Archetypes of Digital Twins

Hendrik van der Valk · Hendrik Haße · Frederik Möller · Boris Otto

Abstract Currently, Digital Twins receive considerable attention from practitioners and in research. A Digital Twin describes a concept that connects physical and virtual objects through a data linkage. However, Digital Twins are highly dependent on their individual use case, which leads to a plethora of Digital Twin configurations. Based on a thorough literature analysis and two interview series with experts from various electrical and mechanical engineering companies, this paper proposes a set of archetypes of Digital Twins for individual use cases. It delimits the Digital Twins from related concepts, e.g., Digital Threads. The paper delivers profound insights into the domain of Digital Twins and, thus, helps the reader to identify the different archetypical patterns.

Keywords Digital Twin · Archetypes · Dimensions · Characteristics · Cluster analysis

1 Introduction

Digital Twins are gaining much attention in research and with practitioners (Detecon Consulting 2019; PwC 2020), which also becomes apparent in a steep rise in academic publications (Scopus 2020). Yet, there is still ambiguity regarding the term's precise definition (Cimino et al. 2019). Instead, a spectrum of *termini* has emerged, including *Digital Shadows*, *Digital Threads*, or *Digital Models* (Helu et al. 2017; Schuh et al. 2019; Urbina Coronado et al. 2018).

Each term is defined differently and often used synonymously to *Digital Twin*, further blurring the concept. Our research addresses precisely that issue, as it proposes a clear distinction of types of Digital Twins. More specifically, we suggest foundational and distinguishable types, so-called archetypes, based on a taxonomical analysis and morphological characteristics. The significant advantage of that approach is that we can differentiate Digital Twins based on morphological characteristics and representative patterns which we can use to distinguish archetypes. To achieve that goal, our study pursues a mixed-method design.

First, we developed initial characteristics and archetypes purely based on findings from a structured literature review, carried out following the recommendations of Webster and Watson (2002) and Vom Brocke et al. (2009). Next, we validated, extended, and triangulated our findings by means of a qualitative interview study. In the study, we collected data from 15 industry experts from various industries in semi-structured interviews. The interviews brought to light rich findings of high relevance for the field of Digital Twin design and archetypical patterns. For example, when asked, "*what do you think a Digital Twin looks like in general*", the overwhelming majority answered: "*A Digital Twin is the digital picture of a physical asset*". The statement's uniformity somewhat clouds the more detailed specifics distinguishing one instance of a Digital Twin from another. Illustratively, points of differentiation include whether the Digital Twin

H. van der Valk (✉) · F. Möller · B. Otto
Chair for Industrial Information Management IIM, TU
Dortmund University, Joseph-von-Fraunhofer-Straße 2-4,
44227 Dortmund, Germany
e-mail: Hendrik.van-der-valk@tu-dortmund.de

H. Haße · F. Möller · B. Otto
Fraunhofer Institute for Software and Systems Engineering
ISST, Emil-Figge-Straße 91, 44227 Dortmund, Germany

is a basis for information (e.g., data repository) or a real-time representation of an asset's physical state. Interestingly, the understanding and definition of Digital Twins varied across the study's participants according to their industrial sector. Those findings further added to our motivation to harmonize the understanding of Digital Twins and standardize it to some degree.

For that purpose, we develop archetypes that differentiate understandings of Digital Twins. Archetypes are a form of standardization (Thesaurus 2020). They originate from Greek and describe original patterns that are "[...] a primordial image, character, or pattern of circumstances that recurs throughout literature and thought consistently enough to be considered a universal concept or situation" (Encyclopedia Britannica 2011, p.1). Given the usefulness of the archetypical differentiation of Digital Twins, as per the argumentation given above, we derive the following research questions (RQ):

RQ1   What clusters of Digital Twins can be derived from the literature corpus?

RQ2   What are the design characteristics of Digital Twins in research and practice?

RQ3   What are the archetypes of Digital Twins?

The paper is structured as follows: After a brief overview of the definitions of Digital Twins, their usages, and origins, we will describe our research methods, i.e., the *structured literature review*, the *development of a* *taxonomy*, the *qualitative research approaches*, and the method of *deriving archetypical patterns*. After this, we introduce the different dimensions and characteristics of Digital Twins, followed by the derivation of the archetypes. The archetypes are discussed and evaluated before we provide the conclusion, limitations, and a brief outlook on future research.

2 Digital Twin Types

The concept of twins is a well-known and widely used technology in classical manufacturing processes since NASA used physical twins as a copy of space vehicles in the Apollo project (Rosen et al. 2015). Since then, the concept has been approached from various points of view, which has lead to different types of Digital Twins. The different definitional approaches become apparent when drawing from the literature corpus (see Table 1). Yet, uniformity in Digital Twin definitions is still lacking. There is only a vague understanding of the concept of Digital Twins (Haag and Anderl 2019). For that matter, Wagner et al. (2019) state that the definition of Digital Twins highly depends on the individual use case. Evidently, using the Digital Twin concept in healthcare use cases (e.g., see Rivera et al. (2019) for the use case *digital patients*) requires a different set of specialized characteristics than a Digital Twin in manufacturing (e.g., see Kritzinger et al. (2018)). Yet, both use cases might have shared underlying

Table 1 Commonly used definitions of digital twins

#	A Digital Twin…:	Author
1	…is "an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin. The Digital Twin is ultra-realistic […] integrates sensor data […] maintenance history and all available historical and fleet data obtained"	Glaessgen and Stargel (2012, p.7)
2	…consists of "a) physical products in Real Space, b) virtual products in Virtual Space, and c) the connections of data and information that ties the virtual and real products" and is "a set of virtual information constructs […] captured from actual sensor data, current, past actual, and future predicted"	Grieves (2014, pp. 1–4); Grieves and Vickers (2017, pp. 94–95)
3	…is seen as the next step in the development of simulation. It supports the simulation as a core functionality along the entire life cycle, e.g., supporting operation and service with direct linkage to operation data	Rosen et al. (2015)
4	…contains "three parts: physical product, virtual product, and connected data that tie the physical and virtual product". Furthermore, they demand a real-time data flow	Tao et al. (2018, p. 3566; 2019)
5	… is a concept, in which "the data flows between an existing physical object and a digital object are fully integrated in both directions […] the digital object might also act as controlling instance of the physical object. There might also be other objects, physical or digital, which induce changes of state in the digital object. A change in state of the physical object directly leads to a change in state of the digital object and vice versa"	Kritzinger et al. (2018, p. 1017)
6	…"has a one-to-one correspondence to its real-world counterpart, enabling product-centric information management" and additionally, a Digital Twin is "a modular entity"	Autiosalo et al. (2020, p. 1204)

technology and characteristics, for instance, *real-time updates*. The composition of the interview partners, therefore, reflects our desired use case independence. Each interviewee described a Digital Twin that was specific to the individual use case. For example, interviewee 1 described a use case in which the Digital Twin was used as a tool to gather, check, and maintain master data. The use case presented by interviewee 2 was an application of a Digital Twin in a warehouse to heighten the transparency and to analyze and improve the processes inside the warehouse. Further use cases included the utilization of Digital Twins in production environments (e.g., interviews 4, 11, 12, or 15), the application of Digital Twins to monitor products over their life cycle (e.g., interviews 10 or 13), the usage in the healthcare sector (e.g., interviews 16 and 17), and in the supply chain management (interview 18).

Based on our literature review and findings (van der Valk et al. 2020), Table 1 shows the most relevant definitions. From these, we synthesize a working definition that guides our research.

The definitions above show a baseline of understanding of a Digital Twin, which leads us to the preliminary definition of a Digital Twin which was used in the analysis of the literature and interviews as a working definition:

Definition 1 The Digital Twin is a virtual construct that represents a physical counterpart, integrates several data inputs with the aim of data handling and processing, and provides a bi-directional data linkage between the virtual world and the physical one. Synchronization is crucial to the Digital Twin in order to display any changes in the state of the physical object.

At this point, we have to stress the fact that more recent reviews brought the kind of data linkage into focus. Kritzinger et al. (2018) proposed that a Digital Twin should contain an automatic data linkage. This approach is backed by several reviews, e.g., the review of Errandonea et al. (2020). Fuller (2020) comes to a similar conclusion in his review. Nevertheless, he also describes the discrepancy between what is called a Digital Twin and what is a Digital Twin per the definition of Kritzinger et al. (2018). We also noticed that many so-called Digital Twins do not provide an automatic data linkage and, therefore, should not be labeled as a Digital Twin. However, at this point we will include descriptions of so-called Digital Twins which do not mandate an automatic data flow to gain deeper insights into this discrepancy.

Digital Twins possess many overlaps with other digitization technologies, and also a variety of synonymously used terms is noticeable. For example, similarities exist with the concepts of Digital Models (Urbina Coronado et al. 2018), Digital Shadows (Kritzinger et al. 2018), and

Digital Threads (Helu et al. 2017). In the following, we aim to provide definitions of these concepts and to stress the differences to a Digital Twin.

Definition 2 A Digital Model poses as a virtual representation of a physical product that may contain a data linkage between both (Kritzinger et al. 2018). However, this linkage is manually at best. The Digital Model will not replicate a change of state of the physical object.

As a representation of the physical object as well as of the changes of any state are crucial to a Digital Twin, the Digital Model cannot be seen as a Digital Twin. Furthermore, a Digital Model lacks the opportunity to handle and process any kind of data. We see a Digital Model as part of a Digital Twin in the sense that it provides the virtual picture of the physical object. Furthermore, a Digital Model does not provide a bi-directional data linkage per se.

Definition 3 The Digital Shadow provides highly accurate representations of processes with the aim to create a real-time picture based on the relevant data (Bauernhansl et al. 2016).

Digital Shadows do not possess automatic bi-directional data links (Kritzinger et al. 2018). Furthermore, an internal data processing is not seen as mandatory for a Digital Shadow (Schuh et al. 2019). Hence, the Digital Shadow is a digital construct on the way to a Digital Twin, but not an actual twin.

Definition 4 Digital Threads connect various data sources along the life cycle of a product and enable a data linkage between physical assets and software products, but do not further process the data (Helu et al. 2017).

The main differences between Digital Threads and Digital Twins are the twins' ability to process the data instead of just gathering it and that a Digital Twin shall represent a physical product. Data feedback, i.e., a bi-directional data flow, is not mandatory for a Digital Thread. Hence, the term Digital Threads may not be used synonymously to Digital Twins.

As related works we have to emphasize the work of Enders and Hoßbach (2019), who developed a taxonomy of different Digital Twin applications, Josifovska et al. (2019), who created a framework for Digital Twins in cyber-physical systems, and as the most recent Jones et al. (2020), who conducted a literature review and detected research gaps regarding Digital Twins. To a certain extent, the different reviews show a convergence, e.g., when portraying the automatic data flows. However, these studies focus on a narrow branch or specialized fields, like cyber-physical systems or manufacturing contexts. Hence, we aim to provide a more general view on Digital Twins which is independent of branches and use cases but contributes a

broader classification of integral parts of Digital Twins. Therefore, it allows the classification of different, domain-independent Digital Twin types.

3 Mixed-Method Design

The paper aims to engineer theoretical and practical descriptions of Digital Twins in reverse, which were collected through a structured literature review and an interview study. We aim to synthesize comprehensive Digital Twin archetypes based on the literature and a qualitative interview study with industry professionals. Due to the large-scale research objective, our research design is a combinatory approach subsuming multiple, mixed methods. Figure 1 illustrates our research process, consisting of two qualitative (literature review, qualitative interview study) and one quantitative (cluster analysis) sections that are organized in action steps (Action 1–Action 7). Greene et al. (1989, p. 256) define mixed-method research processes as those "that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words)". The mixed-method approach has the clear advantage of triangulating results by using multiple data sources instead of just one. In our case, these data sources are the literature corpus on Digital Twins and practitioners from different industries. Summarizing the above, our research approach includes the following steps (see Fig. 1):

Action 1: Exhaustive literature review following Webster and Watson (2002) and Vom Brocke et al. (2009). **Action 2:** Development of a taxonomy for Digital Twins based on van der Valk et al. (2020). **Action 3:** Cluster analysis of the underlying data of the taxonomy. **Action 4:** Qualitative interview study with 15 industry experts to

triangulate the findings. **Action 5:** Cluster analysis with industry experts. **Action 6:** Synthesis of the qualitative and quantitative results. **Action 7:** Evaluation through a second interview series and finalization of archetypes. Figure 1 graphically illustrates the seven-step research process while simultaneously indicating its qualitative and quantitative parts.

3.1 Structured Literature Review

The literature review uses the method of Vom Brocke et al. (2009). Vom Brocke et al. (2009) recommend a five steps when conduction a literature review: first, the definition of the review scope, second, the conceptualization of the topic, third, the actual search process, fourth, the analysis of the literature, and, lastly, the revision of the research agenda. Accordingly, in the first step, we defined the literature review's scope to consider only peer-reviewed publications dealing with the topic of Digital Twins. Because the literature about Digital Twins is growing exponentially (Scopus 2020), we limited the research scope to the scientific databases AIS eLibrary, ACM Digital, IEEE Xplore, Science Direct, and JSTOR. By selecting these databases, we cover the research in the fields of information systems and engineering. In the second step, we conceptualized the topic of the literature review. To do so, we especially searched for definitions of Digital Twins. As Cimino et al. (2019) highlight, there is a wide variety of definitions for Digital Twins. However, we could identify specific definitions, often used in many publications (see Table 1).

In step three, we searched the databases for publications with the search string *"Digital Twin"*. In total, we found 579 publications which contain the term Digital Twin in some context. During the analysis (step 4), we applied several filtering mechanisms, which we drew from Cooper (1988) and Vom Brocke et al. (2009) and which focus on relevance, accessibility, and removal of duplicates. The filtering mechanisms are threefold; first, we consider the relevance mechanism, meaning that the publication must explicitly deal with the Digital Twin. Therefore, we eliminated every paper which does not mention Digital Twins at least in the title, the abstract, or keywords. The second filtering mechanism eliminated all papers that were not accessible. As many publications are published in multiple databases, the results contained a not-neglectable number of duplicates, which we eliminated. Lastly, the third filtering mechanism is quality-related. The publications have to include a comprehensive argumentation, consistent use of established research methodologies and must deal with Digital Twins in a non-trivial fashion (Levy and Ellis 2006). To ensure that papers adhere to the quality criteria mentioned above, we analyzed each paper in-depth. All in
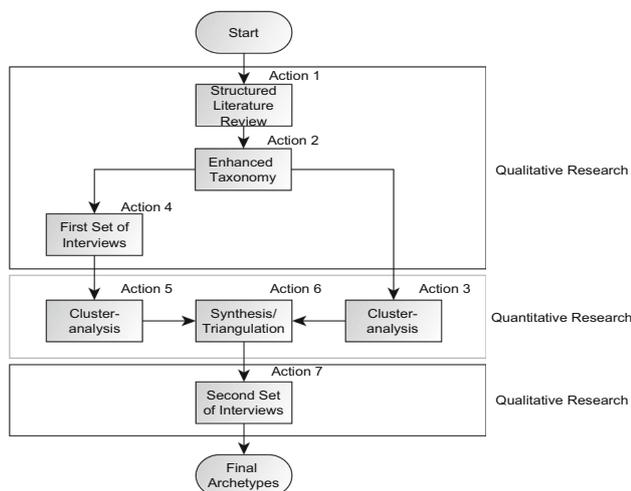
Fig. 1 Research process

all, very few papers did not meet our inclusion requirements. Following Webster and Watson (2002), we added relevant papers cited in the literature corpus through a backward search. Finally, we identified 233 publications as appropriate for our research purpose. Figure 2 shows the distribution of the publications about Digital Twins during the time frame from 2012 to 2020. One can notice the exponential growth of the literature, with more than 80% of the total papers on Digital Twins published in the last two years. That steep rise paints an illustrative picture of the importance and conception of Digital Twins in academia today. The quantitative distribution of papers about Digital Twins (see Fig. 2) weighs towards the recent time period. As each description of a Digital Twin, i.e., each paper, has the same weight and, therefore, the same impact in our analysis, the more recent papers gain a bigger influence in this research due to their amount outweighing the older papers. Hence, we see the influences of the development of Digital Twins over time as considered in the cluster analysis.

3.2 Taxonomy Design

Starting from the literature base created during action 1 (see Fig. 1), we developed a taxonomy of Digital Twins. We applied the method of Nickerson et al. (2013), which has emerged as the de facto standard for taxonomy design in Information Systems (Oberländer et al. 2019; Szopinski et al. 2019). The methodology helps to create the taxonomy comprehensively and transparently (Oberländer et al. 2019). In general, a taxonomy can classify and structure a given field of interest (Glass and Vessey 1995). As a taxonomy enables an empirical structuring of the area of interest, we preferred using a taxonomy over a conceptualization via typologies or ontologies (Bailey 1994).

For the development of the taxonomy and the definition of the taxonomy´s purpose, one has to determine the *meta-characteristic* **(Step 1)**, define *ending conditions* **(Step 2)**, and choose the *empirical-to-conceptual* or the *conceptual-to-empirical* approach **(Step 3)**. This decision
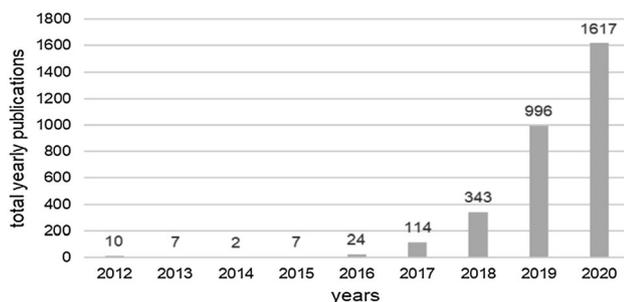
Fig. 2 Yearly publications about digital twins from 2012 to 2020 (Scopus 2020)

predetermines the steps 4e to 6e, or 4c to 6c, respectively. The conceptual-to-empirical approach requires to define characteristics a priori **(Step 4c)** before the analysis of objects **(Steps 5c and 6c)**. The empirical-to-conceptual approach starts with studying a subset of objects **(Step 4e)** and crystallizing characteristics from their comparative analysis **(Steps 5e and 6e)**. Both approaches are iteratively and mutually pursued for as long as the ending conditions have not been reached **(Step 7)**. Once they are fulfilled, the taxonomy has reached its final state. Figure 3 shows the four iterations we conducted during the research process. In the 1st, 2nd, and 3rd iteration, we followed the empirical-to-conceptual way and only analyzed the literature corps (see van der Valk et al. (2020)). During the 4th iteration, we proceeded with the conceptual-to-empirical approach and only analyzed the interview manuscripts. As we met all ending conditions after the 4th iteration, we stopped developing the taxonomy and continued with the cluster analysis.

3.3 Qualitative Research Design

To evaluate the literature base's conceptual insights, we expanded our research by menas of a qualitative study with expert interviews. In total, we conducted 18 interviews in two interview series (see Table 2). The qualitative research follows the approach from Sarker and Sarker (2009). First the interviewees were identified. In line with the 'known sponsor approach' (Patton 2002), we got in touch with the interviewees through senior personal contacts. In preparation for the interview, we provided each interviewee with a brief overview of the research project. The study consists of 18 interviews with industry experts with diverse backgrounds and from different industries. The interviews followed a semi-structured approach, as we only prescribed the superordinate areas of the questions (Myers and Newman 2007; Patton 2002). The research guide includes questions about the general, individual understanding and definition of the interviewees regarding Digital Twins. Mirroring the literature-based taxonomy, we presented each interviewee with the taxonomy of van der Valk et al. (2020). Each interviewee could add or dismiss dimensions or characteristics. Furthermore, we asked the participants which characteristics would be part of their individual configuration of a Digital Twin. This approach allows for a discussion between the interviewer and the interviewee while ensuring comparability between the personal interviews (Merton and Kendall 1946; Myers and Newman 2007; Patton 2002). Each interview was recorded and transcribed. After the first interview series, we analyzed the interviews' transcriptions and coded them accordingly to the Grounded Theory Methodology. The transcripts provide profound access to the full information potential and
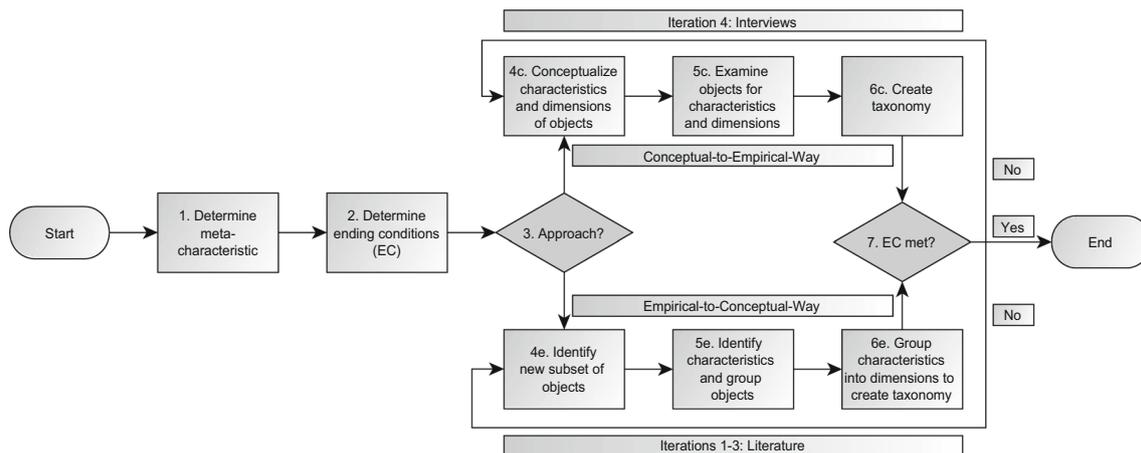
Fig. 3 Creating a taxonomy following Nickerson et al. (2013)

Table 2 Interview partners by sector and research action

#	Position	Sector	Research action
1	Head of Master Data Management	Mechanical Engineering	(4)
2	Global Innovation Manager	Logistics	(4)
3	Enterprise Architect	Electrical Engineering	(4)
4	Technology Director	Electrical Engineering	(4)
5	Enterprise Architect	Electrical Engineering	(4)
6	Production Management	Mechanical Engineering	(4)
7	Senior Research Scientist	Electrical Engineering	(4)
8	Enterprise Architect	Telecommunication	(4)
9	Organizational Developer	Network Operator	(4)
10	Product Line Manager	Electrical Engineering	(4)
11	Quality Manager	Electrical Engineering	(4)
12	Head of Production Technology	Mechanical Engineering	(4)
13	Innovation Manager	Logistics	(4)
14	Product Manager	Electrical Engineering	(4)
15	Head of Data Management	Chemical Engineering	(4)
16	Product and Market Manager	Healthcare	(7)
17	Head of Research and Design	Healthcare	(7)
18	Head of Strategic Management	Logistics	(7)

are the first step towards a thorough analysis (Lapadat and Lindsay 1999; Ochs 1979). Following the recommendations of Iivari et al. (2020), we had a second round of interviews with a smaller set of experts to validate our findings.

3.4 Cluster Analysis

Archetypes are a "typical example of a certain […] thing" (Oxford Dictionary 2020, p. 1) and have emerged as purposeful results in Information Systems (e.g., see Möller et al. (2019) or Weking et al. (2018)). The literal Greek translation for archetype is "first-molded as a pattern" (Liddell et al. 1940), which we aim to achieve in this paper.

Cluster analysis organizes patterns into clusters (Jain et al. 1999). We try to sort the patterns along the structure given by the taxonomy of Digital Twins. For the cluster analysis, we choose the statistical language *R*, using the *daisy* function (to identify dissimilarities between data sets in the data matrix), the Gower measurement coefficient, and the library *cluster* to analyze and visualize the data (Gower 1971; Maechler et al. 2019). For the clusters' partition, we used the *k-means* algorithm, which is the most popular hierarchical algorithm (Jain 2010). The algorithm "finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized" (Jain 2010, p. 653). Therefore, in an iterative process, the algorithm sorts the data points into clusters

that contain the minimum error. To define the appropriate number of clusters, we used the *elbow-method*. We triangulated the preliminary results using a mixed-methods approach (Greene et al. 1989) so that we could synthesize the final results into archetypical patterns. Denzin (2017) recommends triangulating the results, as just one empirical method cannot provide a valid result. Therefore, we compare two sets of clusters, eliminate duplets, and condense them into non-optional characteristics. We synthesize the clusters into aggregated cluster types, from which we derive the archetypes by their configuration of each characteristic.

4 Taxonomy of Digital Twins

The following section discusses the literature-based taxonomy of Digital Twins (see van der Valk et al. (2020)) and provides the foundation for the first cluster analysis. The taxonomy required four design iterations (see section *Taxonomy Design*). In total, we identified eleven relevant dimensions with multiple characteristics. The dimensions are grouped into meta-dimensions. These meta-dimensions are arranged along the way that data move through a Digital Twin, i.e., the data collection, the data handling and distribution, and the conceptual scope.

To achieve the goal of integrating qualitative data from industry experts into the taxonomy, the 4th iteration of its design is embedded in a qualitative interview study. We presented each expert with the taxonomy from the 3rd iteration (see van der Valk et al. (2020)) and allowed them to adjust it and illustrate archetypical configurations. Even though the experts agreed on the validity of the dimensions and characteristics, naturally a spectrum of different archetypical designs reflects each expert's unique perspective. Subsequently, the resulting taxonomy should reflect design characteristics and dimensions relevant to archetype design.

Given that understanding, a Digital Twin must contain mandatory and optional characteristics, as well as mutually exclusive dimensions. Table 3 describes the individual classification of the designation:

The categorization into mandatory, mutually exclusive, not relevant, and optional is based on the literature review and from an understanding derived from the expert's insights. In the following, we will describe the different dimensions along with their classification into the meta-dimensions *data collection*, *data handling and distribution*, and *conceptual scope*. We derived the meta-dimensions inductively based on the dimensions' perceived similarity to each other (Bronowski 1953). Table 4 illustrates the final taxonomy. (Table 5).

Interestingly, the taxonomy shows certain differences to the working definition provided in Sect. 2. Occurring differences will be discussed below under the corresponding dimension.

4.1 Meta Dimensions: Data Collection

The meta-dimension **Data Collection** describes all processes to collect data. This category's dimensions are *data acquisition*, *data source*, *synchronization*, and *data input*.

Data Acquisition: Following Sect. 2, one would expect only an automated data acquisition. Some descriptions of Digital Twins merely mention a manual or semi-manual data acquisition (Miller et al. 2018). However, it was apparent that most publications only describe an automated data acquisition, e.g., through sensors (Cai et al. 2017). Nonetheless, the interview study showed the contrast between the literature and the industrial opinion, as a *semi-manual* data acquisition was demanded.

> "Whereby data acquisition, if you see human–machine-interfaces as input-interfaces, then I would of course also allow manual and semi-manual data acquisition." – Interview 9

Table 3 Designation of characteristics as mandatory, mutually exclusive, not relevant, and optional

Type	Choice	Description
Mandatory Dimension/ Characteristic	None	During the literature analysis, it became clear that some characteristics are so important that there is no designable decision to exclude or include them. We term these dimensions *mandatory* to fix their position as basic dimensions unalterable in Digital Twin design
Mutually Exclusive Dimension	One characteristic per dimension	We deem two dimensions as mutually exclusive. Therefore, a Digital Twin can only possess one of the several characteristics. The presence of the overall dimensions is, however, mandatory
Not Relevant	None	This characteristic is part of the taxonomy in order to complete the taxonomy. However, the characteristic is not relevant for archetypical derivations
Optional Design Characteristic	Free	These dimensions are optional and not substantial for a Digital Twin. One may choose freely whether a dimension is considered and, if any, how many characteristics are included

Table 4 Taxonomy of digital twins (M = mandatory, ME = mutually exclusive, N = not relevant, O = optional)

Meta-Dimension	Dimension	Characteristic		
Data Collection	Data Acquisition	Automated (M)	Semi-Manual (O)	
	Data Source	Multiple Sources (M)	Single Source (N)	
	Synchronization	With (M)	Without (**N**)	
	Data Input	Raw Data (M)	Preprocessed Data (O)	
Data Handling and Distribution	Data Governance	Rules Applied (ME)	Rules Not Applied (ME)	
	Data Link	Bi-Directional (M)	One-Directional (N)	
	Interface	HMI (O)	M2M (O)	
	Interoperability	None (ME)	Via Translator (ME)	Fully (ME)
	Purpose	Processing (M)	Transfer (O)	Repository (M)
Conceptual Scope	Accuracy	Identical (N)	Partial (N)	
	Conceptual Elements	Independent (N)	Bound (N)	
	Time of Creation	Digital First (N)	Physical First (N)	Simultaneously (N)

Table 5 Final characteristics with clusters, M = mandatory, O = optional

Literature (L) or Interview (I) Based		Clusters							
		L	L	L	L&I	L	L&I	L	I
Dimension	Charachteristic	C1	C2	C3	C4 & C10	C5	C6 & C9	C7	C8
M									
Purpose	Internal Processing	X	X	X	X	X	X	X	X
	Repository	X	X	X	X	X	–	X	X
Data Input	Raw Data	X	X	X	X	X	X	X	X
Data Link	Bi-Directional	X	–	X	X	X	–	X	X
Synchronization	With	–	–	X	X	X	X	X	X
O									
Purpose	Transfer	X	–	–	X	X	–	–	–
Data Input	Preprocessed Data	X	X	–	X	X	X	X	X
Interface	HMI	X	–	X	X	–	–	X	X
	M2M	X	–	–	X	X	–	–	X
Share of total Literature		2%	10%	25%	21%	14%	19%	9%	0%
Share of total Interviews		0%	0%	0%	17%	0%	25%	0%	58%
Suitable for archetypes		No	No	Yes	Yes	Yes	No	Yes	Yes

Yet, (semi-) manual options are only additional options to mandatory automated data acquisition. Hence, we continue with the decision of whether a Digital Twin acquires its data fully automatically or semimanually, which consists of a mixture of automated and manual processes.

Data Source: In this context, *single* data sources do not mean that just one device gathers data, but that only one type of device, e.g., sensors, is used. Multiple data sources include different types of sources. For example, a Digital Twin used in the chemistry sector receives its live data from sensors attached to the physical asset, historical data from external databases, and humidity or temperatures from the national weather services. As Digital Twins overwhelmingly use multiple data sources, the analysis omits exclusively single sources. The interviewees backed this postulation, as they corroborated a Digital Twin notion, necessarily, requiring data from multiple sources instead of a single source.

Synchronization: As most definitions mandate a synchronization between the Digital Twin and the physical part, the option of *without synchronization* is somewhat surprising (Kritzinger et al. 2018). Nevertheless, there are

some examples in which a Digital Twin is described as a not synchronized digital object (e.g., Banerjee et al. (2017), Grube et al. (2019)). However, concepts without any kind of synchronization contradict definition 1. Hence, we deem synchronization as mandatory.

Interestingly, the dimension synchronization brought forward different views during the interviews, as, though the dichotomous division into discrete and continuous is valid, it is a matter of realizability.

> "It is our reality in the industry that we cannot permanently transfer all data. […] we can't always guarantee the transmission or a WLAN in the depot. That's why the question is: what minimum data must be transmitted and which data can we do without, and of course it is not easy to agree with my demand that the Digital Twin is always up to date." – Interview 6.
> "Well, I think there are many cases where the real-time connection is not crucial. And where on the other hand it would cost you a lot of money to implement." – Interview 15

The industry partners identify hurdles for the implementation of real-time synchronization. Implementation is expensive and depends on the availability of local mobile networks. Correspondingly, the experts stress that real-time synchronization should be implemented when it generates adequately high benefits.

Data Input: We distinguish between *raw* and *preprocessed* data. Raw data is unprocessed data. These data may stem from sensors, data collection devices, or databases. (Pre-)processed data contains all data which comes from software tools, i.e., analytical tools, applications, or smart devices. In most cases, the Digital Twin integrates both data types for internal data processing (Boschert and Rosen 2016; Shangguan et al. 2019).

4.2 Meta-Dimension: Data Handling and Distribution

The meta-dimension **Data handling and distribution** deals with the dimensions *data governance*, *data link*, *interface*, *interoperability* and the *purpose* of a Digital Twin.

Data Governance: Data governance is one of the most critical aspects of data flows (Otto and Weber 2018). Data governance was an umbrella term for everything related, e.g., data security, data sovereignty, or access control. Therefore, more detailed consideration was not possible, and we evaluated data governance as necessary for a Digital Twin. However, the descriptions did not make it clear which specific data governance rules were applied in each case. Hence, we divided this dimension into *rules applied* or *not applied*. The dimension data governance was highlighted as very relevant during the interviews, and suggestions for extensions to more detailed sub-dimensions, e.g., ownership of the Digital Twin, data accessibility, cyber-security, or data quality management, were provided for further research.

Data Link: We consider a data flow to be bi-directional when the Digital Twin communicates with the physical asset and gives feedback to the physical twin. A *one-directional* data link means a data flow just from the physical asset onwards. However, a one-directional data link does not fulfill definition 1. A Digital Twin must provide a bi-directional data link which is, therefore, mandatory. This dimension is especially important, as the data link provides the foundation for a Digital Twin. Furthermore, a bi-directional data link is one of the enablers for autonomous management of the physical asset through the Digital Twin. A multidirectional data link is conceivable when considering a network of a physical asset, Digital Twins, and supplementary systems (downstream and upstream). Data will flow from multiple sources into multiple systems with the Digital Twin as the center of gravity. Nonetheless, this characteristic is not part of a generic twin, and hence, it is not considered any further.

Interface: The dimension *interface* defines through which gateways the data and information leave the Digital Twin. At this point, we only consider two characteristics of interfaces to be relevant for data output, as data input through machine-to-machine interfaces is mandatory. The first one is a *human–machine interface* (HMI) that allows any operator or user of a Digital Twin to access the output data. We do not go into more detail on purpose, as several options, i.e., augmented reality (Tao et al. 2019), dashboards, light- or audio-signals, and more, seem possible (Lutters and Damgrave 2019; Ma et al. 2019). The second option for the output interface is a *machine-to-machine interface* (M2M). This interface provides the possibility for the Digital Twin to communicate with the physical asset directly. This is the primary enabler for an autonomously operating Digital Twin. We do not define the exact design of the M2M interfaces, as they can be manifold (Martinez et al. 2018; Merkle et al. 2019). Additionally, a Digital Twin can possess both interfaces simultaneously (Petrova-Antonova and Ilieva 2019).

Many companies state problems in safety-relevant, infrastructural sectors with machine-to-machine interfaces when it comes to the interviews. Exemplarily, direct integration with a digital tool via machine to machine-to-machine interfaces is forbidden.

> "This is certainly due to the special nature of railroad technology systems, but other critical infrastructures will also have this. We are always forced to prove that there are no retroactive effects." – Interview 6

As the interviewees were given a choice to manipulate the *status quo* of design dimensions and characteristics, we

introduced the dimension **interoperability** during the interviews.

> Interviewer: "So in your opinion, an additional dimension of'interoperability' would be necessary. What could the characteristics look like then?"
> Interviewee: "Non-interoperable, interoperable with translation interface and interoperable per se. Standards play a role here. [...] So interoperability is about standards and the degree of interoperability." – Interview 8

Interoperability guarantees standards for data transfer. The Digital Twin must be able to understand data, especially data that have been preprocessed by others. Foundationally, the Digital Twin must have interoperable interfaces to represent the physical objects continuously. From the interviews, we derive the dimensions *non-interoperable*, *fully interoperable*, or *interoperable* via *a translator*.

Purpose: This dimension integrates a variety of purpose options. Different tasks with their percentage of occurrence follows:

> Simulation (64%), Condition Monitoring and Analysis (50%), Forecast and Prediction (44%), Optimization (38%), Representation (15%), Data Transfer and Storage (10%), Controlling (8%), Machine Learning (7,5%), Decision Making (5%), and Cost Reduction (2,5%)

Due to the high number and variety of tasks we saw the need to further aggregate them. Subsequently, we opted for a threefold classification into *data processing*, *data transfer*, and *repository*. We see this dimension as mutually non-exclusive. A Digital Twin can process, store, and transfer data at the same time.

4.3 Meta-Dimension: Conceptual Scope

Finally, the last three dimensions belong to the meta-dimension **conceptual scope**. This meta-dimension contains *accuracy*, the *conceptual elements*, and the *time of creation*.

Accuracy Accuracy deals with the model part of a Digital Twin. With this dimension, we aim at the scope of the model. We divided accuracy into *identical* and *partial*. Identical accuracy describes a physical asset fully comprehensively, while partial means that a physical asset is reduced to the crucial parts. However, the dimension **accuracy** includes the idealized characteristic *identical*, which designates an exact digital representation of physical objects. As full model accuracy is a state that is likely not attainable, the interviewees suggested that there is no merit in further considering this characteristic.

> "I believe that this [identical accuracy] will never be achieved, because there are so many different characteristics that we have, because we always have only one model. And a model can never be complete. I can always think of something, which is part of it." – Interview 15

Hence, we deem identical model accuracy in analogy to the single data source as not relevant for the taxonomy's practical usage.

Conceptual Elements We divided this dimension into *physically independent* and *bound*. The former describes only the virtual representation, whereas the latter includes the physical aspect in the Digital Twin concept. This dimension focuses on the used definition in a publication. It does not affect the connection between the physical asset and the Digital Twin or the presence of a physical asset, and hence is not of relevance for the development of the archetypes.

Time of Creation the last dimension consists of three mutually exclusive characteristics, namely, *digital-first*, *physical first*, and *simultaneous*. They describe the point in time when the Digital Twin is created. As the creation of an artifact is a process, a discrete point in time cannot be determined. However, we regard the initial creation process as completed when the developed object's commissioning takes place. This point in time is the time of the creation of the Digital Twin. We evaluate whether the digital representation or the physical asset was developed and, therefore, commissioned. Rarely, both objects were commissioned simultaneously. The time of creation marks the point in time when the object is completed. Digital-first means that the Digital Twin is usable before the physical asset's main development steps. On the other hand, physical first means that the physical asset exists before the digital representation. All other issues fall under the characteristic simultaneous. In alignment with recent literature (Boschert and Rosen 2018), the experts agreed that Digital Twins are designed after the physical assets. Nonetheless, the experts identified that the dimensions of the conceptual scope have no merit for archetype derivation and, hence, are excluded from the further analysis.

Figure 4 graphically shows the remaining characteristics and how they relate to each other. Especially, the illustration emphasizes the differentiation between inputting raw data (e.g., from sensors) and pre-processed data (e.g., from external sources), as well as the feedback of data and information as an output of the Digital Twin. Structurally, the twin consists of a digital representation, the data flow, the internal processing, and the internal repository.

4.4 Digital Twin Clusters

We analyzed the database consisting of 233 publications and 15 interviews using the statistical software *R,* Gower (1971)'s coefficient, and the *k-means* algorithm (see section *Cluster Analysis*). First, we analyzed the 233 publications. For a sound analysis, we had to eliminate the outliers (Punj and Stewart 1983). As two dimensions (data acquisition and data source) did not contain relevant distinguishing characteristics, they do not influence archetype design. We designated these dimensions as not relevant for further analysis. As stated above, the conceptual scope dimensions concern the definitional scope of a Digital Twin but not the actual architecture. Thus, we marked these dimensions as applicable but not highly relevant.

Furthermore, we eliminated *data governance* and *data interoperability* as a general concept, which led us to five highly relevant dimensions with distinguishing characteristics. The dimensions *data link, purpose, interface, data input,* and *synchronization* remained for the cluster analysis. We rated every publication, which did not reveal at least three of the five dimensions, as irrelevant and omitted them. We proceeded with 187 publications. Several iterations of the cluster analysis showed that we gained the best results with seven clusters, which was in line with the elbow method. With only twelve additional runaways, we could proceed with 175 objects.

The seven clusters distinguish themselves from one another, as there are no duplicates. Each cluster is denoted in the same way. To better understand the designations, we labeled the characteristic one-directional as without feedback because the data do not flow back to the physical asset.

The individual configurations of a Digital Twin from the interviews were analyzed analogously to the literature. The cluster analysis could be conducted with 12 of the 15 interviews of the first interview series, as three interviews were outliers and did not provide usable results for the analysis. The analysis revealed three clusters of Digital Twins with essential differences from each other. However, we can identify overlaps between the literature-based clusters and interview-based clusters by comparing the three new clusters with Cluster 1–Cluster 7. The second cluster of the interviews provides the same configuration as Cluster 6. Analogously, the third interview cluster is the same as Cluster 4. Lastly, Cluster 8 (first interview cluster) is a narrowed-down version of Cluster 5. In general, both configurations are designed in the same way. However, the interview-based configuration does not provide the option to transfer data in downstream systems.

Following definition 1 (see Sect. 2), a Digital Twin has to provide the mandatory characteristics. As clusters 1, 2, 6, and 9 lack crucial, mandatory characteristics, we exclude them from the development of the archetypes. Cluster 1 is missing synchronization with the physical world. Therefore, the digital part may control the physical world, but it cannot regulate the physical object in dependence of any state changes. The cluster contains only 2% of all reviewed literature, and none of the interviewees described a Digital Twin belonging to cluster 1. Hence, this cluster is of little
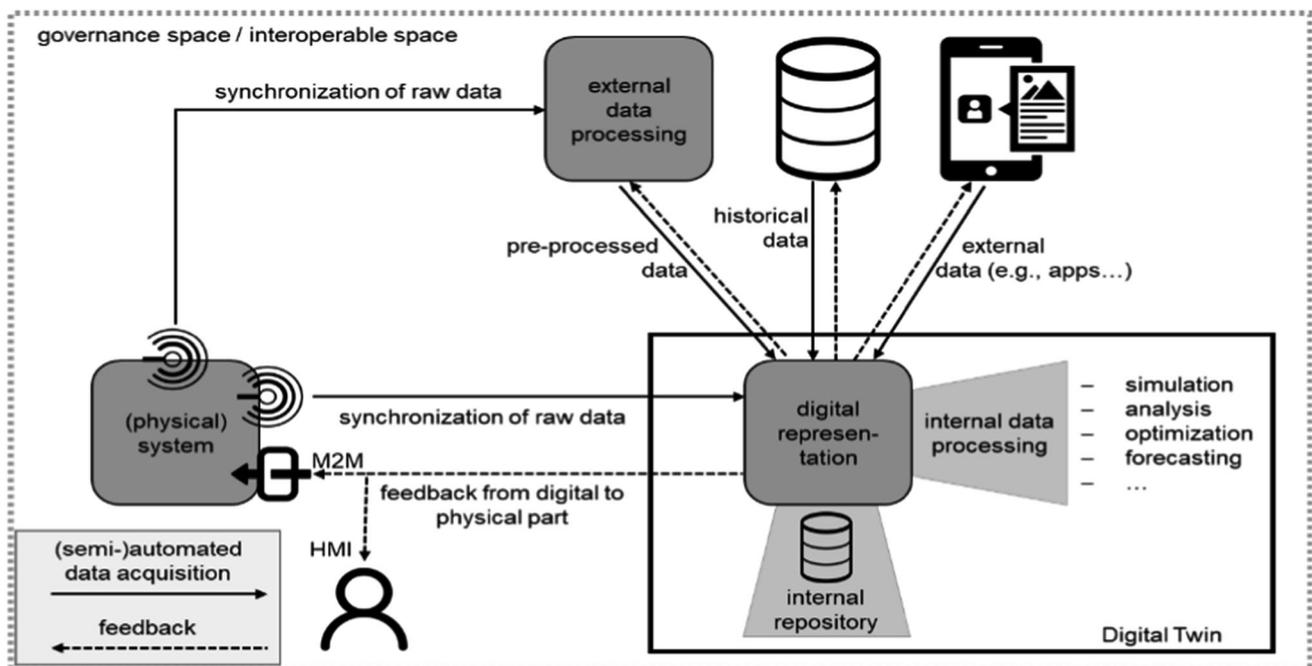
Fig. 4 Conceptual model of a digital twin

relevance and dismissed in further analysis. Examples for this cluster can be found in Beregi et al. (2018), or Lohtander et al. (2018).

Cluster 2 lacks two critical features, namely the bi-directional data linkage as well as the synchronization. Even though 10% of the analyzed literature described a Digital Twin concept that belongs to this cluster, it provides fewer characteristics than cluster 1. It does not fulfill the requirements stated in definition 1. This cluster describes a digital artifact that gathers and stores data, for instance, databases. Schluse et al. (2017) or Radchenko et al. (2018) provide examples for this cluster.

Finally, identical clusters 6 and 9 have to be excluded from the further analysis. Again, there is no option to provide a bi-directional data link between the physical and digital parts. Furthermore, the concept does not offer the ability to store data. Nearly 20% of the literature and 3 out of 12 interviews described an artifact of this cluster. Nevertheless, the mandatory characteristics from definition 1 were not met. Examples for these clusters provide a.o. Buldakova and Suyatinov (2019), and the Interviews 2, 10, and 12. This leaves six clusters for further analysis, namely the derivation of archetypes.

5 Archetypes of Digital Twins

Having illustrated archetypes generated through cluster analysis from the literature and a qualitative interview study, we triangulate our findings by synthesizing methodological approaches (Denzin 2017). The triangulation evaluates each of the qualitative and quantitative

research results through each other (Hammersley 1996). Here, we evaluate the cluster analysis through the interviews and vice versa.

The clusters 3 to 5, 7,8, and 10 describe possible Digital Twins according to definition 1. The optional characteristics provide the distinction between the clusters. For the development of the archetypes, we will proceed with these clusters. Comparing them, cluster 3 describes the Digital Twin with the least capabilities and clusters 4 and 10 with the most ones. All clusters are designated and described in Table 6:

6 Evaluation

We conducted a second series of interviews with experts from different industrial fields to evaluate the five archetypes. We presented the archetypes with the individual characteristics as shown in Table 5. The evaluation interview series confirmed these archetypes, however with minor tweaks:

> For the time being, I do not find any contradiction. I believe that AT 5 would not be accepted in our industry [Healthcare] today. Technically, it can be painted on faster than it can be used. I lack the belief that it would be accepted today. In my world or in our world, this would mean that we would have to take action in the customer system. And I believe that no operator would authorize us to do so. The operator would be interested in the information but would not

Table 6 Definitions of the archetypes

#	Archetype	Definition
AT 1 (Cluster 3)	Basic Digital Twin	Besides the mandatory characteristics, this Digital Twin provides an HMI. Hence, it extends the Digital Twin by just one characteristic and is therefore deemed as basic
AT 2 (Cluster 7)	Enriched Digital Twin	Based on AT 1, this Digital Twin enriches its database by preprocessed data from supplementary systems. Furthermore, many objects in this cluster describe the option of a semi-manual data acquisition
AT 3 (Cluster 8)	Autonomous Control Twin	This twin is an advancement from AT 2. It offers autonomous control, but at the same time, it contains a human–machine interface for the option to intervene. As direct communication with another (virtual or physical) machine is possible, this archetype needs at least interoperability via a translator interface
AT 4 (Cluster 5)	Enhanced Autonomous Control Twin	Enhancing AT 3, the Digital Twin offers autonomous control over a physical asset while integrating external, downstream data processing systems. The interoperability needs to be secured at least over a translator as well. On the downside, this archetype does not offer a human–machine interface
AT 5 (Clusters 4 & 10)	Exhaustive Twin	A Digital Twin with exhaustive data acquisition options, data processing, and control over a physical asset. This archetype provides the user with all options. The twin is able to work and control autonomously. Still, at each point, humans have the ability to intervene or to enrich the database and, hence, provide a semi-manual data linkage. This archetype demands a fully interoperable data linkage to downstream systems as well as to the physical asset itself

wish for/allow active intervention in his processes. – Interview 17

The problem seen here is that a highly developed archetype like AT 5 is technically possible. However, the practical agreement and regulatory aspects may hinder the realization of highly developed archetypes. Another interview-partner agreed with the archetypes but also saw minor issues with the highly developed archetypes due to the high costs while implementing:

> The archetypes seem to make sense. I would make the Digital Twin as simple as possible. I usually have an incredible complexity in the surface, but if I have a lot of things that can be done on a thin budget, I don't think I can afford the luxury of having an expensive complex Digital Twin. – Interview 18

Besides, the evolutionary process from archetype one to five is described:

Today, we are at AT 1 for interlogistics. Maybe also partly AT 2, the topic of preprocessing data is already happening. If I look at intralogistics, we already have AT5 today, but as soon as we talk about industrial borders, about interlogistics between different destinations, then it still takes time. But within a test bed, no question, this [AT 5] is already possible today. – Interview 18

Hence, we conclude that the archetypical patterns AT 1 to AT 5 show an evolutionary process for Digital Twins. The patterns contained a high degree of validity through the application of the triangulation research process. Furthermore, they were confirmed by the evaluation interviews. Additionally, the archetypes represent a sizeable number of papers and interviews (see Table 5).

While AT 3 dominates the industrial view on Digital Twins with a 58% share, the stakes between the archetypes based on the clusters from the literature analysis are more evenly distributed and range between 9 and 25%.

Table 7 Examples for the archetypes

#	Archetype	Example	Description
AT 1	Basic Digital Twin	Burnett et al. (2019), Wang et al. (2020), a.o	It is difficult to find matching industrial examples, as almost all software tools provide at least a baseline of the optional characteristics. Therefore, we have to grade the industrial relevance for AT 1 as low. This is in accordance with the fact that no consulted company described a Digital Twin that would fit in AT 1. The examples describe concepts with exemplary applications in research labs. No industrial usage could be identified
AT 2	Enriched Digital Twin	AutoMod, Ansys, a.o	This archetype is equally hard to identify in usable software tools as AT 1. The nearest solutions are simulation tools like AutoMod or Ansys. These simulation tools provide the mandatory characteristics but are further developments of classical simulations software. As most simulation software is developed towards AT 5, i.e., a holistic tool, AT 2 is more of a theoretical archetype
AT 3	Autonomous Control Twin	Interviews 1, 3, 6, 7, and 13—15	This archetype stems from 7 interviews. The Digital Twins are used in manufacturing contexts. They accompany the production process. The solutions are customized developments. Hence, due to the anonymization, no concrete examples can be given. The difference to the software tools presented in AT 5 level consisted in the missing transfer option. The interviewed companies were not ready for a further sharing of data and, therefore, did not need a transfer option
AT 4	Enhanced Autonomous Control Twin	Lutters and Damgrave (2019), Eckhart and Ekelhart (2019), a.o	For this archetype, industrial appliances are hard to identify. No interviewee could be sorted into this archetype. Additionally, the identified literature describes conceptual approaches or pilots in research contexts. The main hindrance with this archetype is the missing HMI. As HMIs are deemed crucial by industrial experts, no solution is developed without one
AT 5	Exhaustive Twin	Predix (GE Power Digital Solutions 2016 Electric), Siemens MindSphere (Siemens 2018)	Besides the mandatory characteristics which designate the product as a Digital Twin, Predix provides all optional characteristics. Predix is used to optimize and analyze power plants and is seen as an established industrial appliance
			Siemens MindSphere is another industrial example of this archetype. Similar to Predix, Mindsphere provides all mandatory as well as optional characteristics in its full application. MindSphere is a modular environment and may be used in a lighter version, which means that some optional characteristics must not be chosen

Nevertheless, a high interest in the exhaustive Digital Twin is obvious, as this percentage is the second-highest amongst the literature clusters. Additionally, it corresponds to one of the interview clusters. The interest of the industry experts stretches from the autonomous control twin to the exhaustive twin. This is as one would expect, as the different archetypes can be seen as development steps towards the exhaustive one.

Especially the question of interoperability is a highly discussed one within the industry, but is neglected in the research focusses. This shows a particular gap between the theoretical understanding and the practical use of Digital Twins. Therefore, we provide the industrial relevance for each archetype by supplying industrial examples fitted to the archetypes (Table 7).

The archetypes are a reflection of recent trends and developments in Digital Twins. For example, the mandatory characteristics are echoed by the existing literature corpus (e.g., see Kritzinger et al. (2018), Jones et al. (2020), or van der Valk et al. (2020)). Consequently, we see the mandatory characteristics as the smallest common denominator that is a potential baseline for a common understanding of Digital Twins. Beyond that baseline, our archetypical representations enhance the prevailing understanding in the literature through an extensive and in-depth interview study that produces optional characteristics in dependence on various individual use cases and industrial applications. Exemplarily, we point to the issue of interoperability, which was discussed prominently in the interview study, yet neglected in the literature. This shows a conceptual disconnection between the existing theoretical understanding of Digital Twins and their practical application in industry.

With these insights, we extend the definition 1 in the following:

Definition 5 The Digital Twin is a virtual construct that represents a physical counterpart, integrates several data inputs with the aim of data handling, data storing, and data processing, and provides an automatic, bi-directional data linkage between the virtual world and the physical one. Synchronization is crucial to the Digital Twin to display any changes in the state of the physical object. Additionally, a Digital Twin must comply with data governance rules and must provide interoperability with other systems.

7 Conclusion, Limitations, and Contributions

This paper developed archetypes based on Digital Twins characteristics derived from a sound literature base and extended through interviews with industry experts. From this database, we derived clusters of Digital Twins (**RO1**).

Each cluster possesses a particular configuration of characteristics. We could identify seven clusters, which showed specific patterns in their configurations. From these patterns, we were able to derive characteristics that each cluster contains (**RO2**). Denoted as preliminary mandatory characteristics, we could identify that a Digital Twins should contain an automated data acquisition, multiple data sources, the appliance of data governance rules, a data processing and repository, and raw data input.

The interview series provided some interesting insights. Most characteristics could be confirmed. However, some new aspects appeared, such as the semi-manual data acquisition and the dimension interoperability. The analysis of the configurations from the experts showed more mandatory characteristics. The additional mandatory characteristics are a synchronization between the Digital Twin and the physical asset and a bi-directional data link.

Furthermore, we could identify six optional characteristics. This leads to the identification of five archetypes for Digital Twins (**RO3**). These archetypes build upon each other. All archetypes contain the mandatory parts, but they show different configurations in the optional parts from an Assistance Twin to an Exhaustive Twin. Furthermore, we recognize the most important identifying characteristics, which distinguish a Digital Twin from other concepts, i.e., Digital Threads or virtual models, as the presence of synchronization and bi-directional data linkage between the Digital Twin and its physical counterpart. Additionally, the archetypes represent a development of Digital Twins from a more Basic Twin towards the Exhaustive Twin. Hence, the different archetypes may act as a maturity model for the overall development of Digital Twins.

Our work is subject to certain **limitations**. As the definition of the review scope for the literature analysis is subjective, other research teams might define other scopes and, therefore, might find other results. Secondly, in a similar way to coding this process is prone to subjective influences. This research provides several contributions. As **scientific contribution**, this paper analyzes patterns of Digital Twins through the derivation of archetypes. It lays a profound framework for the classification of Digital Twins. We provide an ample contribution to the scientific knowledge base of Digital Twins, which is established by the generalized nature of archetypes. With the derived archetypes, one can sort the differentiating streams in research on Digital Twins. This lays the foundation for further research. Starting from this conceptualization, further scientific contributions could focus on one particular archetype and provide a deeper understanding and elaboration of each archetype. For example, reference models or design principles, including specific technical regulations for implementation, are conceivable.

As our work is based on and partly evaluated through input from industry experts, it provides ample **managerial contributions**. It can be used as a guideline for the development of Digital Twins in commercial environments. Practitioners can compare their understanding of the archetypes and may find a perfect fit with additional information on supplementary modules of a Digital Twin. At the very least, practitioners will gain insights into the fast-growing field of Digital Twin research. Additionally, one can compare the development process's position with the different developed Digital Twins groups. The groups' size will make it possible to conclude how far the development processes have progressed.

References

Autiosalo J, Vepsalainen J, Viitala R, Tammi K (2020) A feature-based framework for structuring industrial digital twins. IEEE Access 8:1193–1208. https://doi.org/10.1109/ACCESS.2019.2950507

Bailey KD (1994) Typologies and taxonomies: an introduction to classification techniques. SAGE, Thousand Oaks

Banerjee A, Dalal R, Mittal S, Joshi KP (2017) Generating digital twin models using knowledge graphs for industrial production lines. In: Proceedings of the 2017 ACM Web Science Conference, Troy. ACM, New York, pp 425–430

Bauernhansl T, Krüger J, Reinhart G, Schuh G (2016) WGP-Standpunkt Industrie 4.0. Wissenschaftliche Gesellschaft für Produktionstechnik WGP e.V., Darmstadt

Beregi R, Szaller Á, Kádár B (2018) Synergy of multi-modelling for process control. IFAC-PapersOnLine 51:1023–1028. https://doi.org/10.1016/j.ifacol.2018.08.473

Boschert S, Rosen R (2018) Next generation digital twin. In: Horváth I, et al (eds) Tools and methods of competitive engineering: Proceedings of the 12th International Symposium on Tools and Methods of Competitive Engineering, Las Palmas de Gran Canaria. Delft University of Technology, Delft, pp 209–218

Boschert S, Rosen R (2016) Digital twin – the simulation aspect. In: Hehenberger P, Bradley D (eds) Mechatronic futures. Springer, Cham, pp 59–74

Encyclopedia Britannica (2011) Archetype. https://www.britannica.com/topic/archetype. Accessed 6 September 2021

Bronowski J (1953) The common sense of science. Harvard University Press, A Harvard paperback

Buldakova TI, Suyatinov SI (2019) Biological principles of integration information at big data processing. In: Proceedings of the 2019 International Russian Automation Conference, Sochi. IEEE, Piscataway, pp 1–6

Burnett D, Thorp J, Richards D, Gorkovenko K, Murray-Rust D (2019) Digital twins as a resource for design research. In: Khamis M, et al (eds) Proceedings of the 8th ACM International Symposium on Pervasive Displays. ACM Press, New York, pp 1–2

Cai Y, Starly B, Cohen P, Lee Y-S (2017) Sensor data and information fusion to construct digital-twins virtual machine tools for cyber-physical manufacturing. Procedia Manuf 10:1031–1042. https://doi.org/10.1016/j.promfg.2017.07.094

Cimino C, Negri E, Fumagalli L (2019) Review of digital twin applications in manufacturing. Comput Ind. https://doi.org/10.1016/j.compind.2019.103130

Detecon Consulting (2019) Companies preparing use concepts for digital twin. https://www.detecon.com/en/detecon-study-companies-preparing-use-concepts-digital-twin. Accessed 6 September 2021

Cooper HM (1988) Organizing knowledge syntheses: a taxonomy of literature reviews. Knowl Soc 1:104–126. https://doi.org/10.1007/BF03177550

Denzin NK (2017) The research act: a theoretical introduction to sociological methods. Methodological perspectives. Routledge, Piscataway

Oxford Dictionary (2020) Archetype. https://www.lexico.com/definition/archetype. Accessed 6 September 2021

Eckhart M, Ekelhart A (2019) Digital twins for cyber-physical systems security: state of the art and outlook. In: Biffl S et al (eds) Security and quality in cyber-physical systems engineering. Springer, Cham, pp 383–412

Enders MR, Hoßbach N (2019) Dimensions of digital twin applications - a literature review. In: Proceedings of the 25th Americas Conference on Information Systems, Cancun, pp 1–10

Errandonea I, Beltrán S, Arrizabalaga S (2020) Digital twin for maintenance: a literature review. Comput Ind 123:103316. https://doi.org/10.1016/j.compind.2020.103316

Fuller A, Fan Z, Day C, Barlow C (2020) Digital twin: enabling technologies, challenges and open research. IEEE Access 8:108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358

GE Power Digital Solutions (2016) GE Digital Twin. Analytic engine for the digital power plant. http://www.ge.com/digital/sites/default/files/download_assets/Digital-Twin-for-the-digital-power-plant-.pdf. Accessed 6 September 2021

Glaessgen E, Stargel D (2012) The digital twin paradigm for future NASA and U.S. Air Force vehicles. In: structures, structural dynamics, and materials and co-located conferences: 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference. American Institute of Aeronautics and Astronautics, Reston, 1–14

Glass RL, Vessey I (1995) Contemporary application-domain taxonomies. IEEE Softw 12:63–76. https://doi.org/10.1109/52.391837

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857–871

Greene JC, Caracelli VJ, Graham WF (1989) Toward a conceptual framework for mixed-method evaluation designs. Educ Eval

Policy Anal 11:255–274. https://doi.org/10.3102/01623737011003255

Grieves M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen F-J et al (eds) Transdisciplinary perspectives on complex systems: new findings and approaches. Springer, Cham, pp 85–113

Grieves M (2014) Digital twin: manufacturing excellence through virtual factory replication. Michael W. Grieves LLC

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

Haag S, Anderl R (2019) Automated generation of as-manufactured geometric representations for digital twins using STEP. Procedia CIRP 84:1082–1087. https://doi.org/10.1016/j.procir.2019.04.305

Hammersley M (1996) The relationship between qualitative and quantitative research: Paradigm loyalty versus methodological eclecticism. In: Richardson JTE (ed) Handbook of qualitative research methods for psychology and the social sciences. British Psychological Society, Leicester, pp 1–27

Helu M, Hedberg T, Feeney AB (2017) Reference architecture to integrate heterogeneous manufacturing systems for the digital thread. CIRP J Manuf Sci Technol 19:1–5. https://doi.org/10.1016/j.cirpj.2017.04.002

Iivari J, Rotvit Perlt Hansen M, Haj-Bolouri A (2020) A proposal for minimum reusability evaluation of design principles. Eur J Inf Syst 30(3):286–303. https://doi.org/10.1080/0960085X.2020.1793697

Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recognit Lett 31:651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain AK, Murty MN, Flynn PJ (1999) Data clustering. ACM Comput Surv 31:264–323. https://doi.org/10.1145/331499.331504

Jones D, Snider C, Nassehi A, Yon J, Hicks B (2020) Characterising the digital twin: a systematic literature review. CIRP J Manuf Sci Technol 29:36–52. https://doi.org/10.1016/j.cirpj.2020.02.002

Josifovska K, Yigitbas E, Engels G (2019) Reference framework for digital twins within cyber-physical systems. In: 2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems. IEEE, Montreal, pp 25–31

Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 51:1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474

Lapadat JC, Lindsay AC (1999) Transcription in research and practice: from standardization of technique to interpretive positionings. Qual Inq 5:64–86

Levy Y, Ellis T (2006) A systems approach to conduct an effective literature review in support of information systems research. Int J Emerg Transdiscipl 9:181–212. https://doi.org/10.28945/479

Liddell HG, Scott R, Jones HS, McKenzie R (1940) A Greek-English lexicon, 9th edn. Clarendon Press, Oxford

Lohtander M, Garcia E, Lanz M, Volotinen J, Ratava J, Kaakkunen J (2018) Micro manufacturing unit – creating digital twin objects with common engineering software. Procedia Manuf 17:468–475. https://doi.org/10.1016/j.promfg.2018.10.071

Lutters E, Damgrave R (2019) The development of pilot production environments based on digital twins and virtual dashboards. Procedia CIRP 84:94–99. https://doi.org/10.1016/j.procir.2019.04.228

Ma X, Tao F, Zhang M, Wang T, Zuo Y (2019) Digital twin enhanced human-machine interaction in product lifecycle. Procedia CIRP 83:789–793. https://doi.org/10.1016/j.procir.2019.04.330

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2019) Cluster: cluster analysis basics and extensions. R Package Version 2.1.0

Martinez GS, Sierla S, Karhela T, Vyatkin V (2018) Automatic generation of a simulation-based digital twin of an industrial process plant. 44th Annual Conference of the IEEE Industrial Electronics Society. Piscataway, Washington D.C. IEEE, pp 3084–3089

Merkle L, Salinas Segura A, Grummel JT, Lienkamp M (2019) Architecture of a digital twin for enabling digital services for battery systems. Proceedings IEEE International Conference on Industrial Cyber Physical Systems. Piscataway, Taipei. IEEE, pp 155–160

Merton RK, Kendall PL (1946) The Focused Interview Am J Soc 51:541–557

Miller AM, Alvarez R, Hartman N (2018) Towards an extended model-based definition for the digital twin. Comput-Aided Design Appl 15:880–891. https://doi.org/10.1080/16864360.2018.1462569

Möller F, Bauhaus H, Hoffmann C, Niess C, Otto B (2019) Archetypes of digital business models in logistics start-ups. In: proceedings of the 27th European conference on information systems, Uppsala. https://aisel.aisnet.org/ecis2019_rp/17

Myers MD, Newman M (2007) The qualitative interview in IS research: examining the craft. Inf Organ 17:2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22:336–359. https://doi.org/10.1057/ejis.2012.26

Oberländer AM, Lösser B, and Rau D (2019) Taxonomy research in information systems: a systematic assessment. In: proceedings of the 27th European conference on information systems, Uppsala. https://aisel.aisnet.org/ecis2019_rp/144

Ochs E (1979) Transcription as theory. In: Ochs E, Schieffelin BB (eds) Developmental pragmatics. Academic Press, New York, pp 43–72

Otto B, Weber K (2018) Data governance. In: Hildebrand K et al (eds) Daten- und Informationsqualität. Springer, Wiesbaden, pp 269–286

Patton MQ (2002) Qualitative research & evaluation methods. SAGE, Thousand Oaks

Petrova-Antonova D, Ilieva S (2019) Methodological framework for digital transition and performance assessment of smart cities. In: 2019 4th international conference on smart and sustainable technologies, pp 1–6. doi: https://doi.org/10.23919/SpliTech.2019.8783170

Punj G, Stewart DW (1983) Cluster analysis in marketing research: review and suggestions for application. J Mark Res 20:134–148. https://doi.org/10.2307/3151680

PwC (2020) Digital factories 2020. Shaping the future of manufacturing. https://www.pwc.de/de/digitale-transformation/digital-factories-2020-shaping-the-future-of-manufacturing.pdf. Accessed 1 June 2021

Radchenko G, Alaasam A, Tchernykh A (2018) Micro-workflows: Kafka and Kepler fusion to support digital twins of industrial processes. In: 2018 IEEE/ACM international conference on utility and cloud computing companion. IEEE, pp 83–88

Rivera LF, Jiménez M, Angara P, Villegas NM, Tamura G, Müller HA (2019) Towards continuous monitoring in personalized healthcare through digital twins. In: proceedings of the 29th annual international conference on computer science and software engineering. IBM Corp, pp 329–335

Rosen R, von Wichert G, Lo G, Bettenhausen KD (2015) About the importance of autonomy and digital twins for the future of

manufacturing. IFAC-PapersOnLine 48:567–572. https://doi.org/10.1016/j.ifacol.2015.06.141

Sarker S, Sarker S (2009) Exploring agility in distributed information systems development teams: an interpretive study in an offshoring context. Inf Syst Res 20:440–461

Schluse M, Atorf L, Rossmann J (2017) Experimentable digital twins for model-based systems engineering and simulation-based development. In: 2017 annual IEEE international systems conference. IEEE, pp 1–8

Schuh G, Kelzenberg C, Wiese J, Ochel T (2019) Data structure of the digital shadow for systematic knowledge management systems in single and small batch production. Procedia CIRP 84:1094–1100. https://doi.org/10.1016/j.procir.2019.04.210

Scopus (2020) Analyze search results - digital twin. https://www.scopus.com/term/analyzer.uri?sid=19082651509480c0dfb47fda36919efa&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22digital+twin%22%29&sort=plf-f&sdt=b&sot=b&sl=29&count=1162&analyzeResults=Analyze+results&txGid=d405e244754866c5b3d614e5d90bda33. Accessed 6 September 2021

Shangguan D, Chen L, Ding J (2019) A hierarchical digital twin model framework for dynamic cyber-physical system design. In: proceedings of the 5th international conference on mechatronics and robotics engineering. ACM Press, New York, pp 123–129

Siemens (2018) MindSphere. Enabling the world's industries to drive their digital transformations. https://www.prolim.com/wp-content/uploads/2019/04/Siemens-MindSphere-Whitepaper.pdf. Accessed 6 September 2021

Szopinski D, Schoormann T, Kundisch D (2019) Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: proceedings of the twenty-seventh European conference on information systems, Uppsala

Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F (2018) Digital twin-driven product design, manufacturing and service with big data. Int J Adv Manuf Technol 94:3563–3576. https://doi.org/10.1007/s00170-017-0233-1

Tao F, Sui F, Liu A, Qi Q, Zhang M, Song B, Guo Z, Lu SC-Y, Nee AYC (2019) Digital twin-driven product design framework. Int J Prod Res 57:3935–3953. https://doi.org/10.1080/00207543.2018.1443229

Thesaurus (2020) Synonyms of "Standard". https://www.thesaurus.com/browse/standard?s=t. Accessed 6 September 2021

Urbina Coronado PD, Lynn R, Louhichi W, Parto M, Wescoat E, Kurfess T (2018) Part data integration in the shop floor digital twin: mobile and cloud technologies to enable a manufacturing execution system. J Manuf Syst 48:25–33. https://doi.org/10.1016/j.jmsy.2018.02.002

van der Valk H, Haße H, Möller F, Arbter M, Otto B (2020) A taxonomy of digital twins. In: AMCIS 2020 proceedings. AIS, Salt Lake City

Vom Brocke J, Simons A, Niehaves B, Reimer K, Plattfaut R, Cleven A (2009) Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: Proceedings of the 17th European Conference on Information Systems. AIS, Verona

Wagner R, Schleich B, Haefner B, Kuhnle A, Wartzack S, Lanza G (2019) Challenges and potentials of digital twins and Industry 4.0 in product design and production for high performance products. Procedia CIRP 84:88–93. https://doi.org/10.1016/j.procir.2019.04.219

Wang W, Zhang Y, Zhong RY (2020) A proactive material handling method for CPS enabled shop-floor. Robotics Comput-Integr Manuf 61:101849. https://doi.org/10.1016/j.rcim.2019.101849

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26:xiii–xxiii

Weking J, Hein A, Böhm M, Krcmar H (2018) A hierarchical taxonomy of business model patterns. Electron Mark 30:447–468. https://doi.org/10.1007/s12525-018-0322-5