Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(2), 8343.
https://doi.org/10.54517/m8343

Article

Data Categories in the Metaverse and Their Application to DNA Data Storage

Martha Wegert[1], **Markus Weinberger**[1,*]

[1] Aalen University, Germany
***Corresponding author:** Markus Weinberger, Markus.Weinberger@hs-aalen.de

CITATION

Wegert M, Weinberger M. Data Categories in the Metaverse and Their Application to DNA Data Storage. Metaverse. 2026; 7(2): 8343. https://doi.org/10.54517/m8343

ARTICLE INFO

Received: 22 December 2025
Accepted: 22 April 2026
Available online: 13 May 2026

Abstract: This study proposes a systematic categorization of Metaverse data and identifies its specific storage requirements. As digital worlds continuously generate massive volumes of data, they place increasing demands on existing data storage solutions. Based on a systematic literature review, the study addresses the research question: Which data categories exist in the Metaverse, and what specific storage requirements do they impose? Building on this analysis, the study evaluates DNA data storage as a future-oriented alternative to conventional storage technologies. The Metaverse is underpinned by key technologies such as the Internet of Things (IoT), Extended Reality (XR), Digital Twins (DT), Artificial Intelligence (AI), Blockchain, 5G, and cloud and edge computing. Depending on their relation to the physical or virtual sphere, these technologies can be functionally grouped into three layers: content, device, and infrastructure. This framework forms the conceptual basis for identifying the primary sources of data generation and classification. The analysis reveals two primary sources of data creation within the Metaverse: physical data, generated by humans, objects, and the environment, and virtual data, produced through digital representations and interactions. A total of 23 data categories were identified and their storage requirements evaluated according to seven criteria: write frequency, access intensity, latency, data volume, protection needs, redundancy, and lifespan. In an example application, the identified data categories and their storage requirements were compared to the properties of DNA data storage. According to current research, DNA data storage is particularly well-suited for long-term archiving of large data volumes due to its exceptional density and durability. However, broader adoption will require significant cost reductions and improvements in access speed. The study concludes that integrating DNA-based storage concepts could enhance the sustainability and scalability of data management in future Metaverse ecosystems.

Keywords: Metaverse; technology; data categories; data types; DNA data storage

1. Introduction

The Metaverse represents the next stage in the evolution of the Internet, where the boundaries between physical and virtual worlds increasingly blur [1,2]. It enables immersive experiences and real-time interaction within persistent digital environments that are accessible anytime and anywhere. For a growing number of users, the Metaverse has become a central space for social interaction across private and professional contexts [3]. It is not a single platform but a convergence of diverse technologies that together form a complex sociotechnical system connecting physical and virtual realities. As a result, enormous amounts of data are continuously generated through human activity, environmental sensing, and digital interaction [4]. These data differ in type, format, and temporal relevance, leading to highly diverse storage requirements. Understanding these requirements is essential for designing efficient and sustainable data storage architectures for the Metaverse.

Despite the huge data volumes and its importance in the Metaverse, the relation of technology and data, as well as data categories and their respective storage requirements are insuffiently researched (see Section 2). This study aims to address this research gap by identifying key Metaverse technologies, proposing a Metaverse framework that integrates these technologies with a data-centric perspective, and applying it to derive the types of data generated in the Metaverse and determine their specific storage requirements. Thus, it also provides a sound basis for deliberate design decisions regarding data storage in the development of Metaverse platforms or experiences.

The resulting data categories and associated storage requirements are then used to assess the applicability of DNA data storage in the Metaverse context. While traditional storage technologies such as hard drives and solid-state drives currently meet most demands, they are expected to reach their physical [5] and energetic limits [6] as data volumes increase exponentially [7]. This raises the question of how to achieve more efficient long-term storage, specifically for archiving Metaverse data. DNA data storage has emerged as a promising alternative. Synthesized DNA offers exceptional density, stability, and longevity, making it a potential medium for sustainable data preservation [8].

The remainder of the article is structured as follows. Section 2 presents related work to motivate the research question and situate it within the existing body of knowledge. Section 3 explains the applied research process, while Section 4 presents the respective results. The application of the results is demonstrated on the example of DNA data storage in Section 5. Finally, Section 6 discusses the results, points out limitations of the presented study, and presents directions for future research.

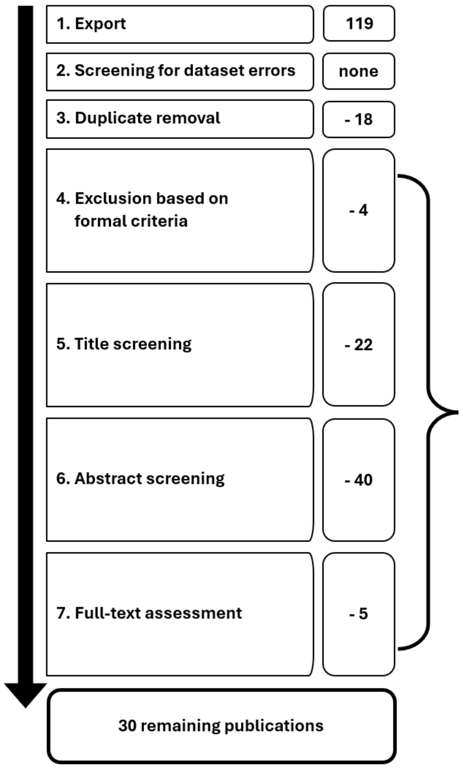
2. Related work

A comprehensive overview of the relevant technologies and data structures is required to establish a solid conceptual foundation for analyzing data categories and their specific storage requirements within the Metaverse. Therefore, a systematic literature review was conducted to identify, classify, and evaluate current research in this area.

Three scientifically recognized and discipline-specific databases were used: ScienceDirect, Web of Science, and SpringerLink. The search combined the terms "Metaverse" AND "data." To ensure a precise, targeted selection, the search was limited to publications that contained both terms in the title ("All in title"). The search phrase was deliberately kept broad to capture a wide range of potentially relevant literature. All searches were conducted in German and English to include the international academic publication landscape. Only sources published between 2021 and mid-2025 were considered in the literature search.

In total, 119 publications were retrieved from the three databases (ScienceDirect: 22; Web of Science: 70; SpringerLink: 27). The significantly higher number of results in *Web of Science* reflects its broader interdisciplinary coverage across computer science, communication studies, psychology, and economics.

Given the number of results, a multistage selection process was applied to identify relevant contributions in a transparent and reproducible manner. The process consisted of seven steps, which are depicted in **Figure 1**. In addition, the figure comprises the applied inclusion and exclusion criteria. After completing all seven steps, 30 publications remained, which were classified as relevant to answering the research question. Most of these were published between 2023 and mid-2025 (2021: 0; 2022: 1; 2023: 13; 2024: 11; 2025: 5) and comprised 23 journal articles, one conference paper, and six book chapters.



Inclusion criteria	
IC1	The respective search term appears in full in the publication title ("all in title").
IC2	Publications whose abstracts addressed the components, data categories, storage requirements, or current storage technologies of the Metaverse and the associated challenges.
IC3	Literature that examines the Metaverse ecosystem.
IC4	Literature with a focus on different data sources in the Metaverse.

Exclusion criteria			
EC1	Publications or books for which no digital access was available.	EC7	Publications with no or very limited citation and source information.
EC2	Literature not written in German or English.	EC8	Literature published before 2021 and after June 2025.
EC3	Master's theses, corrigenda, open letters, and any publications marked as "retracted."	EC9	Literature with a strong focus on data quality, data analytics, machine learning, or AI.
EC4	Analyses that use Big Data but provide no substantive connection to data in the Metaverse.	EC10	Literature with a strong focus on data protection, cryptography, or blockchain technology.
EC5	Literature with a societal or values-based focus (e.g., ethics, culture, gender data gap).	EC11	Literature focusing on only one Metaverse-related single technology (e.g., sensor glove, VR application).
EC6	Literature offering experience-based perspectives on the Metaverse (e.g., sensory perception, service perception).	EC12	Literature with a strong focus on the Metaverse in corporate contexts, supply chain management, healthcare, education, tourism, smart cities, transportation or water-quality applications.

Figure 1. Literature selection process including inclusion and exclusion criteria

Despite the substantial data volumes generated in the Metaverse, current research treats these data in a largely superficial manner. The analysis of the 30 selected publications reveals that no consolidated overview of Metaverse data has yet been established. As a result, the origins, structure, and storage implications of Metaverse data remain insufficiently understood.

To determine storage requirements in the Metaverse, it is first necessary to identify the sources of these data. Because these origins are closely linked to the technologies in use, the question arises of how these technologies can be functionally positioned within the Metaverse.

Different approaches address this question. Two of them, despite following different theoretical foundations, reveal a comparable functional structure. Jung and Jeon (2022) draw on Porter's Diamond Model from economics, which was originally developed to explain the international competitiveness of nations [9]. They transfer its logic to the Metaverse and propose a categorization into infrastructure, devices, content, and platforms. Casale-Brunet et al. (2023), by contrast, base their approach on an architecture-oriented layered model and divide the Metaverse into three closely interconnected levels: technological infrastructure (hardware, software, communication, blockchain, computing resources, and data storage), an interaction layer (AR, content creation, avatars, digital twins), and a user ecosystem (social interactions, digital assets) [10]. Both models therefore distinguish between technological foundations, interactive components, and user-centered content.

An alternative perspective is offered by Zhang et al. (2023) and Ates et al. (2023), who draw a clearer distinction between physical and virtual space [11,12]. At the center of their approach is the bidirectional flow of data: real environments are captured through sensors and interactive devices, which belong to the physical space. The resulting data are processed through an infrastructure layer and transferred into the virtual space, where they generate content, form digital twins, or enable immersive experiences. Compared with the previously mentioned approaches, the infrastructure layer here is understood not only as a technical foundation but also as a connecting element between both spheres.

Although existing classifications offer useful reference points, they are insufficient for this work. They lack an explicit focus on the role of data within the Metaverse. Therefore, this study introduces a new framework that combines technological elements with a data-focused approach.

3. Materials and methods

To derive data categories in the Metaverse systematically and reproducibly, a structured multi-step methodology was applied. It is depicted in **Figure 2**.

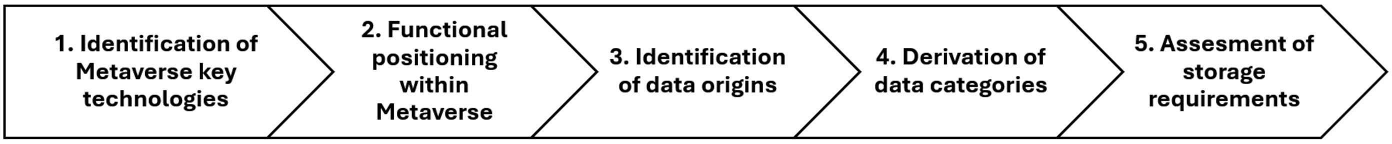


Figure 2. Multi-step research process. (Authors´ own illustration.)

The first step of this methodology focused on identifying the key technologies mentioned in existing Metaverse literature. Across the reviewed literature, nine publications were identified. They explicitly describe technological components of the Metaverse. These works pursue different objectives, such as outlining future technology trends in the Metaverse [13] or analyzing security-related aspects of specific components [14]. They vary notably in the breadth and depth of the technologies they discuss. Despite these differences, a thematic overlap emerges. **Table 1** indicates how many of the nine sources list each technology as a component of the Metaverse.

Based on this comparative analysis, several patterns become visible. Extended reality technologies (XR), blockchain, and AI appear in close to all of the examined approaches. Digital twins, communication technologies such as 5G, and cloud and edge computing are also identified as central components in more than half of the reviewed sources. IoT appears in four of the nine sources, while brain–computer interfaces (BCI) and data mining are mentioned only sporadically and are therefore not examined in detail.

Table 1. Metaverse technologies identified in the reviewed literature. Key technologies printed bold (Authors´ own table.)

Source	**IoT**	HMI	**XR**	BCI	3D Model	**Digital Twin**	Digital Human	**AI**	Data Mining	**Block-chain**	**5G**	**Cloud & Edge**
[1]	X	X	X		X	X		X		X	X	X
[4]			X					X		X		
[13]				X		X	X	X		X		
[15]			X					X		X	X	
[16]	X					X	X	X		X	X	X
[12]	X	X	X					X	X	X	X	X
[14]	X		X			X		X		X		

Continuation Table:

Source	**IoT**	HMI	**XR**	BCI	3D Model	**Digital Twin**	Digital Human	**AI**	Data Mining	**Block-chain**	**5G**	**Cloud & Edge**
[10]			X			X				X	X	X
[11]			X		X						X	X

It is important to note that certain technologies are subsumed under broader concepts. Digital human reconstruction is treated as a subdomain of digital twin technology. Similarly, 3D modelling – which is essential for creating virtual environments and avatars – can be considered part of XR. Human–machine interaction (HMI) is likewise not treated as an independent technology in this work. It describes the interface between humans and digital systems, encompassing technologies such as XR, IoT, and BCI.

Based on these considerations, seven key Metaverse technologies were derived: IoT, XR, digital twins, AI, blockchain, communication technologies (primarily 5G), and cloud and edge computing.

In the second step of the research process (see Figure 2), the identified key technologies are compiled into a consistent framework. To link this technological structure to actual data-creation points, in the next step (3), the phases of the data lifecycle were mapped to the framework.

In the Metaverse, the data lifecycle typically comprises six consecutive phases: collection, storage, transfer, processing, sharing, and deletion [14]. The challenge of heterogeneous data origins becomes apparent already in the collection phase [4,10,11]. Due to the large number of different data sources, the literature often refers to multi-source data [17]. Although storage is defined as a separate phase, it does not represent a closed step. From a technical perspective, data remains stored throughout all subsequent phases, either temporarily—for example, in working or intermediate memory—or permanently in a database. Storage, therefore, forms the foundation for all subsequent phases and accompanies the data in different forms throughout the entire lifecycle.

For this study, the data lifecycle is extended by an additional phase: virtual representation and interaction. This phase is positioned between data processing and data sharing and focuses on data generated exclusively within the virtual sphere. This data creation point is often overlooked in the traditional data lifecycle, which is why it is explicitly stated here (see **Figure 3**).

Based on this foundation, step 4 of the research process (see Figure 2) identifies data categories in the Metaverse.

To evaluate the suitability of different storage technologies, including DNA data storage, the derived data categories must be assessed against clearly defined storage requirement criteria – step 5 of the research process. The aim is to make these specific storage requirements transparent. Seven storage requirement criteria (SC) are used for this purpose. These criteria capture the technical, organizational, and security requirements of a storage system. They address the frequency of access and modification, required availability, data volume, need for protection, need for redundancy, and the expected lifespan of the data. Most criteria include three levels to enable comparative classification. The redundancy criterion is the only exception with two levels because it is formulated as a yes-no decision. **Table 2** provides the exact definition of each criterion.

Figure 3. Phases of the data lifecycle in the Metaverse based on [14], extended by Phase 5 "Virtual Representation & Interaction." (Authors´ own illustration.)

Table 2. Storage requirement criteria with definitions and categories. (Authors´ own table.)

ID	Criterion	Description	Levels
SC1	Frequency of writing	How often are new data produced or existing data modified?	once occasionally continious
SC2	Access frequency	How often are the stored data accessed or used?	rarely occasionally continious
SC3	Latency	How quickly must the data be available?	offline delayed real time
SC4	Data volume	How many units of a data category typically occur?	low moderate high
SC5	Sensitivity	How sensitive is the data with respect to confidentiality?	non-sensitive restricted confidential
SC6	Redundancy	Is it necessary to replicate the data to mitigate failures?	yes no
SC7	Retention period	How long is it necessary to store the data after it is created?	< 1 hour 1 hour - 1 year > 1 year

4. Results

In the following, the results of all steps of the research process depicted in Figure 2 are presented.

4.1. Key technologies

As explained in Section 3, the systematic literature research identified seven Metaverse key technologies, which are explained in further detail in the following.

4.1.1. IoT

Originally, the Internet of Things was designed to capture data from the physical world as comprehensively as possible [18,19]. In the context of the Metaverse, IoT devices function primarily as perception systems [17]. Equipped with sensors, cameras, and similar components, they act as sensory organs that continuously collect data from the real environment [20].

4.1.2. XR

Extended Reality (XR) is an umbrella term for immersive technologies such as virtual reality (VR), augmented reality (AR), and mixed reality (MR), which connect digital and physical worlds in different ways [21,22]. Although XR and IoT have different technical focuses, they cannot always be clearly separated in practice. Many devices now combine functionalities of both technologies and act as human–machine interfaces (HMI), serving as an interactive bridge between users and their technical surroundings [13,18]. Users navigate XR environments using specialized wearables such as headsets and smart glasses [12], as well as through gestures and voice commands. Current development efforts focus on improving resolution, reducing device weight, and lowering costs [9].

4.1.3. Digital Twins

Digital twins represent digital replicas of real-world objects, systems, processes, and humans [4,18]. A digital twin system can be divided into three components: the physical object in the real world, its virtual counterpart in the digital environment, and the data streams that connect both layers and flow bidirectionally between them [23]. Through this data connection, digital counterparts are continuously updated with real-time information [18]. As with IoT devices, the underlying data originates from multiple sources, such as various IoT sensors or user inputs.

4.1.4. AI

AI is considered the driving force of the Metaverse. It refers to the scientific discipline of training computer systems to simulate human intelligence, including abilities such as learning, creativity, and decision-making, as well as image and speech recognition, which are already widely used in Metaverse applications [4,13]. In the context of the Metaverse, AI plays a central role in generating, analyzing, and personalizing content and user experiences [4].

4.1.5. Blockchain

Blockchain technology forms the economic foundation of the Metaverse [4,14,21]. It enables secure, decentralized, and transparent transactions [10,24,25], allowing digital assets to be exchanged and valued.

Future blockchain developments in the Metaverse aim to increase autonomy, intelligence, and scalability. Relevant trends include agent-based intelligent systems, blockchains based on biological identity features, self-sovereign identity (SSI), and non-fungible tokens (NFTs) [13]. An NFT verifies the authenticity of a digital asset and assigns it to a specific owner [4,9]. Examples include digital artworks, music, virtual land, in-game items, or virtual clothing.

4.1.6. Communication Networks

5G communication networks form the foundation for the high-speed, low-latency data transmission required by immersive Metaverse technologies such as XR [9,18]. In addition, 5G ensures stable connectivity for edge devices that continuously capture and transmit real-time data. Without this network infrastructure, high-quality 3D content and interactions within virtual environments could not be delivered with the necessary fidelity and responsiveness [9].

While 5G primarily focuses on high data throughput, the development of 6G aims for a much broader paradigm. 6G seeks to integrate communication, sensing, computing, and storage into a unified information system [11]. Its expected data transmission speed is projected to be 50 times that of 5G and roughly 1,000 times that of 4G. South Korea plans to introduce 6G commercially as early as 2028 to support data-intensive applications such as the Metaverse [9].

4.1.7. Cloud- and edge computing

Modern cloud architectures distinguish compute and storage in the cloud from data collection or presentation on client devices. For Metaverse applications, however, this architectural principle falls short because it does not take into account the processing capabilities of end-user devices [4,17]. IoT and XR devices now also include powerful processors and can perform parts of the data processing independently [4,26]. As a result, data are increasingly processed directly at the point of origin—an approach referred to as edge computing [18,27]. This creates an architecture in which processors such as CPUs, GPUs, TPUs, FPGAs, and ASICs are distributed across the entire network rather than being operated exclusively in centralized data centers [4,18]. Decentralized processing reduces transmitted data volumes, lowers latency, and enables real-time execution of compute-intensive applications [11,26]. In the Metaverse, in particular, dynamic resource orchestration is required to efficiently distribute compute resources between the edge and the cloud [11,18].

4.2. Framework

Step 2 and 3 of the research process (see Figure 2) aim at proposing a framework of Metaverse technologies and data origins. It is depicted in **Figure 4** and incorporates the seven key technologies outlined in Section 4.1. Following Zhang et al. (2023) and Ates et al. (2023), the physical and virtual spheres are considered separately [11,12]. From the models of Jung and Jeon (2022) and Casale-Brunet et al. (2023), the clear distinction between technical and interactive components is adopted [9,10]. The components of the Metaverse are structured into three layers: infrastructure, devices, and content.

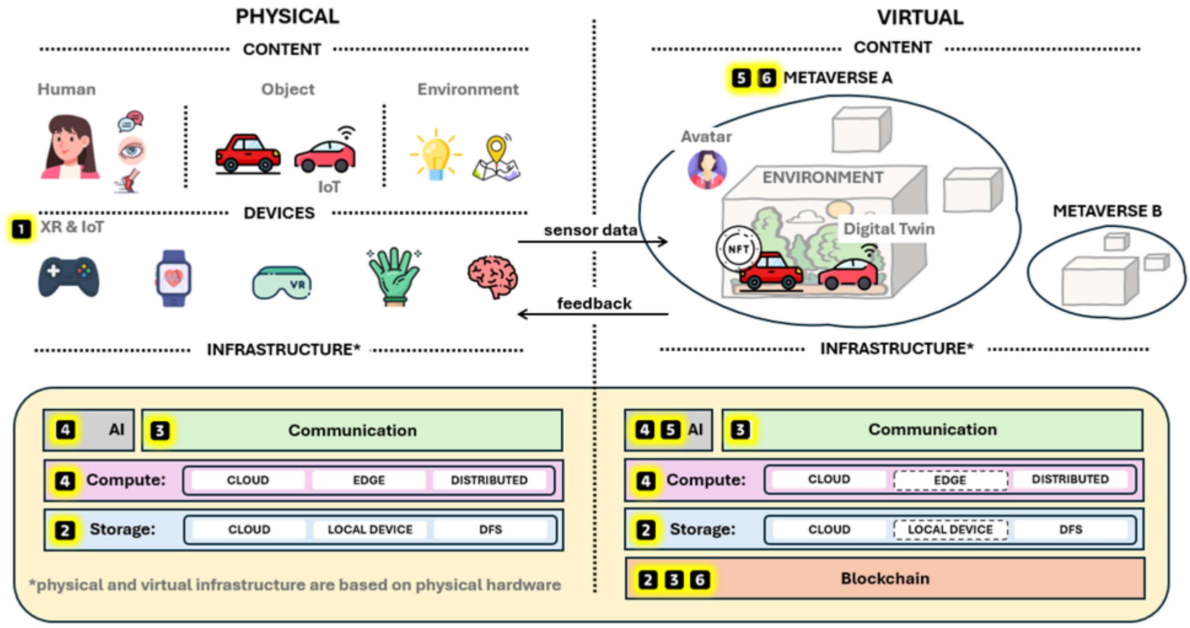


Figure 4. Framework locating key Metaverse technologies and linking data-lifecycle phases to functional components across physical and virtual space. (Authors´ own illustration. Icons by various authors – Flaticon www.flaticon.com).

The **content layer** on the physical side represents the starting point for real-world data that can be captured. Three physical data sources are distinguished: humans, objects, and the environment. Humans serve as a source of motion data, gestures, speech, and biometric attributes. Objects are items observed in the Metaverse and can be represented as digital twins in the virtual space. Connected objects, such as networked vehicles, are also included here. While such objects may be equipped with sensors, a distinction is made between their role as physical entities to be virtually represented and their role as data-capturing devices. When considered within the observed real-world domain, they are classified as objects in the content layer. When functioning primarily as sensor-equipped interfaces for data acquisition, they are assigned to the device layer as IoT devices. The third data source in the physical content layer is the environment. It describes the physical context in which humans and objects exist—factors that influence them but are not part of them. Examples include light exposure, location, and weather data. Virtual content refers to everything that can be "seen" or "experienced" within the Metaverse. Users perceive visual information such as 3D environments, avatars, and digital twins, as well as auditory stimuli generated through simulated sounds and speech. Functional elements such as digital assets and NFTs are also assigned to the virtual content domain.

The **device layer** comprises all devices in the Metaverse that serve as interfaces for capturing or transmitting data from the physical content domain. In addition to XR devices such as VR controllers, headsets, and gloves, this layer includes IoT devices specifically designed to collect data through sensors. An example is the Apple Watch, which measures activity and physiological data. Such information can be used in the Metaverse to adapt content dynamically. In a virtual fitness game, for instance, the difficulty level may adjust based on a user's current heart rate.

Since no physical devices exist in virtual environments to measure or capture data, this layer is absent on the right-hand side of the framework. However, a dynamic exchange occurs between sensor data transmitted from physical devices to the virtual environment and feedback signals that return from the virtual world to the user [14]. While Metaverse devices primarily capture physical data, they respond to virtual events, such as avatar interactions.

The **infrastructure layer** is understood, following Ates et al. (2023), as the technical connector between both spheres [12]. It encompasses key technologies across AI, communication, compute, and storage. Although the hardware supporting these technologies resides in the physical world, it processes and stores data originating from both environments. The origin of the content determines how the infrastructure is applied.

Physically captured data are processed and stored both locally on Metaverse devices and within central cloud systems. Virtual content, by contrast, is primarily managed in the cloud. Local device storage is used only to a limited extent—for example, when frequently visited virtual environments are cached locally to enable offline access. Additionally, the Metaverse uses blockchain technology within the virtual sphere to securely manage NFTs and verify their authenticity.

To enable a coherent derivation of the Metaverse data categories, the phases of the data lifecycle (see Figure 3) are mapped onto the framework. This mapping reveals where data are generated, processed, and stored in both physical and virtual spaces. The numbered elements in Figure 3 indicate the specific lifecycle phase that occurs at each point.

4.3. Data Categories

The Metaverse can be understood as a continuously growing source of data [14]. Based on the developed framework, two central points of data generation can be distinguished: raw data originating from the three physical sources—human, object, and environment—and data generated exclusively within the virtual sphere. Beginning with the physical sphere, each layer of the framework is revisited in the following section, this time with a focus on the data that arises there throughout the lifecycle. These data are presented in an organized manner, grouped into categories.

4.3.1. Physical content layer

In the physical content layer, the data pool is purely potential, as concrete raw data emerge only when states and movements of corresponding devices are captured. Consequently, the first phase of the data lifecycle—data collection—begins only at the device layer. With slight adjustments, the data categories used in robotics by Herath et al. (2024) can be applied to humans as a physical data source in the Metaverse [21].

Biometric data refer to individually identifiable physical characteristics that can be used to identify or authenticate a person. These include facial features, fingerprints, iris patterns, and characteristic voice profiles [21]. They are mostly static or change only very slowly. **Physiological data**, in contrast, describe variable bodily parameters such as heart rate, blood pressure, body temperature, respiratory rate, or oxygen saturation. They are particularly relevant for sports and medical applications, such as health monitoring [26]. While physiological data captures internal bodily states, **motion data** focuses on external movements. They form the basis for spatial interaction and motion tracking. At this point, the categorization used in robotics is adapted by treating **positional data** as part of motion data. This includes spatial position [13], head and hand movements, body posture, gestures, and gait patterns. These data also allow inferences about a person and are sometimes categorized as biometric data in the literature [21].

Behavioral and interaction data extend beyond pure motion information and concentrate on user behavior. They include gaze direction, reaction times, device usage intensity, and input modalities such as touch, speech, or gestures [13,28]. Such information can reveal what users pay particular attention to [28], allowing conclusions about personalization [4] and usability of specific interfaces.

When considering physical objects or the environment as data sources, the literature still lacks a structured classification of Metaverse-related data [18]. **Object recognition data** refer to visual characteristics such as shape, dimensions, color, and surface texture. These data are mostly static because the underlying identification features rarely change. They are captured through camera systems or by scanning barcodes, QR codes, serial numbers, or RFID tags [4]. If the object is movable, further subdivision into **object tracking data** is possible. These dynamic data are also captured by visual tracking systems or integrated sensors and document movements and spatial position over time. In certain domains, such as industry, logistics, or medical technology, **machine condition data** can also be relevant. These data provide information on operational status, error messages, and measurements that change continuously during use. Together, these three data types form the basis for digital twins, whose state is represented in real time in the Metaverse.

To reproduce real-world environmental conditions in the Metaverse, environmental data must be measured. These are predominantly **physical measurements**, such as temperature, light exposure, and humidity, which describe a specific state. They are typically recorded continuously and can change within seconds. Analogous to humans and objects, environmental movements can be captured as well. These include highly variable phenomena such as wave and plant movements, shifting shadows, the bending of a bridge in the wind, or seismic vibrations. Such **environmental dynamics** are recorded using distance, acceleration, infrared, ultrasound, or radar sensors. Another category is **positional data** [13]. These data document the exact location and geographic coordinates of an object, for example, via GPS. Here, positional data are intentionally assigned to the environment because they describe the spatial context in which humans and objects exist, rather than their intrinsic properties. They enable precise spatial alignment in the Metaverse and support navigation and synchronization between the physical and virtual worlds.

4.3.2. Device layer

The device layer marks the starting point of the data lifecycle. Here, all previously described data categories from the content layer are captured as raw data. In addition to these content-related datasets, the Metaverse device itself generates further technical datasets. These include **metadata** that provide contextual information about the raw data, such as the capture time, device location, calibration details, or the file formats used. Some of these metadata remain permanently linked to the raw data to support later interpretation and analysis. Other metadata—such as current signal quality or temporary session IDs—are relevant only for a short period and are continuously overwritten during operation.

Device diagnostic data are generated in parallel. Their content is similar to the machine condition data described above, but they originate not from the observed object itself, but from the Metaverse device in use. They provide information on the device's operational status, battery level, or error messages. While these data are secondary to the content experience in the Metaverse, they are essential for stable technical operation. In most cases, such data are transient and are regularly replaced by updated status information. In exceptional cases—for example, for maintenance or error analysis—they may also be stored permanently. Some device diagnostic data is also displayed within the virtual sphere, such as an on-screen warning when the battery level is low.

4.3.3. Infrastructure layer for physical-sphere data

In the infrastructure layer, several phases of the data lifecycle take place. Phase 2 refers to the storage of the various data categories. It begins immediately after data collection on the local device's memory and extends to additional storage architectures, such as cloud data centers. In Phase 3, data are moved between these storage architectures and transferred from the physical world into the virtual one. This occurs through high-performance communication technologies such as 5G – the key technology introduced in Section 4.1 – as well as Wi-Fi and satellite-based connections. In part, Phase 4 of the lifecycle is already anticipated at the device layer, where edge-computing technologies preprocess raw data directly on the end device [26]. This may involve compressing large files or filtering sensor data before transmission to the virtual sphere [17]. However, most data processing occurs in the infrastructure layer on more powerful server resources, often supported by AI models.

Two cross-cutting data categories emerge from these phases. First, **system and operational data** are produced across all infrastructure components. They are generated to monitor, control, and maintain the infrastructure's technical operations. These data document and safeguard the infrastructure's technical state and include, for example, access logs, bandwidth usage, signal strength, CPU load, and backup logs. A further subdivision of these data by specific infrastructure domains, e.g., storage and network, is not supported by the reviewed literature.

Second, **processed data** forms an essential basis for content within the Metaverse. These data result from transforming and integrating raw data to derive meaningful information. In Phase 4, data from multiple sources are combined and interpreted. In the Metaverse, it is not sufficient to harmonize heterogeneous data streams. Instead, the focus lies on reconstructing events based on these data [4]. For instance, combining a person’s motion, positional, and interaction data can reveal that Avatar A picked up a virtual cup from a table. Processed data include compressed video streams, aggregated sensor measurements, analytical results, or detected objects derived from image data. Raw data from the physical content layer is often processed multiple times, resulting in multiple versions of the same data. Moreover, data may be used for different purposes, leading to varying formats or degrees of preprocessing.

4.3.4. Virtual content layer

In the virtual content layer, every digital element becomes both a data point and a potential data source from the moment it is created. Unlike in the physical world, there is no need to capture data through external devices. Every action automatically leaves a digital trace, ensuring that no event remains undocumented [21]. This continuous system recording results in a highly detailed data foundation and marks the point of data creation within the virtual sphere.

At this point, Phase 5 of the data lifecycle begins: digital representation and interaction. In this phase, the processed data from previous steps is used to generate and further develop digital content. The data generated here can be divided into two main categories: representation data and interaction data, although some data types serve both purposes.

Representation data assigns digital equivalents to the original data sources—human, object, and environment. Human-related data give rise to the **digital identity**, which includes avatar appearance and predefined profile information. Object-related data becomes **digital assets**, such as virtual clothing, tools, vehicles, or buildings. **Environment-related data** describes how a city, forest, or room should be represented virtually. The interpreted raw data from the physical world are not only displayed but also used in real time to control their virtual representations.

At the same time, the virtual content layer is not solely dependent on real-world input. New content can be generated directly within the digital environment. Generative AI technologies can create not only text and images but entire scenarios within the Metaverse. Although the underlying AI models are trained on real-world data, they no longer depend on it to generate synthetic content.

Digital interaction data arises whenever users, virtual objects, and systems interact. These data can be grouped into three categories based on the actors involved.

The first group covers interactions between multiple users. Here, **communication data**, such as chat messages, voice transmissions, video streams, or avatar gestures, is generated. Equally important in the Metaverse are **coordination data**, which allow avatars to remain synchronized and move in relation to one another—for example, through the alignment of avatar movements or gaze directions.

The second group includes interactions between users and virtual objects. **Usage data** record how and when a user performs specific actions, such as switching on a virtual lamp or opening a virtual door. When an object is purchased, sold, or exchanged, **transaction data** is created. In the case of NFT-based digital assets, these are stored on a blockchain.

The third group of digital interaction data relates to interactions between systems and users. These include **feedback data**, which the system returns as visual, haptic, or auditory responses—for example, an object lighting up when selected or a click sound signaling a completed action. In addition to feedback data, **this group includes control data**. While usage data merely records the instruction (e.g., that a door should be opened), control data contains the technical parameters and commands through which the system executes the action.

Within the virtual sphere, all of the above data categories are accompanied by **metadata**. Although not part of the content itself, metadata forms the foundation for rights management and data administration. Static representations contain information about origin, ownership, usage rights, format, and version. For dynamic interaction data, metadata such as timestamps or involved actors are recorded. Metadata plays a crucial role in enabling virtual representations to be used across different Metaverse environments. Building on this, Phase 6 of the data lifecycle begins, in which digital assets are shared, traded, or used across platforms.

4.3.5. Infrastructure layer for virtual-sphere data

The infrastructure for virtual data follows the same principles as for physical data. Virtual content generated in Phase 5 must also be stored, transferred, and processed. It therefore passes through Phases 2 to 4 of the data lifecycle as well. As with physical data, new processed data and new system and operational data are generated in these phases.

A technological particularity is the use of blockchain technology. It plays a connecting role across the phases of storage (Phase 2), transfer (Phase 3), and sharing (Phase 6). In the storage phase, it records transactions and state changes transparently and permanently. In the transfer phase, the blockchain assumes a supportive function. Unlike classical communication technologies, it does not exchange data; instead, it ensures that transactions are authentic and unaltered.

The deletion phase (Phase 7) is not considered further in this context, as it generates no new data.

In total, this analysis identifies 23 data categories that arise across the different phases of the data lifecycle in the Metaverse. **Table 3** provides a summarized overview. Although metadata appear at several points within the Metaverse, they are treated as a single data category in the following analysis.

Table 3. Overview of functional data categories. (Authors´ own illustration.)

Sphere*	Layer	Main category	Relevant phases in data life cycle	Data Category
P	Content	Human-related data	Data source for phase 1	Biometric data Physiological data Motion and posture data Behavioral and interaction data
P	Content	Object-related data	Data source for phase 1	Object recognition data Object tracking data Machine condition data
P	Content	Environment-related data	Data source for phase 1	Physical measurement data Environmental dynamics Positional data
P	Device	Content data	1	Captured raw data from the physical content layer mentioned above
P	Device	Technical data	1,4	Metadata Device diagnostic data
P	Infrastructure	Component-independent data	2,3,4	System and operational data Processed data
V	Content	Representations	5	Digital identities Digital assets Virtual environment data Metadata
V	Content	User-to-user interaction	5,6	Communication data Coordination data Metadata

Continuation Table:

Sphere*	Layer	Main category	Relevant phases in data life cycle	Data Category
V	Content	User-to-object interaction	5,6	Usage data Transaction data Metadata
V	Content	System-to-user interaction	5,6	Feedback data Control data Metadata
V	Infrastructure	Blockchain	2,3,6	Metadata

*(P = physical, V = virtual)

4.4. Storage Requirements

Based on the seven storage requirement criteria presented in Table 2, the 23 identified data categories are evaluated and compared in a matrix shown in **Table 4**.

The assessment of the data categories depends strongly on the specific application and target group, so more than one category may apply in some cases. Because no comparable assessments are available in the literature, this study uses informed estimates. To enhance transparency, selected entries from the matrix are explained in more detail in the following section.

4.4.1. Frequency of writing (SC1)

The frequency of writing (SC1) depends primarily on whether the data are static or dynamic. Static data are rarely modified after their initial generation. This category includes biometric, object-recognition, and transaction data, which are typically captured once. Their virtual counterparts, such as digital identities and digital assets, are also created once but can be updated when needed, for example, when a user adjusts an avatar or purchases a new virtual item. Feedback data are produced intermittently and triggered by specific interactions, so they occur often but not continuously. All remaining categories are classified as dynamic data. These data are continuously updated or newly generated, and include data categories such as object tracking, positional, and usage data.

Table 4. Evaluation of data categories based on storage requirement criteria (SC). Author's own illustration.

Data category	SC1 Frequency of writing	SC2 Access frequency	SC3 Latency	SC4 Data volume	SC5 Sensitivity	SC6 Redun-dancy	SC7 Retention period
Biometric data	once	continious	real time	low	confidential	yes	> 1 year
Physiological data	continious	continious	real time	moderate	confidential	yes	1 hour - 1 year
Motion and posture data	continious	continious	real time	high	restricted	no	< 1 hour
Behavioral and interaction data	continious	occasionally	delayed	high	restricted	yes	> 1 year
Object recognition data	once	occasionally	delayed	moderate	restricted	yes	> 1 year
Object tracking data	continious	continious	real time	high	restricted	no	< 1 hour
Machine condition data	continious	continious	real time	moderate	restricted	yes	1 hour - 1 year

Continuation Table:

Data category	**SC1 Frequency of writing**	**SC2 Access frequency**	**SC3 Latency**	**SC4 Data volume**	**SC5 Sensitivity**	**SC6 Redun-dancy**	**SC7 Retention period**
Physical measurement data	continious	continious	real time	high	non-sensitive	no	1 hour - 1 year
Environmental dynamics	continious	continious	real time	high	non-sensitive	no	< 1 hour
Positional data	continious	continious	real time	high	restricted	no	< 1 hour
Metadata	continious			high	restricted	yes	> 1 year
Device diagnostic data	continious	continious	real time	moderate	restricted	yes	1 hour - 1 year
System and operational data	continious	continious	real time	moderate	restricted	yes	1 hour - 1 year
Processed data	continious	continious	real time	high	restricted	yes	1 hour - 1 year
Digital identities	occasionally		real time	low	confidential	yes	> 1 year
Digital assets	occasionally	continious	real time	moderate	restricted	yes	> 1 year
Virtual environment data	continious	continious	real time	high	restricted	yes	> 1 year
Communication data	continious	continious	real time	high	confidential	yes	1 hour - 1 year
Coordination data	continious	continious	real time	high	restricted	no	< 1 hour
Usage data	continious	occasionally	delayed	moderate	restricted	yes	> 1 year
Transaction data	once	occasionally	delayed	moderate	non-sensitive	yes	> 1 year
Feedback data	occasionally	continious	real time	moderate	non-sensitive	no	< 1 hour
Control data	continious	continious	real time	high	restricted	yes	< 1 hour

4.4.2. Access frequency (SC2)

The access frequency (SC2) was not rated as low for any of the 23 data categories. As with SC1, most dynamic data categories are accessed continuously. This is due to the need for content to be permanently available in the Metaverse to maintain the immersive experience. Digital identities persist in the background but are accessed only during specific events, such as authorization checks. If they take the form of a human digital twin, however, they remain constantly active as long as the user moves within the Metaverse. Some data categories, such as metadata, are accessed only intermittently. Although they are recorded automatically, they are retrieved only when needed, for example, to locate or interpret raw data. The same applies to behavioral and interaction data, user data, and transaction data, which are accessed only in specific situations. Biometric data, by contrast, are continuously accessed for authentication, while object recognition data are retrieved only intermittently, for instance, during inventory processes in industrial settings.

4.4.3. Latency (SC3)

Latency (SC3) varies with the importance of the data to the user experience and system operation. Some data categories, such as behavioral and interaction data, metadata, and usage data, are not critical for immersion and can therefore be delivered with a delay. For transaction data and object recognition data, accurate and verifiable recording is the main priority. As long as transactions are processed reliably, a delay is acceptable here as well. In contrast, continuously used and reactive data streams must be available in real time. They must be retrieved within fractions of a second, as even minor delays can disrupt interactions. This applies to object-tracking, coordination, and feedback data. Real-time availability is essential not only for maintaining an uninterrupted user experience but also for technical stability. System and operational data and device diagnostic data, therefore, require real-time access as well to ensure the functionality of the Metaverse.

4.4.4. Data volume (SC4)

The data volume (SC4) is rated as very high for most data categories. This applies to data captured continuously by sensors in the physical world, such as motion and posture data, physical measurements, and environmental dynamics. Processed data also contributes substantially to the overall data volume in the Metaverse, as numerous intermediate products are generated during processing. In contrast, the volume of technical data from system operations or device diagnostics is considered lower. Metadata typically contains descriptive information and is usually generated alongside other data categories, such as transaction data. The volume of transaction data varies significantly with user activity and can range from a few entries to a very large number. Static categories such as object recognition data, digital assets, and digital identities produce only a limited number of data points because they are rarely modified or updated. Biometric data and digital identities fall into the lowest-volume categories, as they are typically captured once per user and accessed only occasionally.

4.4.5. Sensitivity (SC5)

The sensitivity (SC5) of data in the Metaverse is generally assessed as high. Personal data is particularly sensitive, including biometric and physiological data, digital identities, and communication data. They contain confidential information and may cause severe harm in cases of identity theft. Data categories without a personal reference, by contrast, are considered low risk. These include physical measurements, transaction data, feedback data, and environmental dynamics. The latter may, however, require increased protection if they contain information from private spaces. The same applies to object recognition data when they allow conclusions about private property or company-owned machinery. All remaining data categories require restricted access, meaning that most data in the Metaverse must be considered sensitive. This includes digital assets, positional data, and system and operational data. They enable indirect inferences about user behavior, system states, or critical processes and are therefore susceptible to misuse.

4.4.6. Redundancy (SC6)

A high redundancy requirement (SC6) can be seen as an indicator that certain data are particularly relevant to either users or Metaverse providers. Losing this data would have serious consequences. Personal or system-critical data should be stored redundantly. Hard-to-reconstruct data, such as transaction or communication data, can also be important from a user perspective. To ensure traceability, backups of this data should also be created. Little or no redundancy is required for data that can be easily captured again or for which individual values have limited relevance. This includes motion and posture data, environmental dynamics, and feedback data. An exception applies to physical measurements when a non-repeatable natural event is documented.

4.4.7. Retention period (SC7)

The retention period (SC7) is often closely linked to the redundancy requirement.

The more important the data are for future access and analysis, the longer they generally need to be stored. Examples include biometric data, digital assets, behavioral and interaction data, and usage data. In contrast, some data categories consist of individual data points that are processed immediately after capture. These data are not accessed again and therefore do not require long-term storage. In many cases, each generated data point is overwritten by the next one. Such short-lived data include positional, coordination, and certain usage data. Between these two extremes are data categories with a medium retention period. Physiological data, for example, helps users identify anomalies. When longer time series, such as heart-rate patterns, are analyzed, they can inform personalized adjustments of workload levels during virtual training sessions. Data in this medium category is neither overwritten immediately nor archived permanently; their informational value decreases over time. Additional examples include system and operational data, machine condition data, and device diagnostic data.

For long-term trend analyses, extended storage can also be helpful for categories originally classified as short- or medium-term. The classification is highly context-dependent.

5. Application to DNA data storage

Deoxyribonucleic acid (DNA) is the primary source of information for the structure and function of living organisms [29] and is therefore described as the carrier or storage medium of genetic information [30,31]. At present, DNA is considered the most compact and durable data storage medium available [31]. It represents a cross-fusion technology at the interface of IT and biotechnology [32]. These characteristics make DNA a promising medium for the long-term storage of digital information.

From a biological perspective, DNA is classified as a macromolecule, meaning a very large molecule composed of many individual building blocks [33,34]. Each of these so-called nucleotides (nt) comprises one of four bases: adenine (A), thymine (T), cytosine (C), or guanine (G) [29,34]. The sequence of these four bases along a single DNA strand carries the genetic information of an organism and makes DNA an exceptionally efficient storage medium [33] – not only in nature but also for technical applications. A single nucleotide is approximately 0.3 nanometers in size and represents the smallest unit of genetic information [31]. The helical outer structure is shaped by a backbone of alternating deoxyribose and phosphate groups [32]. This backbone does not itself carry genetic information. The actual information lies in the space between the two strands, where the four bases pair through hydrogen bonds [35]. The quaternary genetic code inspired the concept of DNA data storage, which builds on the idea that any binary information can be encoded in a DNA sequence composed of A, T, C, and G [7].

DNA exhibits several characteristics that make it an attractive medium for data storage. **Table 5** compares key properties of DNA data storage to other technologies. The most significant potential of DNA data storage lies in its far higher storage density [33,36] and capacity compared with traditional storage technologies [5]. The entire human genome, with its approximately 3.2 billion base pairs—corresponding to 6.4 billion nucleotides—fits inside the nucleus of a single human cell [8]. The literature reports widely differing estimates of the theoretical storage density per gram of DNA, ranging from about 215 GB (2.15×10^{11} bytes) per gram of DNA [5] to 2×10^{21} bytes per gram [37]. These substantial differences arise primarily because some studies calculate theoretical maxima, while others incorporate experimental conditions and practical limitations. Nevertheless, even conservative estimates show that DNA enables storage densities far beyond those of the most advanced conventional technologies.

An additional advantage is DNA's extreme stability, as demonstrated by up to 300,000-year-old archaeological samples [8,38,39]. Compared to other data storage technologies, DNA does not require copying data at the end of the life span. Furthermore, it is considered environmentally friendly as it remains stable without an energy supply and can be synthesized without significant use of natural resources [8,30,32,39–41]. Finally, DNA is considered a highly secure storage medium [32]. Once stored, information remains readable as long as the corresponding sequences are not physically or chemically destroyed. Built-in error-correction mechanisms, such as intentionally inserted redundant segments, protect against data loss and ensure long-term integrity [8,27]. As an offline medium, DNA is independent of digital networks and thus protected from cyberattacks. Access to the stored data requires physical possession of the DNA sample, which adds an additional security barrier.

On the other hand, DNA data storage comes with drawbacks compared to other data storage media. The synthesis of synthetically encoded DNA remains a major technological bottleneck. Current challenges primarily relate to three areas. The cost of synthesis remains exceptionally high. Overall processing speed — especially writing — is comparatively low [8,30,42]. Third, efficient rewriting mechanisms are lacking, making it difficult to modify or overwrite data once it is stored [43].

Table 5. Comparison of data storage technologies. (Authors´ own table with reference to [5,7,30,31,44–48].)

	Optical	**Hard Disc**	**Solid State**	**DNA**
Storage density	8×10^{14} bit/cm^3	10^{13} bit/cm^3	10^{14} bit/cm^3	10^{19} bit/cm^3
Typical storage capacity per unit	~ 128 GB	~ 10 TB	~ 1 TB	> 100 EB
Life span	~ 10 – 25 years	~ 5 – 10 years	~ 10 – 15 years	> 1000 years
Write Speed (order of magnitude)	MB/s	GB/s	GB/s	MB/h
Cost	< 100 USD/TB	< 100 USD/TB	< 100 USD/TB	~800 USD/MB

The potentials and challenges of DNA data storage will now be compared with the storage requirements of the identified data categories. This comparison allows the potential application areas of DNA data storage in the Metaverse to be derived. The following paragraphs examine each storage requirement criterion and identify the most suitable category for DNA data storage.

At present, the writing and reading processes for DNA data storage are not automated and remain highly time-consuming. Moreover, data stored in DNA cannot be modified afterward. Changes must therefore be applied externally and then written back into DNA, requiring a repeat of the costly synthesis step. For this reason, data categories with low frequency of writing (SC1) and low access frequency (SC2) are more suitable candidates. In addition, DNA-stored data cannot be retrieved in real time with today’s technologies. Consequently, data categories evaluated with the latency (SC3) options “offline” or “delayed” are most appropriate for further consideration.

By contrast, the data volume considered under SC4 imposes virtually no limitations. The storage capacity of DNA far exceeds that of current technologies, allowing even huge data volumes to be stored without difficulty. DNA is also regarded as a highly secure storage medium. Data categories classified as "confidential" under SC5 are therefore particularly well suited for DNA storage. The most important criterion for assessing the applicability of DNA in the Metaverse is SC6. The literature primarily praises DNA as a medium for "cold" data – data that change infrequently and are accessed infrequently [30,31,48]. Archiving and backup storage, therefore, represent the prime examples of DNA data storage applications.

Maintaining DNA is comparatively inexpensive and environmentally friendly. Nevertheless, the costs of writing data to DNA must be weighed against the costs of maintaining it in conventional storage. The longer the retention period (SC7) of a data category, the more economical DNA storage becomes.

A classification by file format or degree of structure is irrelevant for answering the research question. Because DNA storage is based on binary data, any data format – such as text, audio, or image—and any level of structure can be stored.

Comparing the above-explained properties of DNA data storage to the requirements of Metaverse data categories (see Table 4) reveals that only "object recognition data" and "transaction data" might be potential applications for DNA data storage. Moreover, neither of these categories would benefit from DNA's very high storage capacity, as their data volumes are not considered very large. As a result, the advantages over conventional storage technologies would not be fully realized, and economic viability would remain questionable. Unless the challenges related to financial feasibility and faster data access are significantly reduced in the coming years, DNA storage is unlikely to be used as the sole storage solution for any Metaverse data category.

The assessment differs when DNA data storage is considered as a complementary technology alongside conventional storage systems. In many data categories that generate very large volumes, such as behavioural and interaction data, metadata, virtual environment data, or communication data, the redundancy requirement (SC6) is also very high. For precisely these large-scale archiving needs, many authors view DNA as a promising solution. DNA could alleviate pressure on existing storage technologies by storing backups and archives of large volumes of important or sensitive information. This would free up additional capacity for short-lived data.

In summary, DNA data storage is unlikely to replace current storage methods soon. Hybrid solutions are possible, where cloud services manage data needing quick access, and DNA is used for long-term archiving and backups. It's also worth noting that progress in this field is rapid. As the common limitations and current constraints diminish, the potential uses for DNA storage within the Metaverse could grow significantly.

6. Discussion

In the context of this study, the metaverse is understood as a further development of the internet, in which the boundaries between physical and virtual worlds are becoming increasingly blurred. It is available anytime, anywhere, and features immersive experiences, social interactions, personalized content, and social rules. Digital identities and a digital economy are also defining elements, supported by low computing and transmission latency and persistent structures. While research has already focused on related technologies, the data perspective, with its origins and requirements, was not well understood.

This study conducted a systematic literature review to identify key Metaverse technologies, including IoT, XR, digital twin, AI, blockchain, 5G, and cloud and edge computing. Depending on their relationship to the physical or virtual sphere, these technologies can be functionally divided into three levels: content, device, and infrastructure. This resulted in a framework for the metaverse's components.

Applying the seven phases of the data life cycle to this regulatory framework revealed two central points of data creation: on the one hand, data is created physically by people, objects, and the environment; on the other hand, it is made virtually through digital representations and interactions. In this work, 23 data categories were identified, and their storage requirements were analyzed based on the following seven selected criteria: (SC1) frequency of writing, (SC2) access frequency, (SC3) latency, (SC4) data volume, (SC5) sensitivity, (SC6) redundancy, and (SC7) retention period.

Thus, the study contributes to research by offering the first systematic classification of Metaverse data. This classification clarifies which data categories emerge, how they relate to underlying technologies, and which requirements they impose on storage systems. As Metaverse data volumes are expected to grow substantially, existing storage technologies will reach technical and economic limits. An application use case of the study results might therefore be the assessment of alternative storage approaches. The analysis supports this by establishing a basis for evaluating the suitability of emerging technologies such as DNA data storage. The classification also provides practical value, offering organisations and developers a structured overview of relevant data categories and their associated storage requirements. Developers of Metaverse platforms or designers of immersive experiences gain an overview of involved technologies and corresponding data from the Metaverse technology and data framework (see Figure 3). This helps make deliberate design decisions to fulfill data security and retention requirements, for example.

At the same time, the analysis has limitations. The systematic literature review aimed at creating a sound scientific basis for the analysis. Nevertheless, selecting relevant literature remains subjective. The inclusion and exclusion criteria allow interpretive flexibility, as does the decision about which information from the sources to include. The thematic categorisation of this information is also affected by individual judgment and is therefore not fully traceable.

The chosen search term combination “Metaverse” AND “data” focused on data-related literature, allowing the results to be systematically narrowed. However, this restriction limited the thematic breadth and depth of the analysis. The reviewed literature did not provide systematic listings of data categories, data sources, or Metaverse components, necessitating the development of an independent methodological approach. The data categories were therefore derived from the functional composition of those key technologies most frequently mentioned in the reviewed sources. This selection is not exhaustive and reflects the technological emphasis of the included publications. Although the study focused on data, additional data categories could be identified by incorporating further technologies or components of the Metaverse. Future studies may use alternative search combinations such as “(key) technology” AND “Metaverse” or “components” AND “Metaverse” to broaden the scope.

Another constraint is the inconsistent terminology used in the Metaverse domain, which makes it difficult to distinguish data categories and system components.

Consequently, studies are only partially comparable, and some important sources might be overlooked. Terms like "key technologies" and "data categories" are used differently across various publications. Developing a standardized Metaverse glossary could promote more consistent understanding and application of key concepts in future research.

In light of the aforementioned limitations in the research process, which relied to a large extent on a systematic literature review but also comprised subjective assessments, further validation can be a topic for future studies. These could include further case studies, beyond DNA data storage, or expert evaluation.

Future studies should also consider alternative approaches to data categorisation. For example, data could be derived from specific application domains such as education or healthcare. Additionally, future research should determine whether additional indicators beyond the seven criteria used in this study are relevant for assessing data storage technologies. The selected scale levels could also be refined. The example of retention period (SC7) shows that the economic threshold at which DNA storage becomes viable must first be determined. Only once it is known whether this threshold is reached after a few months or only after several years, an appropriate evaluation scale can be developed.

Data volumes in physical and digital environments are expected to rise exponentially as Metaverse participation increases. This amplifies the need for sustainable storage technologies. Although DNA data storage is not yet technologically mature, its potential compared with traditional storage approaches is already evident. Advancing research and reducing the costs of DNA synthesis are essential steps for its practical implementation.

A related future research direction concerns the role of AI as a key technology. AI significantly shapes the development of the Metaverse, generates new data categories, and enables more efficient storage solutions. These developments underscore the need to evaluate storage and processing technologies in combination rather than isolation.

Author contributions: Conceptualization, Weg and Wei; methodology, Weg and Wei; validation, Weg and Wei; formal analysis, Weg; investigation, Weg; data curation, Weg; writing—original draft preparation, Weg; writing—review and editing, Wei; visualization, Weg; supervision, Wei; project administration, Wei.

Funding: This research received no external funding.

Conflict of interest: The authors declare no conflict of interest.

References

1. Lee, L.-H. *All One Needs to Know about Metaverse*: *A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda*, 1st ed.; Now Publishers: Norwell, MA, **2024**, ISBN 978-1-63828-444-4.
2. Weinberger, M. What Is Metaverse?—A Definition Based on Qualitative Meta-Synthesis. *Future Internet* **2022**, *14*, 310, doi:10.3390/fi14110310.
3. Weinberger, M.; Gross, D. A comparison of virtual worlds based on the metaverse maturity model. *ITU J-FET* **2023**, *4*, 537–548, doi:10.52953/REVF5378.
4. Zhang, B.; Chen, G.; Ooi, B.C.; Shou, M.Z.; Tan, K.-L.; Tung, A.K.H.; Xiao, X.; Yip, J.W.L.; Zhang, M. Managing Metaverse Data Tsunami: Actionable Insights. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 7423–7441, doi:10.1109/TKDE.2024.3354960.
5. Şatır, E. A DNA Data Storage Method Using Spatial Encoding Based Lossless Compression. *Entropy (Basel)* **2024**, *26*, doi:10.3390/e26121116.
6. Anders S G Andrae. *Projecting the chiaroscuro of the electricity use of communication and computing from 2018 to 2030*,

2019. doi: 10.13140/RG.2.2.25103.02724
7. Dimopoulou, M.; Antonini, M. Data and image storage on synthetic DNA: existing solutions and challenges. *J Image Video Proc.* **2022**, *2022*, doi:10.1186/s13640-022-00600-x.
8. Akhtar, Z.B. Unraveling the Promise of Computing DNA Data Storage: An Investigative Analysis of Advancements, Challenges, Future Directions. *JAAI* **2024**, *2*, doi:10.18178/JAAI.2024.2.1.122-137.
9. Jung Sang Hee; Jeon In-oh. A study on the components of the Metaverse ecosystem. *Journal of Digital Convergence* **2022**, *20*, 163–174, doi:10.14400/JDC.2022.20.2.163.
10. Casale-Brunet, S.; Mattavelli, M.; Chiariglione, L. Exploring blockchain-based metaverses: Data collection and valuation of virtual lands using machine learning techniques. *Digital Business* **2023**, *3*, doi:10.1016/j.digbus.2023.100068.
11. Zhang, L.; Du, Q.; Lu, L.; Zhang, S. Overview of the Integration of Communications, Sensing, Computing, and Storage as Enabling Technologies for the Metaverse over 6G Networks. *Electronics* **2023**, *12*, doi:10.3390/electronics12173651.
12. Ates, E.C.; Bostanci, E.; Guzel, M.S. Evaluation of Open Source, Open Data Sharing, and Data Anonymization Concepts in the Development of the Metaverse. In *Metaverse*; Esen, F.S., Tinmaz, H., Singh, M., Eds.; Springer Nature Singapore: Singapore, **2023**; pp 83–98, ISBN 978-981-99-4640-2.
13. Zhang, H.; Lee, S.; Lu, Y.; Yu, X.; Lu, H. A Survey on Big Data Technologies and Their Applications to the Metaverse: Past, Current and Future. *Mathematics* **2023**, *11*, doi:10.3390/math11010096.
14. Maedeh Mosharraf. Data governance in metaverse: Addressing security threats and countermeasures across the data lifecycle. *Technology in Society* **2025**, *82*, doi:10.1016/j.techsoc.2025.102910.
15. Song, J. *Blockchain-Metaverse.*: *The Next Stop of the Internet.;* Beijing, China, Guosheng Securities, **2021**.
16. Zhao, G.; Yi, H.; Xu, Y. *Metaverse;* Chinese Translation Publishing House: Beijing, China, **2021**.
17. Wang, X.; Wang, Y.; Yang, J.; Jia, X.; Li, L.; Ding, W.; Wang, F.-Y. The survey on multi-source data fusion in cyber-physical-social systems: Foundational infrastructure for industrial metaverses and industries 5.0. *Information Fusion* **2024**, *107*, doi:10.1016/j.inffus.2024.102321.
18. Kumar, A.; Lovén, L.; Arshad, M.T.; Pirttikangas, S.; Tarkoma, S. Towards a data fabric framework for industrial metaverse integration. *Sustainable Computing: Informatics and Systems* **2025**, *47*, doi:10.1016/j.suscom.2025.101132.
19. Han, Y.; Niyato, D.; Leung, C.; Miao, C.; Kim, D.I. A Dynamic Resource Allocation Framework for Synchronizing Metaverse with IoT Service and Data. *Proceedings of the IEEE ICC, 2022*; pp 1196–1201.
20. Nazir, A.; He, J.; Zhu, N.; Wajahat, A.; Ma, X.; Ullah, F.; Qureshi, S.; Pathan, M.S. Advancing IoT security: A systematic review of machine learning approaches for the detection of IoT botnets. *Journal of King Saud University - Computer and Information Sciences* **2023**, *35*, doi:10.1016/j.jksuci.2023.101820.
21. Herath, H.M.K.K.M.B.; Herath, H.M.S.S.; Karunasena, G.M.K.B.; Yasakethu, S.L.P. Metaverse Meets Robotics: Addressing Data Protection and Privacy in Robotic Environment. In *Data Protection*; Hewage, C., Yasakethu, L., Jayakody, D.N.K., Eds.; Springer Nature Switzerland: Cham, **2024**; pp 181–211, ISBN 978-3-031-76472-1.
22. Ahmed, N.; Afyouni, I.; Dabool, H.; Al Aghbari, Z. A systemic survey of the Omniverse platform and its applications in data generation, simulation and metaverse. *Front. Comput. Sci.* **2024**, *6*, doi:10.3389/fcomp.2024.1423129.
23. Deng, Y.; Weng, Z.; Zhang, T. Metaverse-driven remote management solution for scene-based energy storage power stations. *Evol. Intel.* **2023**, *16*, 1521–1532, doi:10.1007/s12065-022-00769-0.
24. Ghosh, P.K.; Chakraborty, A.; Hasan, M.; Rashid, K.; Siddique, A.H. Blockchain Application in Healthcare Systems: A Review. *Systems* **2023**, *11*, doi:10.3390/systems11010038.
25. Fang, G.; Sun, Y.; Almutiq, M.; Zhou, W.; Zhao, Y.; Ren, Y. Distributed Medical Data Storage Mechanism Based on Proof of Retrievability and Vector Commitment for Metaverse Services. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 6298–6307, doi:10.1109/JBHI.2023.3272021.
26. Qureshi, S.S.; He, J.; Zhu, N.; Nazir, A.; Fang, J.; Ma, X.; Wajahat, A.; Ullah, F.; Qureshi, S.; Dhelim, S.; et al. Enhancing IoT security and healthcare data protection in the metaverse: A Dynamic Adaptive Security Mechanism. *Egyptian Informatics Journal* **2025**, *30*, doi:10.1016/j.eij.2025.100670.
27. Hornik, J.; Ofir, C.; Rachamim, M.; Graguer, S. Fog Computing-Based Smart Consumer Recommender Systems. *JTAER* **2024**, *19*, 597–614, doi:10.3390/jtaer19010032.
28. Marabelli, M. *AI, Ethics, and Discrimination in Business;* Springer International Publishing: Cham, **2024**, ISBN 978-3-031-53918-3.

29. Kuzdraliński, A.; Miśkiewicz, M.; Szczerba, H.; Mazurczyk, W.; Nivala, J.; Księżopolski, B. Unlocking the potential of DNA-based tagging: current market solutions and expanding horizons. *Nat. Commun.* **2023**, *14*, doi:10.1038/s41467-023-41728-2.
30. Jo, S.; Shin, H.; Joe, S.-Y.; Baek, D.; Park, C.; Chun, H. Recent progress in DNA data storage based on high-throughput DNA synthesis. *Biomed. Eng. Lett.* **2024**, *14*, 993–1009, doi:10.1007/s13534-024-00386-z.
31. Yu, M.; Tang, X.; Li, Z.; Wang, W.; Wang, S.; Li, M.; Yu, Q.; Xie, S.; Zuo, X.; Chen, C. High-throughput DNA synthesis for data storage. *Chem. Soc. Rev.* **2024**, *53*, 4463–4489, doi:10.1039/d3cs00469d.
32. Wang, J.; Chen, R.; Lv, Z. DNA Computing-Based Multi-Source Data Storage Model in Digital Twins. *ACM Trans. Multimedia Comput. Commun. Appl.* **2023**, *19*, 1–16, doi:10.1145/3561823.
33. Doricchi, A.; Platnich, C.M.; Gimpel, A.; Horn, F.; Earle, M.; Lanzavecchia, G.; Cortajarena, A.L.; Liz-Marzán, L.M.; Liu, N.; Heckel, R.; et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* **2022**, *16*, 17552–17571, doi:10.1021/acsnano.2c06748.
34. Li, K.; Chen, H.; Li, D.; Yang, C.; Zhang, H.; Zhu, Z. Empowering DNA-Based Information Processing: Computation and Data Storage. *ACS Appl. Mater. Interfaces* **2024**, *16*, 68749–68771, doi:10.1021/acsami.4c13948.
35. Milenkovic, O.; Pan, C. DNA-Based Data Storage Systems: A Review of Implementations and Code Constructions. *IEEE Trans. Commun.* **2024**, *72*, 3803–3828, doi:10.1109/TCOMM.2024.3367748.
36. Ceze, L.; Nivala, J.; Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456–466, doi:10.1038/s41576-019-0125-3.
37. Blawat, M.; Gaedke, K.; Hütter, I.; Chen, X.-M.; Turczyk, B.; Inverso, S.; Pruitt, B.W.; Church, G.M. Forward Error Correction for DNA Data Storage. *Procedia Computer Science* **2016**, *80*, 1011–1022, doi:10.1016/j.procs.2016.05.398.
38. Church, G.M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628, doi:10.1126/science.1226355.
39. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80, doi:10.1038/nature11875.
40. Zhang, Y.; Ren, Y.; Liu, Y.; Wang, F.; Zhang, H.; Liu, K. Preservation and Encryption in DNA Digital Data Storage. *Chempluschem* **2022**, *87*, doi:10.1002/cplu.202200183.
41. Yan, Z.; Zhang, H.; Lu, B.; Han, T.; Tong, X.; Yuan, Y. DNA palette code for time-series archival data storage. *Natl. Sci. Rev.* **2025**, *12*, doi:10.1093/nsr/nwae321.
42. Hilbert, M.; López, P. The world’s technological capacity to store, communicate, and compute information. *Science* **2011**, *332*, 60–65, doi:10.1126/science.1200970.
43. Pan, C.; Tabatabaei, S.K.; Tabatabaei Yazdi, S.M.H.; Hernandez, A.G.; Schroeder, C.M.; Milenkovic, O. Rewritable two-dimensional DNA-based data storage with machine learning reconstruction. *Nat. Commun.* **2022**, *13*, doi:10.1038/s41467-022-30140-x.
44. Hwu, E.E.-T.; Boisen, A. Hacking CD/DVD/Blu-ray for Biosensing. *ACS Sens.* **2018**, *3*, 1222–1232, doi:10.1021/acssensors.8b00340.
45. Kim, S.J.; Jung, W.-B.; Jung, H.S.; Lee, M.-H.; Heo, J.; Horgan, A.; Godron, X.; Ham, D. The bottom of the memory hierarchy: Semiconductor and DNA data storage. *MRS Bulletin* **2023**, *48*, 547–559, doi:10.1557/s43577-023-00510-x.
46. Akash, A.; Bencurova, E.; Dandekar, T. How to make DNA data storage more applicable. *Trends Biotechnol.* **2024**, *42*, 17–30, doi:10.1016/j.tibtech.2023.07.006.
47. Xu, X.; Wang, W.; Ping, Z. Biotechnological tools boost the functional diversity of DNA-based data storage systems. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 624–630, doi:10.1016/j.csbj.2025.02.002.
48. Baek, D.; Joe, S.-Y.; Shin, H.; Park, C.; Jo, S.; Chun, H. Recent Progress in High-Throughput Enzymatic DNA Synthesis for Data Storage. *BioChip J* **2024**, *18*, 357–372, doi:10.1007/s13206-024-00146-2.