Device health	Monitoring of deviations in data from sensors due to calibration issues, fouling, etc.	PLC Programmer, Process Engineer
D&C.3	Failover arrangements when Digital Twin fails	Ensuring that in the event of non-availability of e.g. setpoints from the Digital Twin, a process is in place to fallback to sensible manual settings.	PLC Programmer, Process Engineer

S. No.	Drivers & Concerns	Description	Relevant to Stakeholder
D&C.4	Future operator knowledge and working efficiency	Safeguarding and retaining the knowledge embodied in the operators and ensuring that working practices such as site visits are conducted efficiently.	PLC Programmer, Process Engineer
D&C.5	Help daily working	Provide easily accessible and necessary data on quality parameters and physical assets, particularly when on-site.	Water Quality Control, Operator
D&C.6	Process optimisation	Optimize distribution of water between reactors (hydraulic) and enhance process performance to maintain high water quality including controlling hardness, pH and turbidity along with the efficient utilisation of chemicals (NaOH and CO_2 dosing).	PLC Programmer, Process Engineer
D&C.7	Increased process understanding	Enhance understanding of the softening process operations by having information on unmeasured parameters, performing what-if simulations and conducting scenario analysis.	Process Engineer
D&C.8	Intelligent chemical dosing	Optimizing the use of chemical additives to the drinking water to minimize consumption and maximize efficacy.	Process Engineer, Energy & Environmental Coordinator
D&C.9	Unexpected events	Unforeseen or unplanned events, such as water quality issues or asset problems, can lead to work-related stress and potential disruptions in the continuous supply of drinking water.	Process Engineer

S. No.	Drivers & Concerns	Description	Relevant to Stakeholder
D&C.10	Economics	Overall economic impact in the form of reducing operation costs, minimising chemical dosing expenses and decreasing the lag time for data availability used in decision-making.	Process Engineer, Energy & Environmental Coordinator, Director Innovations
D&C.11	Good water quality	Ensuring water output from the process meets the regulatory and business requirements in terms of quality	Process Engineer, Water Quality Control, Operator, Director Innovations
D&C.12	Heterogenous internal data sources	Minimizing the difficulties associated with integrating disparate data sources; ensuring that data is made available to all tiers of the business.	Process Engineer, Water Quality Control, Energy & Environmental Coordinator, Operator, Data Architect
D&C.13	Data quality	The reliability of data measured by laboratories, online sensors and on-site measurements. This involves automated validation of data that reduces manual labour required to validate data enhancing data quality.	Process Engineer, Water Quality Control, Data Architect
D&C.14	Data unavailable	Unavailability of crucial data due to limited, access, poor data management and lag time in receiving up-to-date data.	Water Quality Control, Data Architect
D&C.15	Data security	Minimizing the risk of unauthorized access to data as well as ensuring confidence in the integrity of the data recorded and security of the communications between components of the Digital Twin.	PLC Programmer, Data Architect

S. No.	Drivers & Concerns	Description	Relevant to Stakeholder
D&C.16	Environmental regulatory changes & climate change goals	Challenges in achieving climate goals arise from increasingly strict regulations and current operations are expected to lead to higher chemical and energy consumption.	Energy & Environmental Coordinator
D&C.17	Unclear responsibilities	Ambiguity in the allocation of responsibilities: a lack of accountability for processes across various departments and ineffective communication means that responsibilities maybe overlooked.	Process Engineer, Water Quality Control
D&C.18	Silos in organisation	Inability or unwillingness to adequately share data and information across the organisation resulting in a lack of common understanding of business processes and interests.	Process Engineer, Operators, Director Innovations, Energy & Environmental Coordinator, Water Quality Control

4.4.4 Digital Twin Goals

By consolidating the D&Cs expressed by key stakeholders, the next stage involves shaping the Digital Twin Goals (DTGs). These objectives encapsulate a distinct purpose and a specific desired end state that stakeholders aim to achieve through the Digital Twin. Consequently, the Digital Twin must be designed to achieve these overarching goals, ultimately addressing the primary D&Cs. In Table 10, the DTGs have been stated along with the D&Cs addressed.

Table 10 Goals for the softening treatment process Digital Twin

S.No.	Goal	D&C Addressed
DTG.1	Increase understanding and optimise softening process operations and associated assets	D&C.3, D&C.4, D&C.5, D&C.6, D&C.7, D&C.8, D&C.11, D&C.17, D&C.18
DTG.2	Assist in intelligent chemical dosing	D&C.3, D&C.4, D&C.5, D&C.6, D&C.7, D&C.8, D&C.9, D&C.10, D&C.11
DTG.3	Reduce softening process operational costs	D&C.1, D&C.2, D&C.3, D&C.4, D&C.5, D&C.6, D&C.7, D&C.8, D&C.10

S.No.	Goal	D&C Addressed
DTG.4	Provide centralised location for relevant heterogenous data	D&C.12, D&C.13, D&C.14, D&C.15, D&C.17, D&C.18
DTG.5	Assist in sustainability goals	D&C.4, D&C.6, D&C.8, D&C.10, D&C.16,

4.4.5 Requirements

In addition to the formulated DTGs, the D&Cs serve as a robust foundation for identifying a set of requirements for the Digital Twin. When a specific goal is adequately decomposed, it yields a series of properties expected from the Digital Twin. These properties, when translated into actionable terms, become the requirements that essentially act as the ‘means’ to achieve the DTGs. Consequently, considering the D&Cs and the DTGs, a comprehensive list of requirements has been compiled, detailed in Table 11. It is important to note that multiple requirements may contribute to the realisation of a single DTG, even if they differ significantly and pertain to different elements or properties. Additionally, these requirements may involve the collaboration of various stakeholders who may not typically work together.

Table 11 List of requirements that the softening treatment process Digital Twin must fulfil

S.No.	Requirements
R.1	Data Availability
R.2	Data Timeliness
R.3	Data Accessibility
R.4	Data Security
R.5	Data Interpretability
R.6	Data Reliability
R.7	Optimise number of reactors online at a given time
R.8	Optimise flow division for available reactor
R.9	Optimise the softening treatment chemical process
R.10	Enable more remote working – minimise on-site attendance
R.11	Monitor sensor reliability
R.12	Availability of unobserved parameters
R.13	Monitoring the quality and quantity of chemicals used (based on process performance)

S.No.	Requirements
R.14	Monitoring the softening treatment operation costs
R.15	Forecasting to inform operational decisions (of the order of days)
R.16	Forecasting to inform tactical decisions (of the order of months)
R.17	Being able to conduct what-if simulations/scenario analysis/decision support
R.18	Human-in-the-loop: flexibility in how monitoring/feedback systems behave such as user-defined alarms
R.19	Dashboard that includes various data sources
R.20	Dashboard and advanced visualisation that contains advanced analytics, model predictions, sensor/device health and scenario analysis
R.21	Dashboard and advanced visualisation containing targeted data on KPIs related to the softening process such as water quality, economics, total quantities of chemical dosing, projections, energy consumption.
R.22	Dashboard and visualisation to view operations and maintenance data.
R.23	Reduce silos and communication (seeing the same data, common language - single source of truth).
R.24	Responsibilities need to be clear.
R.25	Corporate commitment to maintaining and evolving the Digital Twin platform to take into account changes in best practice, changing data landscape etc.

To depict the intricate relationships between the different entities encompassed within the functional requirements, a reference Digital Twin architecture has been designed, as illustrated in Figure 15. Particular attention was given to harmonizing this conceptualized architecture with the current implementation and future aspirations of DWG's data architectural landscape. In the physical entity layer, data sources, as identified by stakeholders, are clearly defined. The data collected from the treatment processes necessitates networked connections with the associated PLC/SCADA systems. Given the critical status of the treatment installation infrastructure, accessing such data involves additional security measures, facilitated through a Demilitarized Zone (DMZ) positioned in the Communication layer. All acquired data finds its place within a Data Lake and Data Warehouses in the Informational layer. Here, the need for reliable and validated data is effectively addressed through automated validation and data reconciliation processes. Specific data is made available to the models deployed for the Digital Twin. Within the Application layer, this data, along with the models, is integrated into application services and the outcomes are relayed to specialized dashboards tailored to meet the unique needs of specific stakeholders.

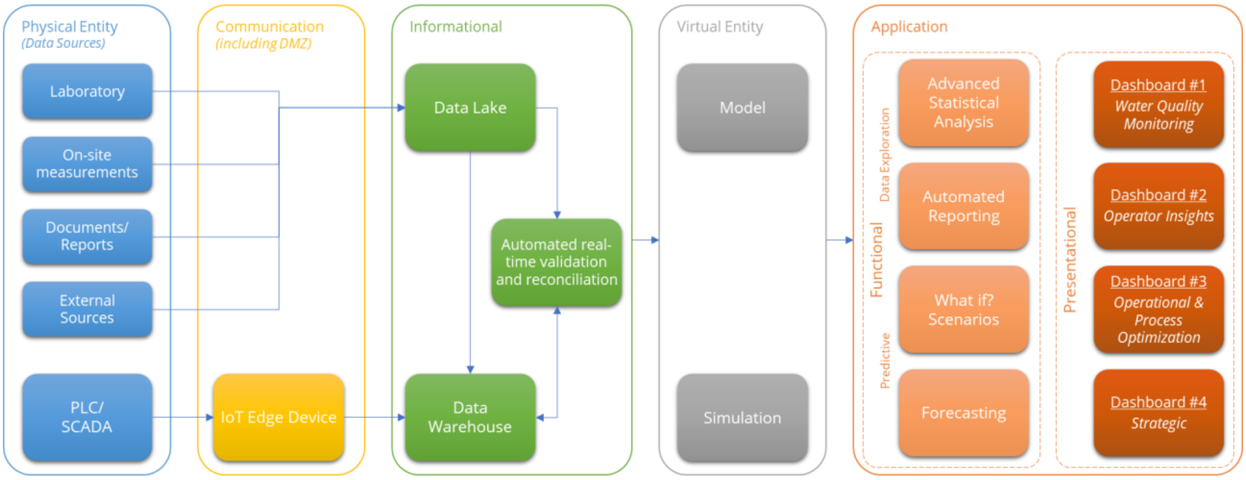


Figure 15 Digital Twin reference architecture based on functional requirements

4.4.6 Design Principles

The design of the Digital Twin application services yields significant benefits when guided by designated design principles. These principles are strongly related to the DTGs and the listed requirements. By following the stipulated guidelines, the Digital Twin will maintain consistency and be easy to use, reducing communication barriers and ensuring a cohesive user experience. Furthermore, the Digital Twin will be easily scalable allowing for its evolution to include additional data, processes and functionalities. The Digital Twin can also be interoperable to integrate with other systems and finally, to be optimised for high performance. Based on the formulated DTGs and requirements, the following Design Principles (DP) have been defined for the softening treatment process Digital Twin:

DP.1 **FAIR, Secure, Efficient & Available Data**

Reinforce the ethical, responsible and efficient use of FAIR data (Findable, Accessible, Interoperable and Reusable). Assure that the data held by an organization is appropriately described by metadata ensuring its timely and accurate retrieval and encouraging its seamless reuse. Provide retrieval mechanisms are documented and publicly available. Ensure that data pertaining to individuals is handled responsibly and ethically by organizations and businesses. Eradicate data silos by presenting the data as a “single source of truth” for the whole of the organization.

DP.2 **Robust Process Monitoring, Optimisation, Forecasting & Understanding**

Emphasise the need for a reliable and real-time monitoring system to assess the performance and behaviour of the softening treatment process. Refining and optimising the process based on collected data, simulations and insights to enhance resource efficiency and water quality. Provide capabilities to anticipate future conditions and potential extreme events while increasing the in-depth understanding of the treatment process, enabling stakeholders to make informed decisions and to address operational and future challenges effectively.

DP.3 **Tailored and Interactive End-User Experience**

Provide tailored user interfaces, features and data displays that are customised based on the stakeholder group requirements. Include interactive elements and features such as real-time data visualisation and user-friendly dashboards. Ensuring stakeholders are presented with relevant, high-quality data aligning to their specific roles and requirements to eradicate communication silos, enable user-adoption and enhance decision-making.

DP.4 **Accountable, Purposeful and Promoting Organisation**

Prioritise the need for accountability and purpose-driven actions within the organisation’s Digital Twin implementation. Align the Digital Twin’s objectives to the strategic goals of the organisation which fosters purposeful development of the system contributing to improved decision-making, efficient operations and streamlined processes. Promote the adoption and future adaptability of the Digital Twin by fostering a culture of constant evolution and embracing emerging technologies to ensure the system remains at the forefront of innovation.

4.4.7 Digital Twin Application Services

While considering the requirements, specific application services have been identified for inclusion in the design of the Digital Twin for the softening process. These applications aim to provide tangible solutions and services to key stakeholders, enhancing their efficiency, support informed decision-making at an operational, tactical and strategic level and contributing to organisation goals related to efficiency, cost optimisation and sustainability. It must be noted, that the identified services primarily reside in the Application and Presentational layers (southbound layers) of the architecture in Figure 15 – layers where maximum interaction with end-users occurs. However, successful deployment of the Digital Twin also depends on the integration and technology deployment within the Physical Entity, Communication and Informational layers. Based on discussions with relevant stakeholders, it is known that ongoing efforts including IT and OT convergence and data management and storage solutions address these needs, particularly regarding fulfilling specific requirements such as R.1-R.5, R.10 and R.23.

Below, the identified Digital Twin Application Services (DT.AS) and the associated requirements that each fulfil are provided.

***DT.AS.1* Automated Real-time Data Validation & Reconciliation**

This service has been envisioned to be part of the Information layer and closely interacting with a Data Lake and Data Warehouse. Data quality compromises various factors, including sensor faults, calibration issues, fouling, connectivity problems between sensors, actuators and the data management system. Additionally, manual data curation and quality assurance is a prohibitively time-consuming and expensive task. Therefore, there is a need to conduct automated data quality control where data packed in batches or in real-time are screened for erroneous measurements. This can be achieved through statistical and rule-based anomaly detection methods to flag erroneous values. Decisions can then be made to reconcile the anomalies, either by replacing the value using interpolation, previous values or an alternative value through model predictions provided by other application services. This service fulfils the requirements R.6 and R.11.

***DT.AS.2* Advanced Statistical Analysis**

This service focusses on harnessing the available data to derive deeper insights, enabling informed decision-making and maximising the value of data resources by transforming data into actionable information, thus facilitating knowledge creation. This can include performing statistical summaries, KPI calculations and advanced statistical analysis such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and clustering techniques to increase understanding of the process. The deployment has been envisioned as a network of microservices that involves the breaking down of the different statistical computations and analyses into small, independent and reusable components that can be deployed, managed and scaled independently. This can be conducted using technologies such as Docker Containers, Kubernetes and Azure Container Instances. In production, the microservices take input data and perform the necessary analysis as defined by the user based on their requirements. This is then relayed to the other services within the Functional and Presentational Layers and can be considered to be part of the goal of returning better, more detailed information about the operation of the process to the operator and other stakeholders. This application service fulfils the requirements R.5, R.13, R.14 and R.18.

***DT.AS.3* Automated Reporting**

The automated reporting service leverages the single-source-of-truth concept to facilitate dissemination of the current system state along with related historical information. The reporting can take the form of basic statistical summaries such as recording historical trends alongside current KPI values. Reports can be automatically compiled into documents for wider dissemination within the organization and stored within a Document Management System or automatically distributed through subscription email or other corporate communication platforms. This service complements the data available through the dashboarding facilities. This application service fulfils the requirements R.1, R.3, R.5 and R.23.

***DT.AS.4* Process Model to Perform What-if? Scenarios**
This application service functions as a computational tool that facilitates the simulation of hypothetical scenarios and evaluates their potential impact on the virtual softening treatment process. Its functionalities can include process optimisation, assessment of chemical dosing strategies and scenarios and validates measured parameters, such as pH, with the model predictions, to assess the health and reliability of sensor devices. Additionally, data and information on unobserved and unmeasured parameters can also be accessed that increases process understanding. By conducting what-if simulations, stakeholders can gain valuable insights into how process changes may affect operational efficiency and water quality, ultimately enabling more informed decision-making and strategy development. This application service can potentially be realised by operationalising the existing offline PhreeqC model, initially as set up by the process engineering department for the softening process. By connecting it with real-time data, it offers a quick-win solution for evaluating its potential in achieving the DTGs. However, it's essential to thoroughly evaluate the model and its deployment requirements, comparing them with alternative modelling options to ensure the most suitable solution. This application service fulfils the requirements R.7-R.9, R.11-R.13, R.17 and R.18.

***DT.AS.5* Forecasting**
A forecasting service for the softening treatment process entails the development of predictive models. Such models can benefit from the use of data-driven technologies using historical data for the training of the models and when deployed, connected with real-time data during production. This service can be used to anticipate and predict variations in the quality of incoming raw water quality, process parameters, water quality attributes in the reactor, operational conditions and cost and economic predictions. Moreover, predictive and proactive maintenance of key physical assets such as sensor devices can be conducted where early detection of anomalies and deviations promote efficient operations. Such a capability can complement and enhance existing maintenance programs. The forecasting service can be considered to offer forecasts across various time horizons, addressing both operational and tactical needs. For operational decisions, insights provided in the order of days, allows for adjustments during operations and maintenance. On a tactical level, forecasts made in the order of months, facilitates long-term planning and resource allocation such as economic decisions and chemical stockpiling. The forecasting service can become critical to address current challenges related to water availability, particularly as water sources may transition from groundwater to alternative sources such as surface water, where quality is more variable and seasonal. The application service fulfils the requirement R.7-R.9, R.11 and R.14-R.16.

***DT.AS.6* Dashboarding and Advanced Visualisation**
This service serves as the principal interface between the Digital Twin and a multitude of stakeholders. Furthermore, a primary requirement stated by stakeholders, it offers the possibility of having access to relevant data in an understandable format, that supports data-driven decisions-making and visualisation. The provision of standardised access to the data and information through this application service plays a crucial role in eradicating communication and organisation silos, to ensure better collaboration among entities within the organisation. Notably, the Digital Twin Roadmap considers dashboards (Simple Dashboard and Multi-source Dashboard with Advanced Analytics) as pivotal milestones for an organisation throughout the Digital Twin's development and deployment. The following dashboards have been identified as integral part of this visualisation application service:

- *DT.AS.6.1 Dashboard #1 – Multi-location & multi-data source water quality monitoring:* Visualising water quality data related to the groundwater and the influent and effluent of the softening treatment process for the purpose of monitoring and reporting. This dashboard is most relevant for the stakeholder – water quality control

- ***DT.AS.6.2 Dashboard #2 – Operator Insights:*** Visualisation of key operations and maintenance data such as historical and scheduled maintenance records, chemical stockpiling, alarms, spare-parts inventory and more. This dashboard is most relevant for the stakeholder – operators.
- ***DT.AS.6.3 Dashboard #3 – Operational & Process Optimisation:*** Visualisation and interactive dashboard containing monitoring data, advanced analytics, model predictions, operational forecasts, what-if scenarios, sensor/device health, KPIs on economics and water quality. This dashboard is most relevant for the stakeholder – process engineer.
- ***DT.AS.6.4 Dashboard #4 – Strategic:*** Containing visualisation of measured and calculated data specifically curated in order to provide strategic information on the KPIs related to the softening process such as water quality, economics, total amount of chemical dosing, projections and energy consumption. This dashboard is most relevant for the stakeholders – energy and environmental coordinator, direction of innovations.

4.4.8 Digital Twin Outcomes and Motivation Stack Example

The Digital Twin Outcomes (DTO), as discussed in Section 4.4.2, play a pivotal role in determining the value proposition of the Digital Twin, both in its design phase and when it is actively used by stakeholders. These outcomes represent the tangible results and benefits that the Digital Twin delivers. They can provide both qualitative and quantitative indicators, which, when consolidated, creates a value proposition that guides subsequent decision-making by the core team.

During the design phase, the DTOs can be defined from two sides. The DTOs can be envisioned or projected based on the intended DTGs and requirements. These DTOs can guide the development of the application services to achieve the specific goals. In contrast, the design and choice of technology used for the application services, can be assessed to what DTOs they provide. These acquired DTOs can then be cross-examined with the defined DTGs to assess whether this iteration and choices for the design meets the ambitions and needs of the stakeholders. A combination of these DTOs will aid in the value proposition creation. Once the Digital Twin is deployed and utilised by stakeholders, real-world outcomes emerge as a result of its functionality. These outcomes are the actual benefits experienced by users and the organization. They might include improved process performance, cost savings, enhanced water quality, or other operational advantages. The assessment of these outcomes is essential for verifying the value proposition of the Digital Twin and evaluating whether it aligns with the initial projections made in the design phase.

In Figure 16 below, an example illustrates the evaluation process during the design of the Digital Twin. This process combines the various motivation elements discussed in the previous sub-sections, DTOs and value propositions for a specific use case focused on optimizing the chemical dosage in the softening treatment process. This illustration is referred to as the *Motivation Stack*. It demonstrates the bidirectional information flow from both a top-down and bottom-up perspective. At the top of the Motivation Stack, relevant stakeholders for optimizing chemical dosage are defined. Based on their Drivers & Concerns, specific Digital Twin Goals have been established. Additionally, the Drivers & Concerns inform the Requirements for the Digital Twin. Application Services are identified to meet these requirements while adhering to the Design Principles. Based on the anticipated performance of the services during the design phase, the application services yield certain Digital Twin Outcomes, which in turn, contribute to the formulation of the value proposition. This value proposition directly informs the Programme Manager of Innovation, who is a core-team member. For illustration purposes, sample outcomes are provided. Anticipated outcomes include a reduction in chemical dosing demand due to process optimisation. However, an increase in operational costs due to chemical dosing is projected due to the adoption of higher-quality chemicals. This, in turn, indirectly results in a lower carbon footprint, as a lower chemical volume is required. These interventions aim to maintain the hardness level in the produced water within a certain range of French degrees.

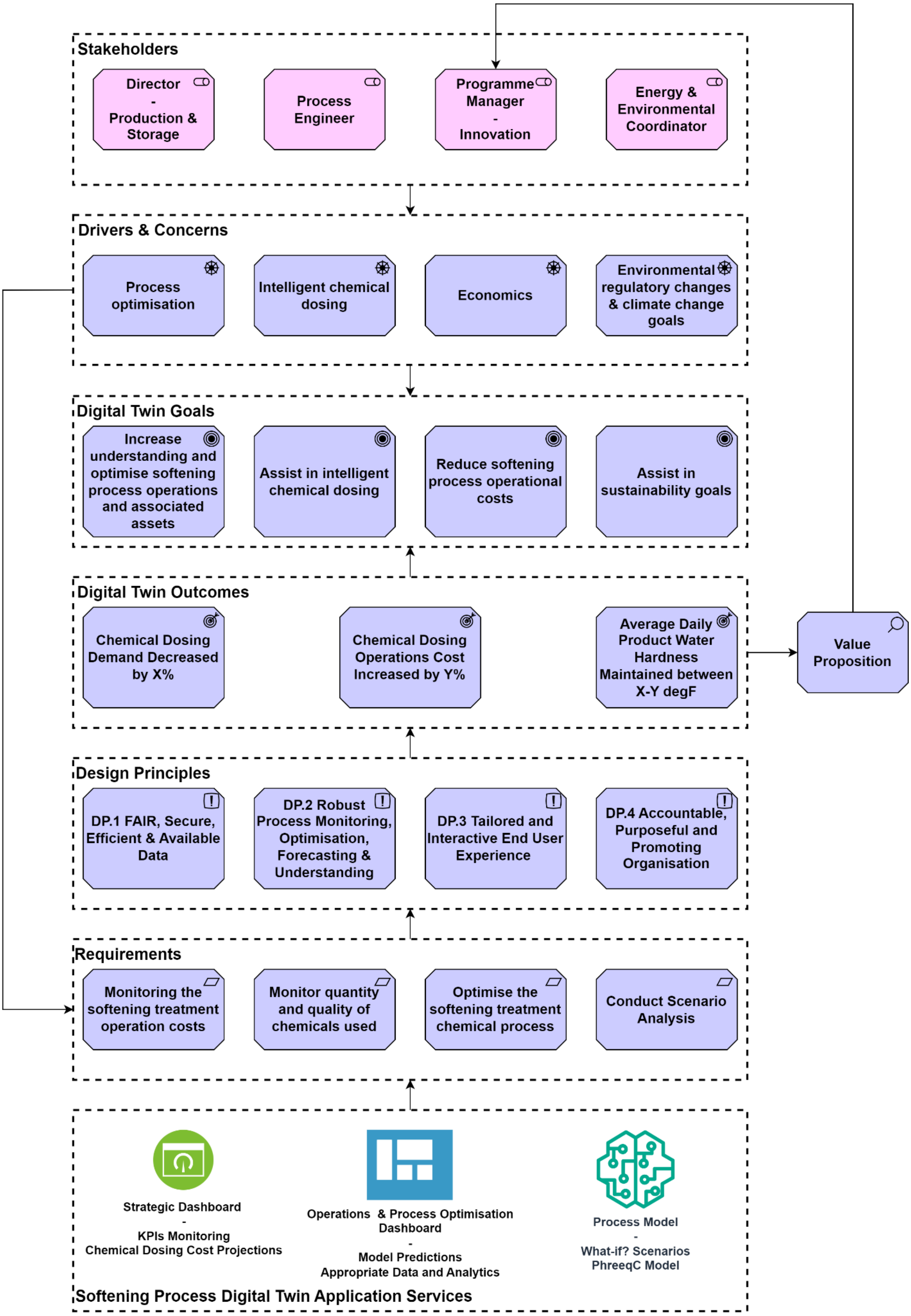


Figure 16 Motivation Stack example illustrating the evaluation process for optimising chemical dosing in the softening treatment process

4.4.9 Assessment of Roadmap Placement for Digital Twin

Finally, to strategically plan the development and deployment of the Digital Twin, it is essential to determine its position in the Roadmap (Figure 8). The analysis of the stakeholders' responses revealed that DWG is currently within the region of *Step 1 – Core team selection* and *Milestone #1 – Simple Dashboard*. In this functional design, various elements of the roadmap have been detailed for the softening treatment case study and identified concrete application services based on key stakeholder requirements and organisational objectives. These outcomes serve as a strong foundation for progressing through the roadmap to reach the agreed-upon endpoint, which, in the case of the Digital Twin for the softening treatment process, is *Milestone #3 – Decision Support Twin*. Table 12 provides a plan for the development and deployment of the application services aligned with the Roadmap Milestones.

Table 12 Development and deployment planning of Digital Twin Application Services (DT.AS) aligned with Digital Twin Roadmap milestones

Roadmap Milestone	Digital Twin Application Services (DT.AS) Maturity Level
Milestone #1: Simple Dashboard	*DT.AS.2 – Proof-of-Concept* Statistical analysis to be conducted has been defined. Initial development and offline tests conducted. *DT.AS.3 – Proof-of-Concept* Reporting formats defined and tests conducted in a testing environment. User-feedback received in an acceptance environment. *DT.AS.6 – Proof-of-Concept* Mock-up layouts for dashboards defined, relevant data sources identified and deployment of a simple release of the dashboards with connection to 1 data source. Collection of user-feedback to inform improvements.
Milestone #2: Multi-Source Dashboard with Advanced Analytics	*DT.AS.1 – Proof-Concept* Methods in data validation and reconciliation tested and validated. Service deployed in a testing environment. *DT.AS.2 – Production ready* Deployed as a network of microservices. Container instances' API endpoints defined and exposed to receive input. Connections with DT.AS.6 service to provide output to end-users. *DT.AS.3 – Production ready* Service deployed and connected with Informational layer, DT.AS.2 and DT.AS.6. *DT.AS.4 – Proof-of-concept* Model concept/type finalised for process model. Training or calibration finalised and testing/validation conducted with historical data. Initial release of best performing model in acceptance environment and integration with real-time data to assess model performance with new data and scenarios. Connections with DT.AS.6 to access and visualise predictions. *DT.AS.5 – Proof-of-concept* Data or target variables prioritised to be considered for forecasting. Input variables or features identified through feature engineering and feature importance. Training of data-driven models to test forecasting capabilities. Deployment of an instance of the model and integration with real-time data. Connections with DT.AS.6 to access and visualise forecasts.

Roadmap Milestone	Digital Twin Application Services (DT.AS) Maturity Level
	DT.AS.6 – Production ready All dashboards deployed and fully operational, with seamless integration of data from multiple sources and the outputs of advanced statistical analysis microservices. Relevant stakeholders have access to the dashboards, which are actively utilised in their day-to-day tasks and operations.
Milestone #3: Decision Support Twin	*DT.AS.1 – Production ready* Data validation and reconciliation service integrated with Data Lake and Data Warehouse and performing automated and real-time data quality control. *DT.AS.4 – Production ready* Process model refinement based on performance data. Protocols for recalibration established. Service enabled to perform scenario analysis. Model fully integrated within Digital Twin, providing crucial information on softening treatment processes to inform operational and long-term strategic decision-making. *DT.AS.5 – Production ready* Forecasting model re-training and refinement based on performance data. Serverless functions deployed to ensure constant updating of model with new training data. Models fully integrated within Digital Twin, providing crucial information on softening treatment processes to inform operational and long-term strategic decision-making.
Milestone #4: Control Twin	Not pursued in the implementation of the Digital Twin for DWG's softening treatment process.

4.5 Targeted Actions and Recommendation

The functional design has provided a comprehensive framework and analytical methodology to understand stakeholder motivations, identify Digital Twin Goals, formulate requirements and method to understand the value proposition for the development and deployment of a Digital Twin and its application services for the softening treatment processes. Provided below are targeted actions and recommendations that can be considered to further expand this analysis and considerations for the Digital Twin design:

- The development of use cases and user stories can be a suitable way to further expand the relations between the Digital Twin Goals, identified requirements and the defined application services. This can also be a meaningful method to communicate back to the stakeholders on how such solutions can help fulfil their requirements.

- It is recommended that *Motivation Stacks* are defined for all defined use cases, as shown in the example provided in Figure 16, to support this process.

- If necessary, certain requirements can be decomposed to more detailed requirements. For example, a requirement stated as *Optimise the softening treatment chemical process* can be further broken to *optimise chemical dosing*, *optimise removal of pellets*, and more. It is recommended that based on the assessment of each requirement, a brainstorming or working session is organised with key stakeholders/experts (in the provided example, that would be the PLC programmer and process technologists) that can detail requirements which directly inform the Digital Twin Application Services. Furthermore, it is recommended to ensure accountability of each requirement to ensure that the developed applications cover all identified requirements assigned to it.

- It is recommended that workshops, in the form of multi-stakeholder forums, are conducted to discuss which data and specific parameters are of importance for the various application services. This will ensure further the fulfilment of certain requirements related to data availability, data interpretability and human-in-the-loop.

- An addition to the current functional design method is the consideration of constraints, with respect to the current system as well as from a technological perspective. Information on the constraints can be an important factor during the design of the Digital Twin application services.

- It is recommended that at every Milestone of the roadmap (or during the management buy-in step), a review of the Digital Twin Goals and Design Principles are conducted to ensure they remain up-to-date. These elements can be updated or upgraded based on what the current state of the system and what was achieved in the previous development cycle.

- Stakeholder feedback highlighted a lack of alignment on the definition of a Digital Twin and the digitalisation strategy. It is recommended that such concepts are clarified and regularly communicated to the stakeholders. This can also be a topic for discussion in a multi-stakeholder forum.

- A potential follow-up activity is to identify implemented digital twins for the softening treatment process by other water companies or utilities and conduct a comparative study to assess the requirements, outcomes and value propositions of this study in relation to their approaches and implementations.

5 Conclusions and Recommendations

5.1 Conclusions

This report outlines a systematic roadmap for implementing Digital Twin technology in drinking water treatment processes, enabling real-time water quality monitoring. Digital Twins, representing virtual models of real systems, mimic physical processes using actual data inputs, offering insights into water quality parameters and enhancing operational efficiency. Furthermore, it addresses the need for data-driven decision-making in water companies while retaining operator knowledge. Digital Twins serve as virtual representations, facilitating this transition. The roadmap not only focuses on technical integration but also emphasizes organizational aspects, including data governance, skill development and structural changes.

The study results in a comprehensive literature review, shared understanding of Digital Twin concepts and a detailed roadmap specific to water treatment processes. It covers defining the purpose, technical integration, organizational recommendations and guidelines for adopting standards and platforms. These outputs empower stakeholders with knowledge essential for successful Digital Twin implementation.

Stakeholders, including process technologists, ICT professionals, data engineers and management professionals are demonstrated to play vital roles in the inception, implementation and adoption of a Digital Twin as well as collectively benefiting from the results. Process technologists gain from real-time insights, while data engineers can simplify analyses using Digital Twins. Collaboration among domain experts is seen to be crucial highlighting the interdisciplinary nature of Digital Twin projects.

The project involved gathering scientific knowledge, exploring existing implementations, identifying relevant KPIs, assessing data infrastructure and developing a detailed roadmap. Stakeholder engagement informed the roadmap development, leading to the creation of a functional design for a case study on water softening. The design includes selecting a core team, identifying a value proposition and clarifying design principles and requirements, illustrating practical application steps.

Overall, the project's outcomes provide a holistic approach to Digital Twin implementation in water treatment processes, bridging the gap between technical integration and organizational readiness. Moreover, it was observed that a roadmap, offering strategic direction and vision, can be effectively applied and operationalized using the developed methodology within the functional design. This activity served to confirm that the concepts outlined in the roadmap are indeed, valid and essential in the pursuit of creating and integrating Digital Twins.

5.2 Recommendations

The following recommendations and potential follow-up activities have been envisioned:

- Validation of the roadmap by key operational and management level stakeholders from the water companies
- Advancing the results of the current functional design for De Watergroep’s softening process to pursue the design and implementation of a Digital Twin, striving to reach Milestone 1 of the roadmap.
- Testing the roadmap further by conducting a more detailed functional design of Digital Twin requirements for other drinking water treatment processes.

6 Bibliography

Abba, S. I., Pham, Q. B., Saini, G., Linh, N. T. T., Ahmed, A. N., Mohajane, M., Khaledian, M., Abdulkadir, R. A., & Bach, Q. V. (2020). Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environmental Science and Pollution Research*, **27**(33), *pp41524–41539*. *https://doi.org/10.1007/s11356-020-09689-x*

Adamenko, D., Kunnen, S., Pluhnau, R., Loibl, A. & Nagarajah, A. 2020. Review and comparison of the methods of designing the Digital Twin. *Proceedings 30th CIRP Design 2020.* Procedia CIRP 91 (2020). *pp27-32*. *https://doi.org/10.1016/j.procir.2020.02.146*

Akinmolayan, F; (2017) *Mathematical modelling of clean water treatment works*. Doctoral thesis , UCL (University College London) https://discovery.ucl.ac.uk/id/eprint/1553176/

Aliashrafi, A., Zhang, Y., Groenewegen, H., & Peleato, N. M. (2021). A review of data-driven modelling in drinking water treatment. *Reviews in Environmental Science and Biotechnology*, **20**(4), *pp985–1009*. *https://doi.org/10.1007/s11157-021-09592-y*

AlSawaftah, N., Abuwatfa, W., Darwish, N., & Husseini, G. (2021). A Comprehensive Review on Membrane Fouling: Mathematical Modelling, Prediction, Diagnosis and Mitigation. *Water*, **13**(9), 1327. MDPI AG. Retrieved from http://dx.doi.org/10.3390/w13091327

Alzamora, F.M., Conejos, P., Castro-Gama, M. & Vertommen, I. (2021). *Digital Twins - A new paradigm for water supply and distribution networks*. IAHR hydrolink. **2**. *pp49-54. https://www.iahr.org/library/infor?pid=10798 [accessed 17-10-2023]*

Armstrong, M. M. (2020). *Cheat sheet: What is Digital Twin. IMB. Here's your quick guide to digital twins, what they are and why they matter for your organization*. *https://www.ibm.com/blog/iot-cheat-sheet-digital-twin/ [accessed 17-10-2023]*

Aspen Technology. (2020). *Hybrid Modeling: AI and Domain Expertise Combine to Optimize Assets* (White paper). https://www.aspentech.com/en/resources/white-papers/hybrid-modeling-ai-and-domain-expertise-combine-to-optimize-assets *[accessed 17-10-2023]*

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2019). Enforcing Analytic Constraints in Neural-Networks Emulating Physical Systems. https://doi.org/10.1103/PhysRevLett.126.098302

Castro-Gama, M., Vertommen, I., Bouziotas, D., & Béen, F. (2020). *Using mobile phone data in digital twins for drinking water distribution networks.* BTO Report **2020**.29, KWR Water Research Institute, Nieuwegein, Netherlands, *43pp*.

Chaux, J.D., Sanchez-Londono, D. & Barbieri, G.A. 2021. Digital Twin Architecture to Optimi2ze Productivity within Controlled Environment Agriculture. *Applied Science*. **11**, *8875*. *https://doi.org/10.3390/app11198875*

Chang, E. E., Yang, S. Y., Huang, C. P., Liang, C. H., & Chiang, P. C. (2011). Assessing the fouling mechanisms of high-pressure nanofiltration membrane using the modified Hermia model and the resistance-in-series model. *Separation and Purification Technology*, 79(3), 329-336. https://doi.org/10.1016/j.seppur.2011.03.017.

Chen, K. L., Song, L., Ong, S. L., & Ng, W. J. (2004). The development of membrane fouling in full-scale RO processes. *Journal of Membrane Science*, 232(1–2), 63-72. https://doi.org/10.1016/j.memsci.2003.11.028.

Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., & Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research*, **171**, 115454. *https://doi.org/10.1016/j.watres.2019.115454*

Cho, K. H., Sthiannopkao, S., Pachepsky, Y. A., Kim, K. W., & Kim, J. H. (2011). Prediction of contamination potential of groundwater arsenic in Cambodia, Laos and Thailand using artificial neural network. *Water Research*, **45**(17), *pp5535-5544*. *https://doi.org/10.1016/j.watres.2011.08.010*

Corbatón-Báguena, M. J., Vincent-Vela, M. C., Gozálvez-Zafrilla, J. M., Álvarez-Blanco, S., Lora-García, J., & Catalán-Martínez, D. (2016). Comparison between artificial neural networks and Hermia's models to assess ultrafiltration performance. *Separation and Purification Technology*, 170, 434-444. https://doi.org/10.1016/j.seppur.2016.07.007

Curl, J. M., Nading, T., Hegger, K., Barhoumi, A., & Smoczynski, M. (2019). Digital Twins: The Next Generation of Water Treatment Technology. *Journal - American Water Works Association*, **111**(12), *pp44–50*. *https://doi.org/10.1002/awwa.1413*

Delgrange, N., Cabassud, C., Cabassud, M., Durand-Bourlier, L., & Lainé, J. M. (1998). Neural networks for prediction of ultrafiltration transmembrane pressure – application to drinking water production. Journal of Membrane *Science*, **150**(1), *pp111-123*. https://doi.org/10.1016/S0376-7388(98)00217-8

Delgrange-Vincent, N., Cabassud, C., Cabassud, M., Durand-Bourlier, L., & Laîné, J. M. (2000). Neural networks for long-term prediction of fouling and backwash efficiency in ultrafiltration for drinking water production. *Desalination*, **131**(1–3), *pp353-362*. https://doi.org/10.1016/S0011-9164(00)90034-1

Donnelly, J., Daneshkhah, A. and Abolfathi, S. (2023). Physics-Informed Neural Networks for Statistical Emulation of Hydrodynamical Numerical Models, EGU General Assembly 2023, Vienna, Austria, 23–28 Apr 2023, EGU23-5445, https://doi.org/10.5194/egusphere-egu23-5445

Duclos-Orsello, C., Li, W., & Ho, C. (2006). A three mechanism model to describe fouling of microfiltration membranes. *Journal of Membrane Science*, **280**(1–2), *pp856-866.* https://doi.org/10.1016/j.memsci.2006.03.005

Fujii, T.Y., Hayashi, V.T., Arakaki, R., Ruggiero, W.V., Bulla, R. Jr., Hayashi, F.H. & Khalil, K.A. 2022. A Digital Twin Architecture Model Applied with MLOps Techniques to Improve Short-Term Energy Consumption Prediction. *Machines*, **10**(23). *https://doi.org/10.3390/machines10010023*

Griffiths, K. A., & Andrews, R. C. (2011). The application of artificial neural networks for the optimization of coagulant dosage. *Water Science and Technology: Water Supply*, **11**(5), *pp605–611*. *https://doi.org/10.2166/ws.2011.028*

Hvala, N., & Kocijan, J. (2020). Design of a hybrid mechanistic/Gaussian process model to predict full-scale wastewater treatment plant effluent. *Computers & Chemical Engineering*, **140**, 106934. https://doi.org/10.1016/J.COMPCHEMENG.2020.106934

Jeldres, R. I., Fawell, P. D., & Florio, B. J. (2018). Population balance modelling to describe the particle aggregation process: A review. *Powder Technology*, **326**, *pp190-207*. https://doi.org/10.1016/j.powtec.2017.12.033.

Jia, X., Willard, J., Karpatne, A., Read, J. S., Zwart, J. A., Steinbach, M., & Kumar, V. (2021). Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. *ACM/IMS Transactions on Data Science*, **2**(3), *pp1–26*. https://doi.org/10.1145/3447814

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, **29**, *pp36-52*. *https://doi.org/https://doi.org/10.1016/j.cirpj.2020.02.002*

Juntunen, P., Liukkonen, M., Lehtola, M., & Hiltunen, Y. (2013). Cluster analysis by self-organizing maps: An application to the modelling of water quality in a treatment process. *Applied Soft Computing Journal*, **13**(7), *pp3191–3196*. *https://doi.org/10.1016/j.asoc.2013.01.027*

Kennedy, M. J., Gandomi, A. H., & Miller, C. M. (2015). Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal. *Journal of Environmental Chemical Engineering*, **3**(4), *pp2829–2838*. *https://doi.org/10.1016/J.JECE.2015.10.010*

Keskitalo, J., & Leiviskä, K. (2014). AISC 322 - Artificial Neural Network Ensembles in Hybrid Modelling of Activated Sludge Plant. 322. https://doi.org/10.1007/978-3-319-11313-5_60

Kockmann, N. (2019). "Digital methods and tools for chemical equipment and plants." *Reaction Chemistry & Engineering,* **4**(9): *pp1522-1529*. *https://doi.org/10.1039/C9RE00017H*

Li, L., Rong, S., Wang, R., & Yu, S. (2021). Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chemical Engineering Journal*, **405**(June 2020), 126673. *https://doi.org/10.1016/j.cej.2020.126673*

Lim, K. Y. H., Zheng, P., Chen, C. H., & Huang, L. (2020). A digital twin-enhanced system for engineering product family design and optimization. *Journal of Manufacturing Systems*, **57**, *pp82-93.* *https://doi.org/10.1016/J.JMSY.2020.08.011*

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies and industrial applications. *Journal of Manufacturing Systems*, **58**, *pp346-361*. *https://doi.org/https://doi.org/10.1016/j.jmsy.2020.06.017*

Maier, H. R., Morgan, N., & Chow, C. W. K. (2004). Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling and Software*, **19**(5), *pp485–494*. *https://doi.org/10.1016/S1364-8152(03)00163-4*

Matheri, A. N., Mohamed, B., Ntuli, F., Nabadda, E., & Ngila, J. C. (2022). Sustainable circularity and intelligent data-driven operations and control of the wastewater treatment plant. *Physics and Chemistry of the Earth*, **126**(November 2021), 103152. *https://doi.org/10.1016/j.pce.2022.103152*

Meierhofer, J., Schweiger, L., Lu, J., Züst, S., West, S., Stoll, O. & Kiritsis, D. 2021. Digital Twin-Enabled Decision Support Services in Industrial Ecosystems. *Applied Science*. **11**, *11418*. *https://doi.org/10.3390/app112311418*

Mondal, S., & De, S. (2010). A fouling model for steady-state crossflow membrane filtration considering sequential intermediate pore blocking and cake formation. Separation and Purification Technology, 75(2), 222-228. https://doi.org/10.1016/j.seppur.2010.07.016.

Negri, E., Fumagalli, L., & Macchi, M. (2017). A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, **11**, *pp939-948*. *https://doi.org/10.1016/J.PROMFG.2017.07.198*

Nwogu, C., Lugaresi, G., Anagnostou, A., Matta, A. & Taylor, S.J.E. 2022. Towards a Requirement-driven Digital Twin Architecture. *Proceedings 55th CIRP Conference on Manufacturing Systems*. Procedia CIRP **107**(2022). *pp758-763*. *https://doi.org/10.1016/j.procir.2022.05.058*

Patriarca, R., Simone, F., & Di Gravio, G. (2022). Modelling cyber resilience in a water treatment and distribution system. *Reliability Engineering and System Safety*, **226**(January), 108653. *https://doi.org/10.1016/j.ress.2022.108653*

Park, S., Baek, S. S., Pyo, J., Pachepsky, Y., Park, J., & Cho, K. H. (2019). Deep neural networks for modeling fouling growth and flux decline during NF/RO membrane filtration. *Journal of Membrane Science*, *587*, 117164. https://doi.org/10.1016/j.memsci.2019.06.004

Perno, M., Hvam, L., & Haug, A. (2022). Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. *Computers in Industry*, **134**, 103558-103558. *https://doi.org/10.1016/J.COMPIND.2021.103558*

Pesantez, J. E., Alghamdi, F., Sabu, S., Mahinthakumar, G., & Berglund, E. Z. (2022). Using a digital twin to explore water infrastructure impacts during the COVID-19 pandemic. *Sustainable Cities and Society*, **77**. *https://doi.org/10.1016/j.scs.2021.103520*

Piron, E., Latrille, E., & René, F. (1997). Application of artificial neural networks for crossflow microfiltration modeling: "black-box" and semi-physical approaches. *Computers & Chemical Engineering*, **21**(9), *pp1021-1030*. *https://doi.org/10.1016/S0098-1354(96)00332-8*

Ruzsa, C. (2021). Digital twin technology - external data resources in creating the model and classification of different digital twin types in manufacturing. *Procedia Manufacturing*, **54**, *pp209-215*. *https://doi.org/10.1016/j.promfg.2021.07.032*

van Schagen, K.M. , Babuška, R., Rietveld, L.C., Baars, E.T. Optimal flow distribution over multiple parallel pellet reactors: a model-based approach. *Water Science & Technology,* 1 February 2006; **53**(4-5): *pp493–501*. doi: https://doi.org/10.2166/wst.2006.160

Scheifele, C., Verl, A., & Riedel, O. (2019). Real-time co-simulation for the virtual commissioning of production systems. *Procedia CIRP*, **79**, *pp397-402*. *https://doi.org/10.1016/J.PROCIR.2019.02.104*

Shetty, G. R., Malki, H., & Chellam, S. (2003). Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks. *Journal of Membrane Science*, **212**(1–2), *pp99-112*. https://doi.org/10.1016/S0376-7388(02)00473-8

Steindl, G., Stagl, M., Kasper, L., Kastner, W. & Hofmann, R. (2020). Generic Digital Twin Architecture for Industrial Energy Systems. *Applied Sciences*, **2020**, *10*, 8903. *https://doi.org/10.3390/app10248903*

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems.* **27**. *https://doi.org/10.48550/arXiv.1409.3215*

Teller, M. (2021). Legal aspects related to digital twin. *Philosophical Transactions of the Royal Society A*, **379**(2207), 20210023-20210023. *https://doi.org/10.1098/rsta.2021.0023*

Tao, F., Zhang, M. & Nee, A. 2019. Five-Dimension Digital Twin Modelling and Its Key Technologies. In: Tao, F., Zhang, M. & Nee, A. *[eds] Digital Twin Driven Smart Manufacturing,* Academic Press, London, UK. *pp63–81*.

The Open Group. (2018). The Open Group Architecture Framework (TOGAF) 9.2: An Enterprise Architecture Methodology and Framework. https://pubs.opengroup.org/architecture/togaf9-doc/arch/

Therrien, J. D., Nicolaï, N., & Vanrolleghem, P. A. (2020). A critical review of the data pipeline: How wastewater system operation flows from data to intelligence. Water Science and Technology, **82**(12), *pp2613–2634*. *https://doi.org/10.2166/wst.2020.393*

Wan ZY, Vlachas P, Koumoutsakos P, Sapsis T (2018) Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PLoS ONE,* **13**(5): e0197704. https://doi.org/10.1371/journal.pone.0197704

Wei, Y., Law, A. W., Yang, C., & Tang, D. (2022). Combined Anomaly Detection Framework for Digital Twins of Water Treatment Facilities. *Water*, **14**(7), 1001. *https://doi.org/10.3390/w14071001*

Appendix I. Interview Questions for the Leading Water Companies

Interviews – Leading (Water) Companies in the Development and Deployment of Digital Twins

Within the BTO Digital Twin Zuivering project, interviews are needed to be conducted with leading (water) companies to gather information on how such a company have envisioned, developed and potentially deployed digital twins within their technical processes or organisation in general. The purpose is to learn from their experience, what was needed during such a transition, what all tools are currently being used and more importantly, what were the major limitations or barriers that affected their execution of such a strategy. Such learnings can then be consolidated by the project team and used in the development of the roadmap along with reporting the results to the broader drinking water sector.

The following document should be considered as a guide that can be utilised during the interviews. Questions have been listed in order that can be followed while conducting the interview and space has been provided to record the answers. The questions have been also been grouped within overarching topics.

Interviewers:

Interviewee(s) Details:

Name:

Position:

Organisation/Company:

Contact Details:

Date:

Location:

1,2,6,7,8

Pilot

I. What is a Digital Twin?

1. What is your or your organisation's definition of a Digital Twin?
2. What functions should the Digital Twin be capable of conducting?

Organisational/Strategic Level

3. What are the main drivers currently for your (water) company to perform a digital transition and for the adoption of Digital Twins? Are there organisational challenges, technological challenges, environmental challenges or customer-related challenges that has driven your company to look into Digital Twins?
   Examples to stimulate a discussion/response:

   Organisational Challenges: Digitising decades of knowledge that are currently in possession within an ageing workforce, increasing efficiency withing organisation, standardising data and information to promote easier sharing among departments.

   Customer-Related Challenges: Streamlining and automating handling of customer complaints, (near) real-time sharing of data/information relevant to customer consumption

   Technological Challenges: Better monitoring of process performance, converting water treatment systems that are *data-rich* to *information-rich*, optimisation of system to better meet targets, plant-wide control

   Environmental Challenges: Tackling climate change related challenges, climate resilience of water treatment assets, future-proofing of water treatment assets.

4. Does your (water) company have a digital strategy to support the digital transition and promote the use of Digital Twins? If yes, could an explanation be provided considering – how many years does the current digital strategy run for; which part of the business is the digital strategy developed for; who is responsible for the implementation of the digital strategy; which product owners have been designated for the implementation?
5. Were there specific milestones defined within the existing digital strategy, or in general, in the development and deployment of the Digital Twins? If so, what were they? Which have been achieved and how long did it take? How many are remaining to be achieved?
   Examples of Milestones that could be mentioned to stimulate discussion/response:

   Milestone 0 – There have been internal conversations on the subject, for the past 1 (or more) year

   Milestone 1 – Proposals and strategic documents have been provided to the management requesting for budget allocation.

   Milestone 2 – Increased deployment of sensors and instrumentation to further increase data availability.

   Milestone 3 – Research projects on developing proof-of-concept data analytics tools, Digital Twins (AI or process models) for given use cases within the drinking water treatment assets.

   Milestone 4 – Further training, evaluation and finalisation of Digital Twin solutions.

   Milestone 5 – Retrofitting or upgrading of IT infrastructure to support the deployment of Digital Twins.

   Milestone 6 – Deployment of Digital Twins into legacy system providing (near) real-time output

6. Which Key Performance Indicators (KPIs) were considered important when developing the Digital Twin?
7. What is the (if any) main obstacle(s) that hamper the deployment of and transformation to a digital twin?
   Depending on the answer please proceed to the correct section (for example):

Challenges with environment, sensors, data management, IT support- Go to section II, III and IV

Challenges with modelling- Go to section III, IV

Challenges with data visualisation, access to data and control of data- Go to section V

Challenges with organizational acceptance, trust of DT and legal regulations- Go to section VI and VIII.

II. IT Infrastructure/Architecture Related

8. Have the Digital Twins been deployed within the legacy system of the company? Or are they currently running in other servers (Virtual Machines for example)?
9. Could a short walkthrough/explanation be provided on how the data pipeline and communication between the various components within the IT infrastructure has been currently setup? From raw data being collected by sensors, to making predictions using the Digital Twin and relaying the outputs into visualisation tools.
10. Was there a need to upgrade or retrofit the current IT infrastructure to support the deployment of Digital Twins? If so, was it conducted internally or sub-contracted to a company?
11. Has the Digital Twin been deployed within a paid cloud computing service, such as Azure? Or are open-source platforms being used and if so, why (or why not)?
12. Are models and other pertinent steps (such as data validation, pre-processing of inputs and outputs, etc.) being containerised (such as the use of Dockers) for the deployment?
13. How is data security being handled with respect to the interaction between the data storage, Digital Twin and end-users within your company? Was the legacy system sufficiently equipped to deal with data security related issues or were there need for upgrades? If upgrades were made, what were the main points of limitation that were prevalent?

III. Data Collection and Sensor Deployment (15 minutes)

14. Which process-based, asset-based and maintenance-based parameters are measured using online sensors? (To be answered as a general overview, not looking to get a list of all parameters).
15. Are there key or 'nice-to-measure' parameters still needed to be measured online? Are there plans to further expand the deployment of sensor and increase the volume of data?
16. From a financial perspective, what is the typical investment needed to be able to deploy new sensors to measure all relevant and key process parameters within a treatment plant? (If a non-water company, question should be tailored to their process/what the digital twin is for)
17. What is the typical time period needed to be receiving reliable data from the sensors? Consider the following activities as an example - acquisition of the new sensors, deployment, testing, calibration, making IoT connections to the legacy systems, storage in the data warehouse.

IV. Models Trained/Calibrated with the Digital Twin

18. What type of process(es) within the treatment or other system is (are) being modelled as part of the Digital Twin? And why these processes?

19. Are first principle models/mechanistic/phenomenological models being calibrated and used? If so, why? And if not, why have they not been considered? What limitations or challenges led you to decide to not use them?
20. Are Artificial Intelligence (AI) models being trained and used? If so, what type of AI models- ML models, neural networks, deep learning based, etc.? Why and what benefits of AI models led you to use such tools? If not used, why not and what challenges within the process or organisation led you to not consider AI models?
21. Have hybrid models been considered (or used) as a Digital Twin? Hybrid models can be loosely defined as models that can combine process-based and data-driven models in a manner that allows for more accurate predictions of given processes by adequately benefiting from the advantages existing within the two schools of modelling.
22. How often are the models being used re-calibrated or updated? Is this being done live or 'on the fly'? If not, is it feasible for live updating? What would be the current barriers or limitations of the models deployed that prevents live updating?
23. Are modelling software being used for the digital twin? Are they paid or open-source? If no software is used, is Python or (another programming language) the main foundation for the development and deployment of the models?

V. Dashboarding and Advanced Visualisation

24. How are the outputs/results from the Digital Twin visualised currently? Have they been included into the legacy system based dashboards? If so, what are the software tools used in the legacy – PowerBi, Tableau, etc.?
25. If no for legacy dashboards in Q21, can the new dashboards developed be described? Are open-source based solutions (such as Grafana) being considered? Are these containerised for the deployment?
26. How is access to the dashboards managed? Are all employees able to see the Digital Twin outputs or only a handful? What type of employees (process operators, process technologists, management, etc.) have been given access?
27. Can control-based updates (such as setpoints of specific assets with the treatment process) be made from the dashboard, based on the outputs of the Digital Twin? As a result, is a form of predictive control being utilised of considered?
28. How important is the dashboard to reach the goal of the Digital Twin?

VI. Perception of the Digital Twin within the Organisation

29. How do process operators value the Digital Twin? Do operators value their work the same, more or less due to the Digital Twin? Is their experience still valued?
30. Were specific interactions (such as workshops) and events organised to provide transparent information to process operators, thereby potentially promoting them embracing (the use of) Digital Twins? How was this handled or was there even a need for it?
31. Were the operators incorporated and made involved in the initial planning for the development of the Digital Twins?
32. Do the colleagues, especially operators, trust the Digital Twin and the outputs it provides? How is it managed?
33. Was any training provided to the operators for the implementation and use of the Digital Twin?

34. How important is the commitment of management to ensure the smooth execution and use of the Digital Twin? And why? How do you foresee their commitment to be in the coming future?

VII. Legality and Regulations

35. Is the Digital Twin used for gathering data or modelling of specific processes for the purpose of reporting and ensuring government compliance?
36. Are (or were) there any legal barriers/issues that impacted the development and deployment of the Digital Twin?

VIII. Digital Twin perception

In the last 10 minutes of the interview discuss the following:

37. Does your digital twin live up to the expectations of your organisation?
38. What are the benefits of the digital twin? How do you qualify these benefits?
39. Is there anything that you would have done differently in your implementation plan for the digital twin?

Relevant Points Made During Discussion:

Appendix II. Roadmap Risk Assessments

Table 13 Risk Assessment for Organization path

Step of Roadmap	ID	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Select multidisciplinary coordination Team	38	Lack of resources	ALL	**4** Likely	**4** Major	**16**	Extreme *(15-25)*
	39	Lack of knowlege	ALL	**2** Unlikely	**2** Minor	**4**	Moderate *(4-6)*
Select users with digital mindset	40	Lack of knowledge	ALL	**3** Possible	**4** Major	**12**	High *(8-12)*
Define what problem DT/Dashboard will solve	41	Lack of knowledge	ALL	**4** Possible	**5** Major	**13**	High *(8-12)*
User Feedback/ training	42	Outdated operational procedures	Operators, supervisors	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
Show value to organization	43	Lack of stakeholder engagement	All	**3** Possible	**1** Negligible	**3**	Low *(1-3)*
Policy Change/ Role of DT in organization in decision making	44	Lack of operator buy-in	ALL	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
	45	Lack of acceptance and incapability to switch to a new way of working	Operators, supervisors	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Evaluate legal ramifications of decision making	46	Ethics of digital twin making decisions. Liability	ALL	**4** Likely	**4** Major	**12**	High *(8-12)*
Convince management	47	Results do not prove necessity of management	All	**3** Possible	**5** Catastrophic	**15**	Extreme *(15-25)*

Table 14 Risk Assessment for Model Development path

Step of Roadmap	ID	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Selection Process	1	Lack of problem definition	ALL	**3** Possible	**4** Major	**12**	High *(8-12)*
	2	Selection of inappropriate task	ALL	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
Select DT Goal	3	Unclear understanding of DT levels	Developers, Modellers, Managers	**2** Unlikely	**2** Minor	**4**	Moderate *(4-6)*
Choose model type (AI, hybrid, mechanistic)	4	Unclear understanding of necessary models	Developers, Modellers	**4** Likely	**4** Major	**16**	Extreme *(15-25)*
Offline model development	5	Incorrect/insufficient data	Developers, Modellers	**5** Almost Certain	**4** Major	**20**	Extreme *(15-25)*
	6	Lack of modellers	Developers, Modellers	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
	7	Cost of modelling	Developers, Modellers, Management	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
	8	Lack of sensors	Developers, Modellers, Operators	**4** Likely	**3** Moderate	**12**	High *(8-12)*
	9	run time is too long	Developers, Modellers, IT	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Live model sensor fusion	10	Setting up required infrastructure is deemed too costly	IT, management	**2** Unlikely	**3** Moderate	**6**	Moderate *(4-6)*
	11	Update interval sensor is too low	developers, modellers data expert	**2** Unlikely	**3** Moderate	**6**	Moderate *(4-6)*
Improve model	12	There has been no proper development pipeline implemented which makes deploying of new versions hard	developers, modellers data expert IT	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Automatic retrain/recalibrate model	13	data validation is not done properly (yet) before stored in the data warehouse		**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
Develop model based control	14	Management buy-in	ALL	**2** Unlikely	**3** Moderate	**6**	Moderate *(4-6)*
	15	insufficient model speed	developer, modellers, IT, process technologist	**3** Possible	**2** Minor	**6**	Moderate (4-6)

Step of Roadmap	ID	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
	16	insufficient model accuracy	developer, modellers, IT, process technologist	**3** Possible	**4** Major	**12**	High (8-12)
	17	lack of knowledge (modellers are not process control experts)		**3** Possible	**3** Moderate	**9**	High (8-12)
Deploy model in operation	18	Management buy in	All	**3** Possible	**4** Major	**12**	High (8-12)

Table 15 Risk Assessment for Data Management and Architecture path

Step of Roadmap	ID	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Capable ICT Team	19	Lack of knowledge	ALL	**3** Possible	**2** Minor	**6**	Moderate *(4-6)*
	20	Lack of Resources	ALL	**4** Likely	**4** Major	**16**	Extreme *(15-25)*
Standard Data formats	21	Lack of architecture	Developers, Modellers, Operators	3 Possible	2 Minor	**6**	Moderate *(4-6)*
	22	Lack of investment in data management plan	Developers, Modellers, Operators, Management	**4** Likely	**4** Major	**16**	Extreme *(15-25)*
Sensor & asset data in Data lake	23	Lack of investment in data management plan	data engineer	**2** Unlikely	**3** Moderate	**6**	Moderate *(4-6)*
Context broker: Real time integration of models and data	24	lack of knowledge (how to set it up)	IT, data management, process engineers	**2** Unlikely	**3** Moderate	**6**	Moderate *(4-6)*
Model outputs (soft sensors)	25	Data errors creating incorrect soft sensor data		**4** Likely	**3** Moderate	**12**	High *(8-12)*
Deploy missing sensors	26	Inability to purchase parts	ALL	**5** Almost Certain	**5** Catastrophic	**25**	Extreme *(15-25)*
Automatic data validation and reconciliation	27	Lack of investment in proper infrastructure (context broker)	ALL	**2** Unlikely	**4** Major	**8**	High *(8-12)*
	28	Inaccurate models for reconciliation	data engineer	**2** Unlikely	**4** Major	**8**	High *(8-12)*
Corporate data warehouse	29	Lack of investment in infrastructure (databases and servers/cloud)	IT data manager, management	**4** Likely	**5** Catastrophic	**20**	Extreme *(15-25)*

Table 16 Risk Assessment of Security path

Step of Roadmap	ID	Risk Description	Resources Impacted	Probability Level	Impact Level	Risk Value	Overall Risk Assessment
Policy	30	Lack of knowledge	ALL	**3** Possible	**1** Negligible	**3**	Low *(1-3)*
	31	Lack of management buyiin	ALL	**3** Possible	**3** Moderate	**9**	High *(8-12)*
	32	Outdated policy	ALL	**3** Possible	**1** Negligible	**3**	Low *(1-3)*
Contraints	33	Policy creating issues on allowability of specific technology	ALL	**3** Possible	**4** Major	**12**	High *(8-12)*
Transport	34	Cyber-security of data	ALL	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Encryption	35	Cyber-security of data	ALL	**3** Possible	**3** Moderate	**9**	High *(8-12)*
Smart monitoring of access	36						
Air gapped communication	37	not able to ensure air gapped communication (e.g. ensure only a machine on premise can change setpoints)		**3** Possible	**5** Catastrophic	**15**	Extreme *(15-25)*