



journal homepage: www.elsevier.com/locate/csbj



Review

The visual story of data storage: From storage properties to user interfaces



Aleksandar Anžel *, Dominik Heider, Georges Hattab

University of Marburg, Department of Mathematics and Computer Science, Marburg 35043, Germany

ARTICLE INFO

Article history:

Received 29 June 2021
 Received in revised form 19 August 2021
 Accepted 19 August 2021
 Available online 24 August 2021

Keywords:

Storage
 Device
 Medium
 Usage
 Capacity
 Lifespan

ABSTRACT

About fifty times more data has been created than there are stars in the observable universe. Current trends in data creation and consumption mean that the devices and storage media we use will require more physical space. Novel data storage media such as DNA are considered a viable alternative. Yet, the introduction of new storage technologies should be accompanied by an evaluation of user requirements. To assess such needs, we designed and conducted a survey to rank different storage properties adapted for visualization. That is, accessibility, capacity, usage, mutability, lifespan, addressability, and typology. Withal, we reported different storage devices over time while ranking them by their properties. Our results indicated a timeline of three distinct periods: magnetic, optical and electronic, and alternative media. Moreover, by investigating user interfaces across different operating systems, we observed a predominant presence of bar charts and tree maps for the usage of a medium and its file directory hierarchy, respectively. Taken together with the results of our survey, this allowed us to create a customized user interface that includes data visualizations that can be toggled for both user groups: Experts and Public.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

| | |
|--|------|
| 1. Introduction | 4905 |
| 2. Results | 4905 |
| 2.1. Timeline by typology, accessibility, and mutability | 4905 |
| 2.2. Timeline by typology, capacity, and lifespan | 4908 |
| 2.3. Novel and alternative media | 4909 |
| 2.4. Industry-based UI and visualizations | 4909 |
| 2.5. Survey results | 4912 |
| 3. Methods | 4912 |
| 3.1. Literature search | 4912 |
| 3.2. Ranking survey | 4912 |
| 3.3. User Interface | 4912 |
| 3.4. Property-based visualizations | 4914 |
| 4. Conclusion | 4915 |
| 5. Discussion | 4916 |
| Author contributions statement | 4916 |
| Declaration of Competing Interest | 4916 |
| Acknowledgements | 4916 |
| References | 4916 |

* Corresponding author.
 E-mail address: aleksandar.anzel@uni-marburg.de (A. Anžel).

1. Introduction

At our current rate, 2.5 quintillion bytes of human and machine-generated data are created every day, and the pace is only accelerating [1]. The amount of created data in 2020 was predicted to reach 35 zettabytes or ZB (i.e., 35 trillion gigabytes or GB). 33 ZB were already reached back in 2018. This led the International Data Corporation (IDC) to make a new estimation for 2025: 180 ZB of new data will be created worldwide [2]. Although such numbers are not astronomically large, they remain too large to fathom. For comparison, the radius of the observable universe is roughly 43×10^{21} kilometers; which is two orders of magnitude smaller than the aforementioned estimate of 180 ZB or 1.8×10^{23} bytes [3]. With the advent of such a historical era, better storage devices and long-term storage solutions will be required. This especially holds true with a projection putting the world population around 10.88 billion in 2100 [4].

A storage device can contain information, process such information, or do both [5]. When the device contains only information, it is called a recording medium. Recording can be done using almost any form of energy, acoustic vibrations (phonographic recording) to electromagnetic energy (magnetic tape and optical discs) [6–8]. Today's storage devices contain different types of magnetized media that are usually ferromagnetic materials (e.g., iron or chromium oxides) [9,10]. Ferromagnetic materials have structurally unpaired spins that are organized into magnetic domains. Many bytes, as the units of information, can be recorded on magnetized media as "ascending" or "descending" spin domains. These correspond to ones and zeros in the binary system. Thanks to the unique property of ferromagnetic materials, these magnetized media retain the data, and they can then be read in the same way. To provide a descriptive overview of storage devices and media, we follow their basic properties: (1) Accessibility, defines how data is organized on a device and how it can be accessed: serial or random [11]; (2) Capacity, defines how much capacity a device has (in bytes) [12], (3) Lifespan, defines how long data can be stored in certain conditions (in years) [13,14], (4) Mutability, defines the functions of a device: write, read, or both; and (5) Typology, defines the categories of storage devices: optical, magnetic, semiconductor or electronic, molecular, etc. Combined with other properties, such as energy use or data density, they affect the adoption and usage costs of a specific device or medium. For example, data density is a measure of the quantity of information bits that can be stored on a given length of track, area of surface, or in a given volume of a storage medium. A higher density is preferred to optimize the given length, surface, or volume of said medium. Altogether, such properties steer the adoption of storage devices and media not only by commercial companies but also by the public. In the context of long-term data storage, devices with volatile memory are faster than non-volatile memory. However, in the event of a power outage, volatile memory is not retained rendering it unsuitable. In the looming possibility of a digital dark age, this rationale extends to digital libraries and renders them obsolete, e.g., tapes or network storage systems (clouds). This, in turn, led to the development of alternative and novel media.

Striking examples of novel storage media are Ribonucleic acid (RNA) and Deoxyribonucleic acid (DNA) molecules. Because of its greater stability, the DNA molecule presents a better storage medium than its counterpart RNA. It carries the biological information necessary for the proper functioning of cells [15–17]. It is not by any means new, as it has been used for billions of years as a carrier for genetic information by living organisms [18]. Compared to other – organic and inorganic – molecules, its capacity puts it in a league of its own [19,20]. DNA has a great potential for information storage with a capacity outperforming existing technologies.

For instance, traditional storage devices such as magnetic hard drives and flash drives have a data density of 10^{13} and 10^{16} bits per cm^3 , respectively [21]. In comparison, DNA reaches a data density up to 10^{19} bits per cm^3 . In other words, three orders of magnitude higher and specifically of 1 billion terabyte per gram. Moreover, DNA has a very good molecular stability, which has been shown in sequencing studies of extinct species, referred to as ancient DNA [22,23].

Aside from their novelty, a yet-to-be-considered aspect concerns the User Interface (UI), which of the aforementioned properties are relevant, and how they are visualized. For instance, knowing if a Hard Disk Drive or HDD device is at 20% or 50% of its usage is often visualized using a horizontal stacked bar chart. However, with the advent of new storage media, such as molecular media, there exist no known standards. That is to say, there is no agreement on which properties are more important or relevant for a potential user. Indeed, user preferences impact which information is judged as relevant, then subsequently visualized. In this work, we argue that certain standards should be considered to prepare for the foreseeable future where such media may be available for the task of long-term archiving or even for day-to-day use. For this purpose, we investigated the industry-wide approaches that implement UIs and visualizations to display some of these storage properties. We reported a historical account of their evolution across operating systems and platforms. We also divided the timeline into three distinct periods (magnetic, optical and electronic, alternative) and reported the TOP 3 media and devices for each time period. To accommodate for the shift of storage types into molecular media, we created a survey to rank the storage properties by importance depending on whether the user is a member of the general public or the research community (i.e., domain expert). The expert pool consisted of members of the largest known research consortium on MOlecular Storage for Long-term Archiving (MOSLA). We summarize our findings below:

- (a) We conducted a successful literature search of historical, currently in use, novel and experimental data storage media and devices,
- (b) we created a survey to determine the most important storage media properties depending on a target group, and
- (c) by relying on our findings, we proposed a user-settable UI to select and display the relevant properties for the right audience.

2. Results

We report the results in five subsections. First, we focus on presenting specific storage media and devices that relate to the typology, accessibility, and mutability properties. We rely on the typology property to delineate periods of time for making the timeline more tractable. Second, we present storage media and devices by capacity and lifespan. In each subsection, only the TOP 3 is reported for the sake of brevity. Third, we briefly describe novel and alternative media as they bring in their own right a new set of challenges. Fourth, we report the results of our survey on user preferences to visualize storage properties. Fifth, we detail our proposed UI that can be toggled depending on the target audience with historically adapted visualizations and user-settable parameters.

2.1. Timeline by typology, accessibility, and mutability

First, we report the timeline of different storage media by their type. Second, and for brevity, we consider the logical condition of

reporting the TOP 3 storage devices and media that are only and only if both randomly accessible and mutable (read and write).

In the beginning, data storage heavily relied on magnetic media. As seen in Table 1 and Figs. 1 and 2, the magnetic time period is observed in blue. This magnetic era slowly shifted to optical and electronic media (in orange and gray). In addition, experimental storage technologies with hybrid typologies saw the light. For example, electro-mechanical and magneto-optical typologies, DAT (in red) and Mini-Disc (in gray and blue), respectively. This optical and electronic era continues to supply remarkable data storage and media. However, in most recent years, we have seen the rise of the novel and alternative media that effectively rely on atoms and molecules, while continuing to make extensive use of previous media types. Hence, the timeline of data storage devices and media can be divided into three time periods or eras.

The first is the magnetic era, where the arrival of magnetic core memory and Hard Disk Drives revolutionized data storage as we

know it. This era spanned ~40 years of almost exclusively magnetic devices. From the introduction of drum memory in 1932 to the bubble memory in 1970, magnetic devices starring this era were the Magnetic-core memory, the Hard Disk Drive (HDD), and the Floppy Disk. **Magnetic-core memory**, or ferrite-core memory, is an early form of computer memory. It uses small magnetic ceramic rings, namely the cores, to store information via the polarity of the magnetic field that they contain. In 1949, the earliest work on core memory was done by Shanghai-born American physicist An Wang, who created the Pulse Transfer Controlling Device. The name referred to the way the magnetic field of the cores could be used to control the switching of the current [77].

A Hard Disk Drive (HDD) stores and retrieves digital data from a planar magnetic surface and relies on rigid rotating platters. The information is written to the disk by transmitting an electromagnetic flux through an antenna or write head that is very close to a magnetic material, which in turn changes its polarization due

Table 1

Timeline of data storage media. In the beginning, data storage heavily relied on magnetic media. This slowly shifted to optical and electronic media. In recent years, we have seen the rise of novel media that rely on atomic properties, biological molecules, and organisms. The magnetic Band (Tape) storage medium is reported in 1952, although BASF introduced the first magnetic tape for audio in 1934. Although BASF supplied the first 50,000 meters of magnetic audiotape in 1932, the magnetic tape is reported in 1952 as it corresponds to the first usage of a tape as a data carrier using the IBM 7 track. Mutability refers to read, write, or both. Note that there exists no device with a write-only mutability. We refer to the binary state of read and write (using the letter x) versus the read-only state [18,24–76].

| Year | Data Storage Name | Type | Accessible | Mutable |
|------|---|------|------------|---------|
| 1932 | Drum memory ^{24–26} | ● | | X |
| 1946 | Williams-Kilburn Tube ^{27,28} | ● | | X |
| 1949 | Magnetic-core memory ^{29,30} | ● | X | X |
| 1952 | Magnetic Band (Tape) ^{26,31,32} | ● | | X |
| 1956 | Hard Disk Drive (HDD) ^{33,34} | ● | X | X |
| 1960 | Magnetic stripe card ^{35,36} | ● | | X |
| 1969 | Floppy Disk ^{37,38} | ● | X | X |
| 1970 | Bubble Memory ^{39–41} | ● | | X |
| 1978 | LaserDisc (LD)/Discovision (DV) ⁴² | ● | X | X |
| 1978 | Solid-State-Drive (SSD) ^{43,44} | ● | X | X |
| 1981 | Compact Disc (CD) ⁴⁵ | ● | X | X |
| 1987 | Digital Audio Tape (DAT) ^{46,47} | ● | | X |
| 1991 | Mini-Disc ^{46,48} | ● | X | X |
| 1992 | Digital Compact Cassette (DCC) ^{49,50} | ● | | X |
| 1994 | Zip Drive ⁵¹ | ● | X | X |
| 1994 | CompactFlash (CF) ⁵² | ● | X | X |
| 1995 | Digital Versatile Disc (DVD) ^{53,54} | ● | X | X |
| 2000 | USB Flash Disk ^{55,56} | ● | X | X |
| 2001 | Secure Digital (SD) card ⁵⁷ | ● | X | X |
| 2002 | Blu-ray Disc (BD) ^{58,59} | ● | X | X |
| 2004 | Holographic Versatile Disc (HVD) ^{60–62} | ● | X | X |
| 2009 | Millennial Disc (M-DISC) ⁶³ | ● | X | |
| 2010 | Synthetic Deoxyribonucleic acid (DNA) ^{18,64–70} | ○ | | |
| 2014 | Cell cultures ^{71–73} | ○ | | |
| 2014 | Colloidal particle clusters ⁷⁴ | ○ | | X |
| 2016 | Chlorine Atomic memory ⁷⁵ | ○ | X | X |
| 2019 | Synthetic Metabolomes ⁷⁶ | ○ | | |

- Magnetic
- Electronic
- Optical
- Electro-mechanical
- Magneto-Optical
- Cathode-ray tube
- Atomic
- Molecular

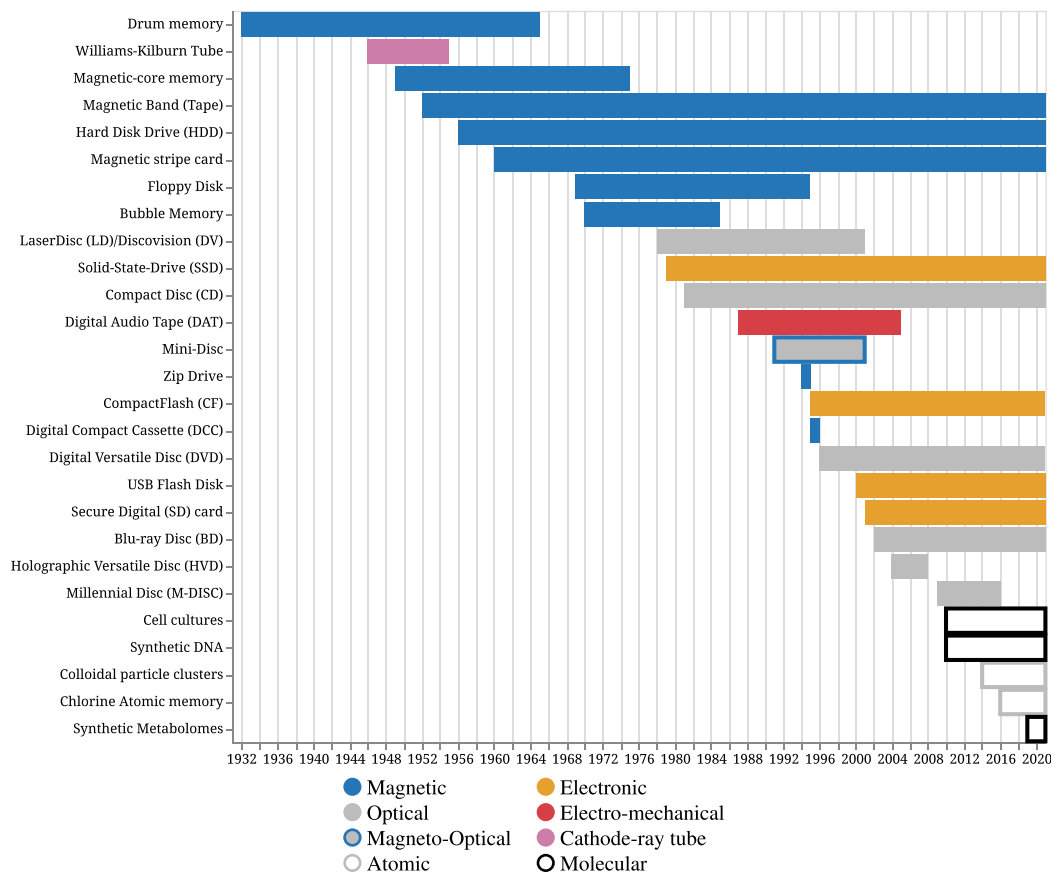


Fig. 1. Timeline of Storage Media and their Usage. The transition to new storage technologies is observable at multiple time points. Today, various storage media from different eras are still in use.

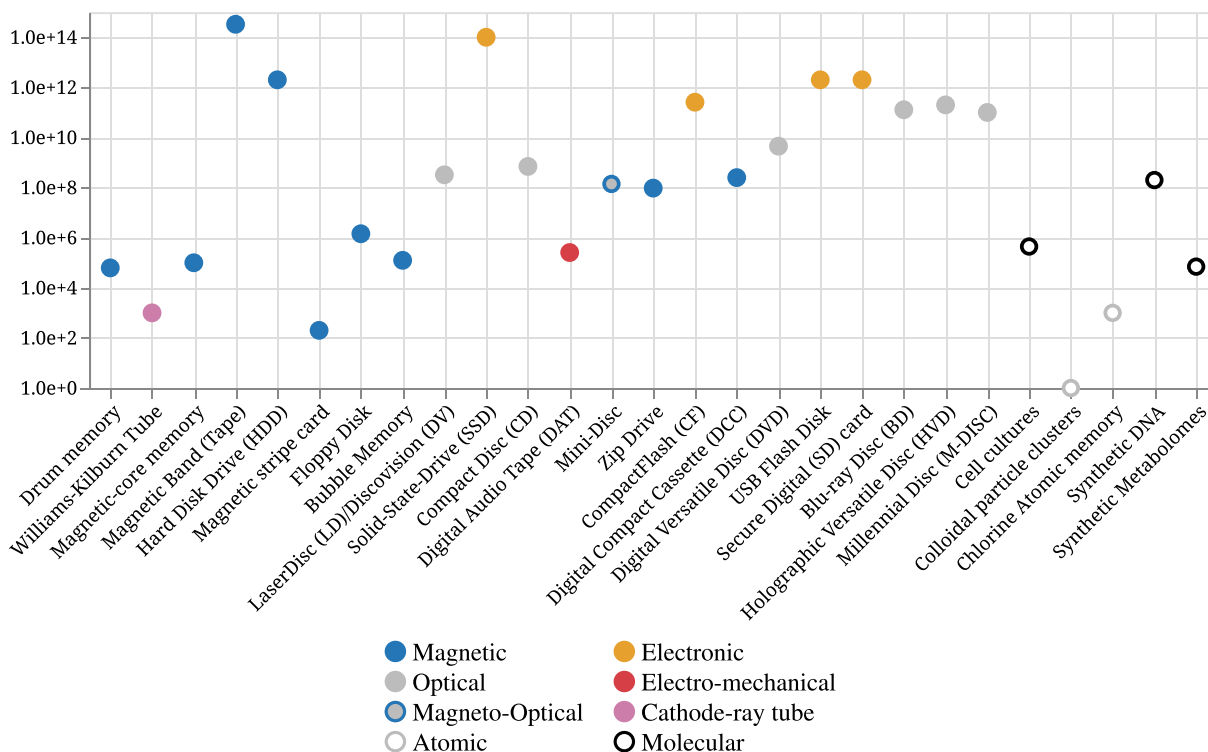


Fig. 2. The Capacity of Storage Media over Time. Values in bytes are reported on a Log scale with a base 10. It is possible to observe an increase over time as novel media improve in capacity. However, new experimental devices or media are often expensive which limits the storage capacity.

to the flux. The first computer with an HDD as standard was the IBM 350 Disk File, introduced in 1956 with the IBM 305 computer. **The Floppy Disk** is a data storage device that is composed of a circular piece of thin flexible magnetic medium encased in a square or rectangular plastic casing. A Floppy Disk is read and written using a floppy disk drive. In 1967, IBM started developing a practical and inexpensive device for easy loading of microcode into their 370 mainframes and for sending out customer updates. The result of this work was a read-only, 8-inch (20 cm) floppy. Initial floppy disks were designed to hold 80 KB and load microcodes into IBM 3330, making them an intermediate device to fill another storage device, i.e., a disk pack file with a 100 MB capacity. Next disks were 5.25 inches, then dimensions changed to 3.5 inches and with added protection thanks to a sliding metal cover to protect the disk medium from direct physical contact. With new floppy sizes and competitive prices, newer and smaller floppies replaced their predecessors very quickly.

The second is the optical and electronic era. Although other hybrid device and media types were introduced in this era, the optical and electronic types were the most prominent. It spanned 24 years, from 1978 to 2002, with many media and devices that were both accessible and mutable. It all started in 1978 with the LaserDisc. **The LaserDisc** (LD) is a home video format and the first commercial optical disc storage medium. In the same year, in 1978, StorageTek® launched the STC 4305, which was the first semiconductor storage device compatible with a hard drive interface, or **the Solid-State-Drive** (SSD), aimed at the IBM mainframe plug-compatible market. It entered the market as a serious competitor to the IBM 2305 HDD system with seven times the speed and half its price. The LD was the predecessor of the Compact Disc (CD), the Digital Versatile Disc (DVD), and the Blu-ray Disc (BD) which were launched in 1981, 1995, and 2002, respectively. While the SSD was the predecessor of flash memory-based media, namely the Compact Flash (CF), the USB Flash Disk, and the Secure Digital (SD) card, which were launched in 1994, 2000, and 2001, respectively. Introduced in 1994, **the Zip Drive** failed to replace the widely adopted 3.5 inches floppy disks although it was the super-floppy of the era with a much greater capacity and performance.

The third is the molecular and atomic storage era. It spans from 2004 to today. Although this era features novel storage media, only Chlorine Atomic memory and Holographic Versatile Disk (HVD) are considered both accessible and mutable. From 2004 to 2008, **the Holographic Versatile Disc** (HVD) was researched and its development was halted. Compared to DVD technology, holographic memory could have stored information at higher density inside crystals or photopolymers [61]. In DVDs, the upper limit of the data density was reached due to the diffraction limit on the writing beams. Although promising, the HVD was never released. In 2006, **Chlorine Atomic memory** consists of arranging functional atoms into extended and scalable atomic circuits. The idea is to create an atomic-scale memory that can be read and rewritten automatically by means of atomic-scale markers using chlorine vacancies on a copper sheet (Cu) [75]. Such vacancies are found to be stable at temperatures up to 77 K (−196,15 °C) and would outperform state-of-the-art HDDs by three orders of magnitude.

Over time, storage devices improved in terms of accessibility. However, preferences for novel upcoming devices and media shifted their adoption and usage. As seen in Fig. 1, if the timeline is based solely on the use of storage devices, we observe a clear separation between storage devices that are still in use and those that are obsolete. For example, Floppy Disks became obsolete in the mid-1990s. From the magnetic era, accessible and mutable devices that survived only include HDDs. From the optical and electronic era, all aforementioned optical and electronic media are still being used today with the exception of the Blu-ray Drive

which was discontinued in 2019, and later on, replaced with UHD Blu-ray which is now predominantly used. From the molecular and atomic storage era, Chlorine Atomic memory is the only accessible and mutable medium. Although this is the case, many other media have been developed with long-term archiving in mind. This led to a polarization of inaccessible and immutable media (e.g., synthetic DNA, synthetic metabolomes).

2.2. Timeline by typology, capacity, and lifespan

In this second subsection, we borrow the aforementioned division of the timeline and report the TOP 3 storage devices in terms of capacity and lifespan properties. Fig. 2 introduces the capacity of storage media over time on a Log scale.

First, magnetic storage devices proved to be resilient to the ever-growing need for more storage space. However, as electronic technology progressed, the capacity of magnetic-based storage devices became greater and the actual magnetic memory became cheaper. Because of their resilience and price, magnetic devices are widely adopted for archival use. **Hard Disk Drives** (HDDs) are the most developed magnetic-based storage medium. Their capacity increased substantially, from 3,750,000 bytes of IBM 350 disk storage unit to around 2×10^{12} bytes of modern HDDs [78]. The lifespan of any storage medium is directly influenced by environmental factors in which that medium resides, as well as the frequency in which it is used. It was long thought that higher temperatures might increase the chance of HDD failure, however, recent studies have found that there was no correlation between physical drive temperature and drive failures [79]. Since HDDs rely on mechanical parts for read and write operations, their lifespan is constrained to the quality and durability of those parts. New technologies of disk coating are being introduced to improve the current durability and data density of HDDs. Carbon-based overcoats are replaced with graphene-based ones to achieve a better reduction in friction and provide superior corrosion and wear resistance. This, in turn, enables the potential of increasing data density up to 4 to 10 times [34]. **The Zip drive** was the least long in use compared to all previously mentioned media and devices. Therefore, not much was done to further develop its capacity and lifespan. At the peak of their development, the maximum capacity of the Zip Drive was around 10^8 bytes. The internal structure of the Zip Drive is almost identical to that of the HDD, consisting of read/write heads hovering over a rapidly spinning floppy disk mounted in a sturdy cartridge. That makes the Zip Drive much like the HDD, prone to failure and with a similar lifespan. **Magnetic-core memory** was introduced in the early days of computer systems and engineering. The properties of this storage device, such as non-volatility and random-accessibility, made it perfect for use as the primary memory of those early systems. The maximum capacity of the Magnetic-core memory was around 10^6 but was not further improved, as the storage medium was replaced with a more technologically advanced static random-access memory.

Second, the optical and electronic era. With the introduction of the first optical storage medium in 1978, **the LaserDisc** (LD) provided a serious alternative for long-term data storage. Its introduction occurred roughly at the same time as the first electronic storage medium, the SSD. As previously mentioned, the LD preceded many optical media. **The Blu-ray disc**, and its modern successor the Ultra HD (UHD) Blu-ray, currently holds the record for highest storage capacity for optical media of this era. When first introduced, the capacity of this storage medium was around 25×10^9 , while it now peaks at around 10^{11} bytes. Unlike the organic dye used in optical disc media, found in DVDs and CDs, a different approach is used to encode data on a Blu-ray disk. A com-

bination of silicon and copper is used to create a layer on which the data is engraved, making Blu-ray disks significantly more durable and resilient than DVDs and CDs, with a lifespan of around 150 years. However, this warrants further research since these media did not exist long enough to either confirm or deny such claims. Alternatively, the current record for electronic media or storage devices of this era is currently held by **the SSD** with a maximum capacity of 10^{14} bytes. In second and third place, **the USB flash disk** and **the SD card** share the record with current maximum capacities of around 2×10^{12} bytes. Indeed, electronic-based storage devices do not contain any moving parts and therefore use entirely different methods to read, write, and store data. Research shows that the life span of devices based on flash technology is greatly determined by the usage in terms of data written on those storage media. Furthermore, it has been shown that SSDs were replaced 25% less often than HDDs [80].

The third era is represented by the introduction of many novel media such as cell cultures, synthetic DNA, and synthetic metabolomes. In addition, the Holographic Versatile Disc (HVD) optical medium was meant to compete with Blu-ray disk on capacity and reliability. However, HVD was costly and incompatible with existing or new storage standards, making its adoption problematic. The current storage standard specifies 2×10^{11} bytes of capacity. Yet with no clear demand, such disks were never put in mass production and the HVD remained in the research phase [81]. This third era is further detailed in the next section.

Over time and across media types, the changes in capacity reached in multiple instances local maximums. For example, optical media have reached a capacity of 10^{11} , as seen in Fig. 2.

2.3. Novel and alternative media

Many novel and alternative media have been proposed in different contexts. They comprise atomic and molecular media. In the case of molecular media, a variety of approaches exist including organic, such as synthetic metabolomes, metallo-organic clusters, synthetic DNA molecules, and inorganic clusters. The most promising molecular medium is **synthetic DNA**, although it requires a variety of steps; e.g., synthesis, encoding, sequencing, error correction, etc. Indeed, some approaches figured out that DNA storage systems are too slow to replace HDDs [64]. Yet more recent breakthroughs have made use of hundreds of thousands of short DNA oligonucleotides for encoding small amounts of data [65]. In that approach, naive encoding without error correction was used, which rendered it unsuitable for long-term archiving due to sequencing errors. To avoid problems in sequencing, a Huffman-encoding with a ternary code was employed by another approach [66]. This specifically helped avoid homo-polymeric DNA sequences which are problematic at the sequencing step. Moreover, they introduced redundancy in the encoding, thus enabling a simple error-correcting procedure. Based on these findings, another approach improved data density in the DNA storage medium by employing a kind of RAID (Redundant Array of Independent Disks) system [68]. In addition, a rudimentary random access procedure based on primer sequences was developed. Although it can be used for direct access of parts of the data, it is not possible to perform semantic searches, for which complex index structures are necessary [82]. For this reason, we did not report synthetic DNA as an accessible medium. Further research into improving DNA storage systems made use of different encodings (e.g., fountain codes [69], forward error correction [83], Reed-Solomon codes [67]). In spite of the fact that they are used for error correction, these encodings only correct or compensate until a certain threshold is reached. When the DNA storage molecule(s) are exposed to different reagents or stimuli, error-correcting codes cannot handle

any potential degradation (e.g., damaged bases, breaks between individual nucleotides, and fractures in the phosphate backbone). The usage of higher codes such as Galois fields codes could help error- and erasure-correcting codes for reliable DNA storage [84,85]. A comprehensive review of the research literature on synthetic DNA storage and its challenges is addressed by Dong et al. [86]. Physical storage of synthetic DNA varies from one laboratory to another. It defines a unit of molecular storage that often relies on sequencing redundancy (i.e., deep sequencing coverage, and having many copies of each sequence) and ranges from, but is not limited to, amber-enclosed spores [18], lyophilized oligonucleotides [66], inorganic silica [67], etc. The capacity value for synthetic DNA is experimentally validated and visually reported in Fig. 2. In this instance, the molecular storage unit is dehydrated synthetic DNA pool.

Significant efforts are being made to use inorganic molecules for data storage. Current efforts are based on the use of four-color printing of clusters to reveal their extreme nonlinear optical properties. These properties should be distinguishable enough for a spectrometer to read them and thus capture the information they contain. The fundamentals of such molecular clusters are described in several works [87–89].

Another aspect of molecular storage is organism-based. This corresponds to maintaining certain **cell cultures**. It effectively relies on integrating short synthetic DNA strands (i.e., oligonucleotides) into a biological organism *in vivo*. In turn, the organisms can store and duplicate the information on an *ad hoc* basis. This special type of medium was researched to insert synthetic DNA fragments encoding data and inserting them into the genome of bacteria, fungi, and plants. This was successfully accomplished in *Escherichia coli*, *Bacillus subtilis*, *Pichia pastoris*, and *Arabidopsis thaliana* [71–73]. Estimates for a single gram of bacteria indicated an advertised storage capacity of more than 900 TB. However, an experimentally validated approach confirms using bacterial cell cultures with 445 KB of digital files in synthetic DNA [73].

Synthetic metabolomes are another promising molecular medium. Metabolomes comprise the complete set of small molecules found in a biological system. Unlike DNA and protein molecules, they are small in mass, abundant, and more structurally and energetically diverse [76]. The synthetic metabolomes consist of a mixture of metabolites (e.g., Galactose, Tryptophan, etc). These are spatially arrayed in thousands of nanoliter volumes on a physical multi-well array or plate. In turn, each resulting volume contains a prescribed mixture from a library of purified metabolites, i.e., a synthetic metabolome. This approach demonstrated the storage of many image data. The largest is a 17,424-bit image requiring approximately 70 KB. This image was written into 1,452 mixtures from a 12-metabolite subset of the library [76].

Another noteworthy addition concerns non-genomic molecular media. They have also been demonstrated yet have not been included in this study due to their experimental nature. One type of such media relies on fluorescent dyes on polymer films or the rotaxane molecular architecture [90,91]. Another type relies on creating nano-structures to obtain an etched crystalline quartz or even a thin diamond layer [92,93].

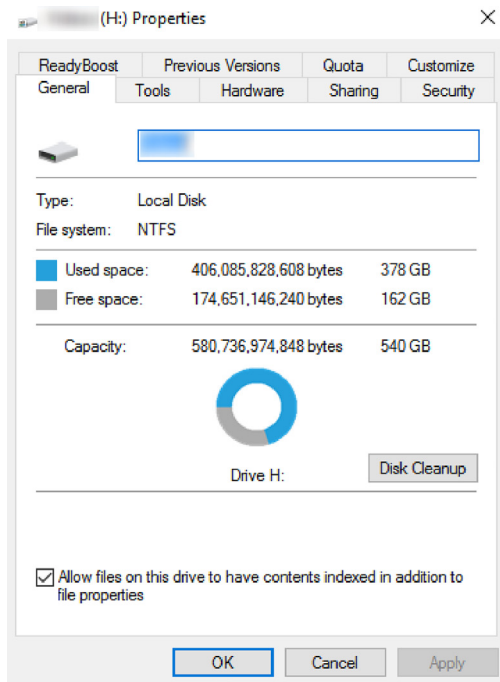
2.4. Industry-based UI and visualizations

From the early days of the computer, the amount of used information in the storage devices was important. Early devices showed the usage of the storage medium in percent or only the occupied space next to the capacity of the medium or both. This information was presented in a textual form and a human-readable format. This refers to encoding the information in ASCII or Unicode text rather than binary data. Many technological breakthroughs, such as nanoscale circuits, enabled more compact and reliable solutions

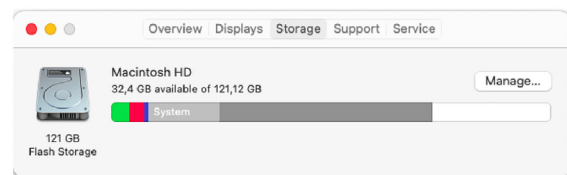
with a higher element density. Various sensors were introduced to different storage types, for example, temperature sensors. This led to the monitoring of meta-information, which in turn allowed the prediction of the device lifespan. In turn, UIs were upgraded with further characteristics such as Temperature or Speed of the storage device. The advancement in graphical processing power also opened up further possibilities. First and foremost, it enabled the visualization of the available information for the capacity property.

Later, it was used to visualize partitioning of the storage medium, file structure, usage, etc. We divide our investigation into early, pre-modern, and modern operating systems.

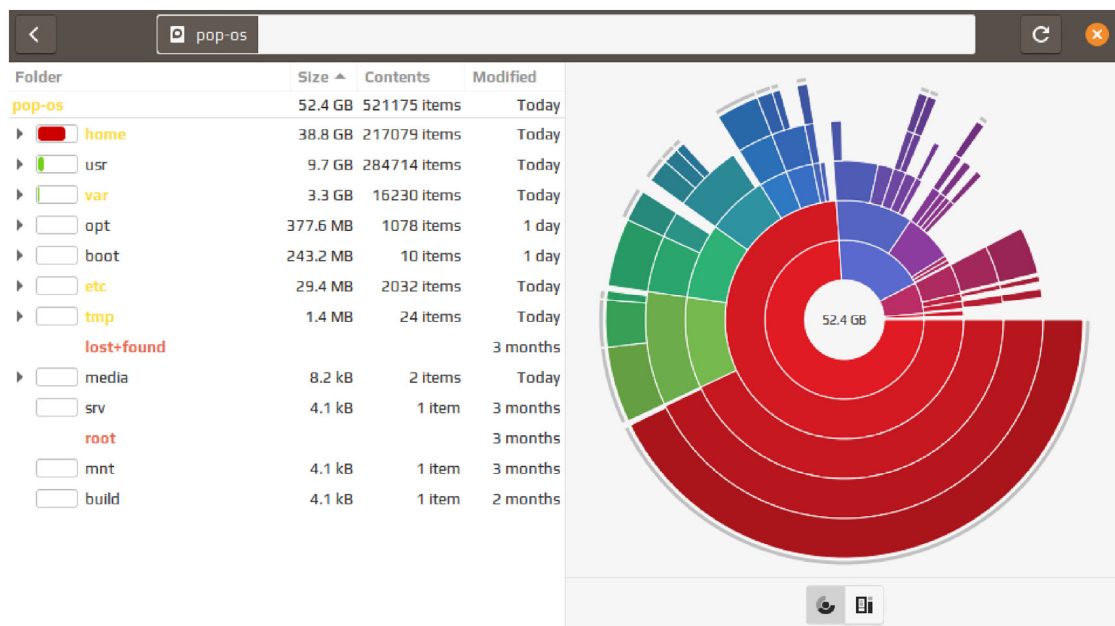
Early systems introduced built-in tools that report detailed information about the user's storage devices. These systems comprised the report of textual information for different properties, yet mainly focused on capacity [94]. Such systems align with the pre-modern arrival of the personal computer with Windows 95®



(a)



(b)



(c)

Fig. 3. Screenshots of data usage visualizations for different Operating Systems. Part-to-a-whole visualization variations are used. (a) Donut chart representing the local disk usage on Windows 10. (b) Stacked bar chart (horizontal) representing the amount of memory used by different file types on Mac OS X Catalina. In the most recent version, Big Sur, only used space is shown in relation to the available space. (c) Sunburst diagram representing the amount of memory used by different directories on Linux (Pop!_OS 20.10).

or System 7[®], and their adoption until Windows XP[®] and Mac OS X[®]. The capacity of a storage device was one of the most important properties and was often visualized as a horizontal stacked bar chart displaying used and free storage space. Modern systems have built-in tools and visualize storage capacity with pie charts and some of its variants, like doughnut charts, or even multi-level pie charts, as seen in Fig. 3.

Along with previously mentioned visualization methods, certain OS offered a built-in representation of used storage space in a tree map chart, as seen in Fig. 4. In most cases, the visualization of usage presents an overview first, details second. Generally, the overview is the stacked horizontal bar chart, while details may be observed in visual and textual form simultaneously. Furthermore, supplementary storage device information, for example,

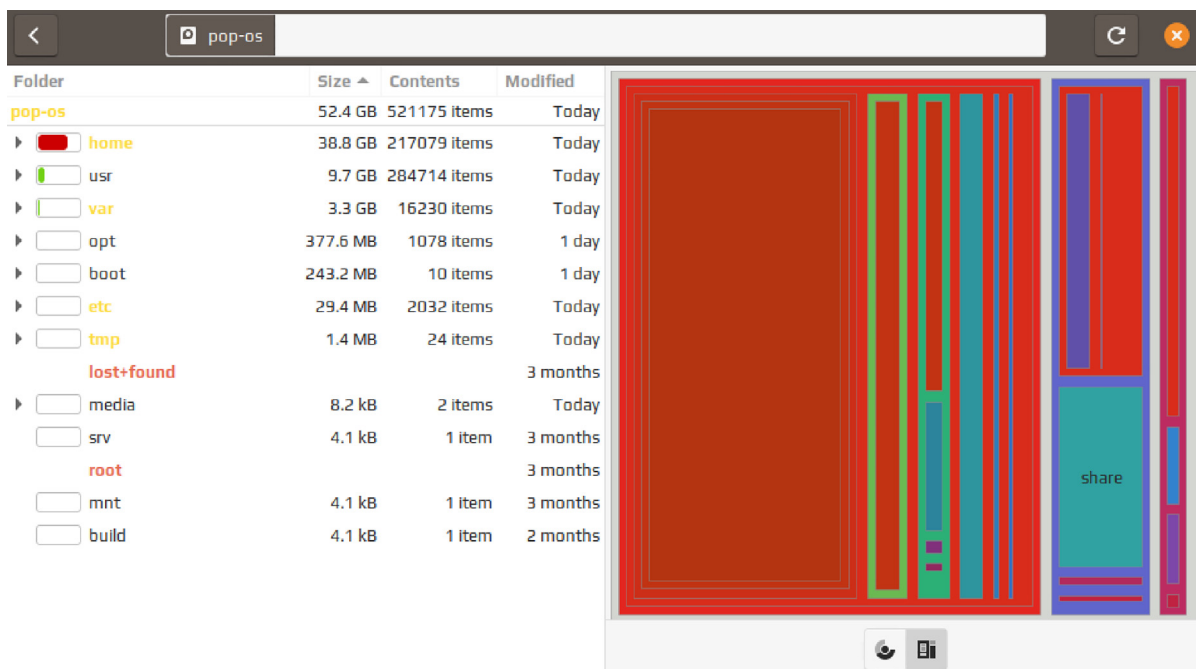


Fig. 4. Alternative visualisation on a Linux distribution. Tree map chart representing the amount of memory used by different directories on Linux (Pop!_OS 20.10).

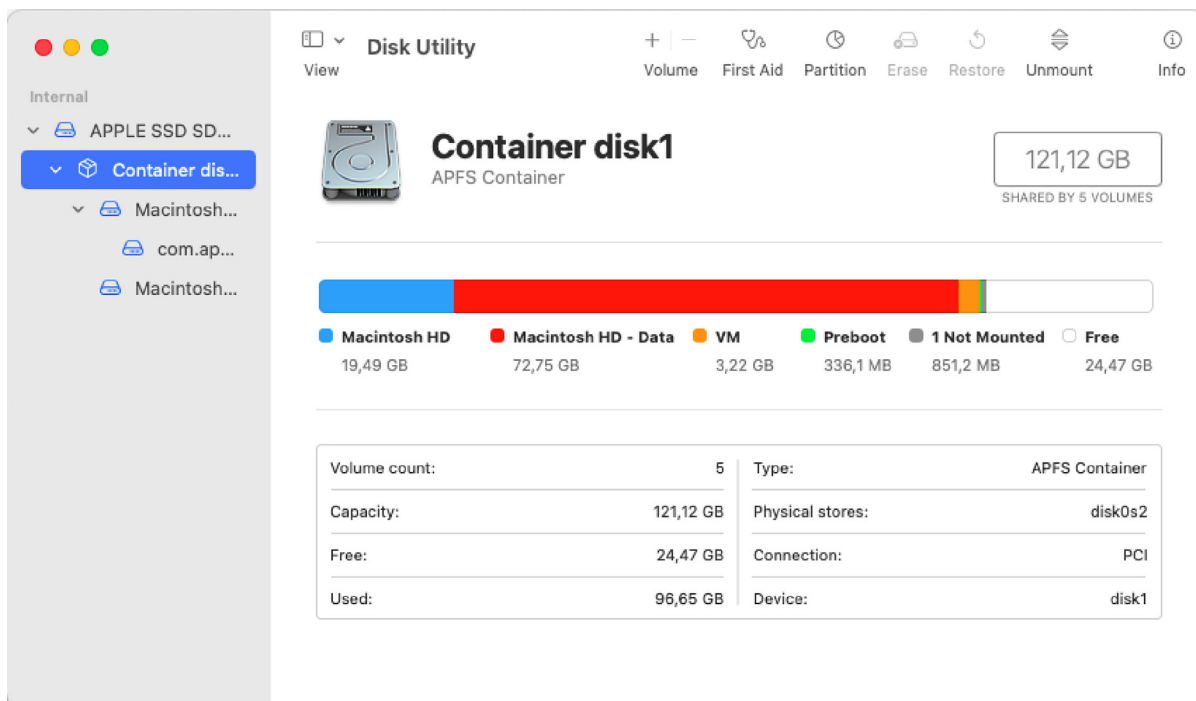


Fig. 5. Additional information on Mac OS X. Supplementary data of how the storage device is partitioned and other important information such as the type of physical connection and the name of the disk are reported in textual form and tabular format on Mac OS X.

Table 2

Survey results. Cells with the highest votes are highlighted. For the expert pool, the TOP 3 properties are lifespan, capacity, and accessibility. When experts are asked about their opinion as members of the public, the order of the TOP 3 changes to: accessibility, capacity, then lifespan. For the general public, the results are the same as when the expert pool stated their opinion as members of the general public.

| Property | The expert pool | | | The experts as public | | | The general public | | |
|---------------|-----------------|----------|----------|-----------------------|----------|----------|--------------------|-----------|----------|
| | 1st | 2nd | 3rd | 1st | 2nd | 3rd | 1st | 2nd | 3rd |
| Accessibility | 5 | 1 | 7 | 8 | 3 | 2 | 17 | 0 | 6 |
| Capacity | 6 | 7 | 4 | 5 | 8 | 2 | 6 | 11 | 6 |
| Lifespan | 8 | 5 | 2 | 2 | 1 | 9 | 5 | 5 | 7 |
| Mutability | 0 | 4 | 3 | 1 | 4 | 4 | 2 | 5 | 6 |
| Usage | 0 | 2 | 3 | 3 | 3 | 2 | 0 | 9 | 5 |

SSD temperature or mount point, is presented in textual form. Fig. 5 depicts this occurrence.

2.5. Survey results

As stated earlier, the survey was conducted on two different groups: (a) domain experts in molecular data storage, and (b) the general public. Nineteen responses were collected from the expert group and thirty from the general public.

In the first group, 68.4% of participants identified themselves as male, and 31.6% as female. Exactly 47.4% put computer science as their field of research, followed by biology at 36.8%, then chemistry at 15.8%, then physics and mathematics at 5.3%. Nearly 60% reported having 10 or more years of experience in their respective field, followed by 31.6% who have 0 to 3 years of experience, and the rest having 3 to 5 years of experience. When ranking the TOP 3 properties, and while considering themselves domain experts, participants chose the following in order of importance: 1. lifespan, 2. capacity, and 3. accessibility. These results changed when considering themselves as members of the general public. The properties of accessibility and lifespan switched places, placing accessibility first and lifespan last. All TOP 3 ranking results are also presented in Table 2.

Exactly 43.3% of the participants belonging to the general public group identified themselves as male, followed by 50% identified as female, while 6.7% preferred not to state their gender. As for the field of study, 60% reported computer science, 16.7% social science, 13.3% biology, 10% medicine, 6.7% mathematics and the rest was split between economics, and business informatics. The general public was not asked the question about their experience. The results gathered from this group, i.e., the general public, are reported in Table 2. In this case, the TOP 3 was: 1. accessibility, 2. capacity, and 3. lifespan.

To support the visualization of the TOP 3 data storage properties, we designed and implemented a User Interface (UI) adapted to both audiences. The methodology for property-based visualizations and the UI is described in the Methods section. The implementation of the UI and the source code used to create all of the figures are uploaded as part of the supplementary material and are also available at the TVSDS GitHub repository: <https://github.com/AAnzel/TVSDS>. The survey and its results are provided with this manuscript as part of the supplementary material.

3. Methods

The literature search was possible thanks to an array of different scholarly literature databases: Google Scholar, Europe PMC, IEEE Xplore. While the experimental part of the paper consisted

of generating a survey, analyzing results, and creating an adequate UI and data visualizations.

3.1. Literature search

Relevant search keywords were used for different parts of this paper. A non-exhaustive list of used keywords is: data storage, novel media, storage device, storage medium, DNA storage, storage visualization, molecular medium, molecular storage. Apart from the literature search, our paper also includes knowledge gathered from several archives, for instance, the IBM Archives, The Internet Archive, and the Museum of Obsolete Media. For certain data storage technologies, like HDDs, HVDs, SSDs, we also incorporated relevant facts stated in their storage standards, that are publicly available online.

3.2. Ranking survey

For the survey, we used Google Forms and presented it as follows. First, a brief explanation of the survey was given followed by a one-sentence description of each data storage property: accessibility, capacity, lifespan, mutability, and usage. Second, participants were asked to rank them in a TOP 3 fashion. The expert group was asked to rank properties as experts in the molecular data storage domain and as members of the general public. Third and last, demographics data was collected. We gathered the results of the survey and proposed a new UI for the property-based visualizations. The survey results laid out in Table 2 are divided into three groups: (a) the domain expert opinion, (b) their view as a public, and (c) the opinion of the general public.

3.3. User Interface

By coupling the results from the survey, the literature search, and the visualization standards to display data storage information, we propose an adapted user interface and state-of-the-art visualizations. The proposed User Interface (UI) consists of two views: the basic view to suit the needs of the general public, and the advanced view for domain experts. The default home view is the basic view. It serves as an overview by presenting the accessibility (sequential versus random), the usage or how much of the media is already in use (in percentage), and the capacity of the media in kilobytes (KB). In addition, a file hierarchy is presented in textual format. The basic view component of the UI can be seen in Fig. 6.

The advanced view can be toggled by a select box button to display further details. It specifically includes additional properties (e.g., lifespan) and further details for the usage and the file hierarchy. This is shown in Fig. 7. The advanced view benefits from

Advanced view



Accessibility: **sequential**

Directory structure:

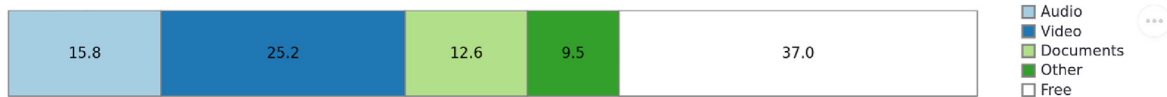
```

/home/user
├── Desktop
├── Documents
├── Downloads
├── Games
├── Music
├── Pictures
└── Videos
    
```

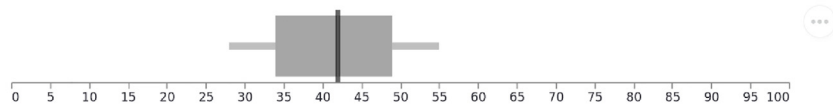
7 directories

Fig. 6. The basic view. It provides an overview of the used versus the free storage space as well as the main properties that are required for the general public.

Advanced view



Lifespan: **42 years**



Accessibility: **sequential**

Temperature: **45°C**

Mutability: **read/write**

Directory structure:

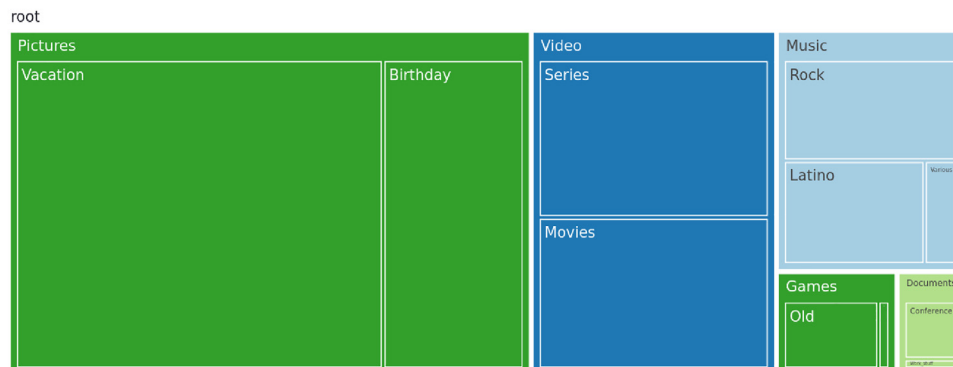


Fig. 7. The advanced view. If the Advanced view checkbox is ticked, the storage medium properties are displayed in order of importance for the expert audience.

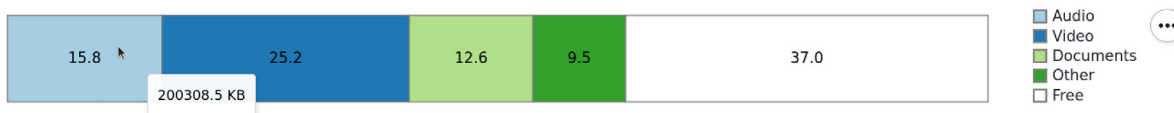


Fig. 8. Capacity/usage tooltip in the advanced view. Supplementary information is shown while mouse-hovering over stacks.

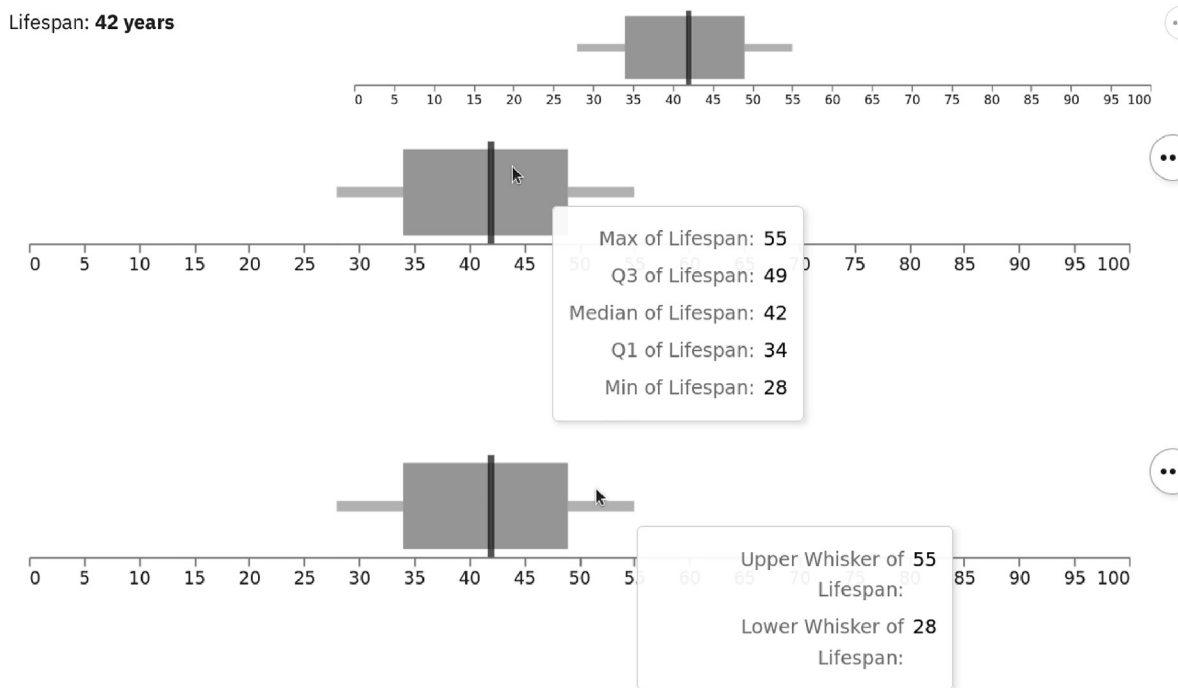


Fig. 9. Lifespan visualization. The lifespan property is presented with textual and visual information. Detailed information of lifespan estimation is presented in an overlay window while hovering over the gray area, as well as over the whiskers.

mouse hover and mouse click events to provide granular details for each visualization. For example, the mouse hover event for the capacity/usage visualization reports the actual size in kilobytes (KB) of a specific file type category (e.g., audio files), as seen in Fig. 8. The mouse hover event for lifespan visualization gives detailed information on lifespan estimation, as shown in Fig. 9.

On the other hand, selecting a directory by a mouse click within the directory hierarchy visualization makes the directory expand to occupy the whole pixel space of the visualization. The chosen directory is now considered a top-level directory, and the visualization shows the hierarchy of its subdirectories, with all previously top-level directories laid on top of the visualization. This is shown in Fig. 10. The UI reports the properties by relying on the TOP 3 ranking results, as seen in Table 2. In light of the survey results, the order of appearance of the visualizations is adjusted to fit the reported importance of the studied properties depending on the audience, and is updated accordingly in its corresponding view (basic vs. advanced). The visualizations present in both UI views dynamically adapt to the current page width, maximizing the ink-to-pixel ratio.

3.4. Property-based visualizations

By considering the state-of-the-art data visualizations for data storage, we propose 3 main visualizations for the ranked properties: usage, capacity, and lifespan. We follow the nested model of visualization to describe the visual encoding [95]. The properties are encoded as follows: capacity as a real number, while usage is

reported as a real number and in percentage (%), lifespan as an integer, and accessibility uses the boolean data type (0 or 1). Textual information is represented as UTF-8 text strings of variable lengths.

First, for both usage and capacity, a horizontal bar chart is employed. Bar charts are very effective at displaying part of a whole, and visually comparing metric values across different subgroups of the data at hand. The basic view shows only two parts or two stacks, free and used space. While the advanced view details which kind of file type categories (i.e., audio, video, documents, other) occupy the storage space, and by how much (in percentage and in KB). Except for the free space stack, each stack is visually encoded using the area and color channels. The area channel represents the used space: each part as a percentage to a whole (i.e., maximum storage capacity in KB). The color channel relies on the categorical or nominal encoding of each file type category. Four file types are considered excluding free space: audio, video, documents, and other are mapped to colorblind-safe categorical colors: #A6CEE3, #1F78B4, #B2DF8A, and #33A02C, respectively. The encoding of the color channel follows the state-of-the-art rules to colorize a data visualization [96]. Moreover, each stack benefits from textual overlays to report the used space in percentage (%). This is presented in Fig. 8.

Second, we used a whisker chart for the lifespan property. It combines a bar and whiskers as well as a vertical line. The latter spans the chart height and encodes the quantitative property (i.e., estimated lifespan) of the medium by relying on the position channel. The estimation of the lifespan is encoded using the area

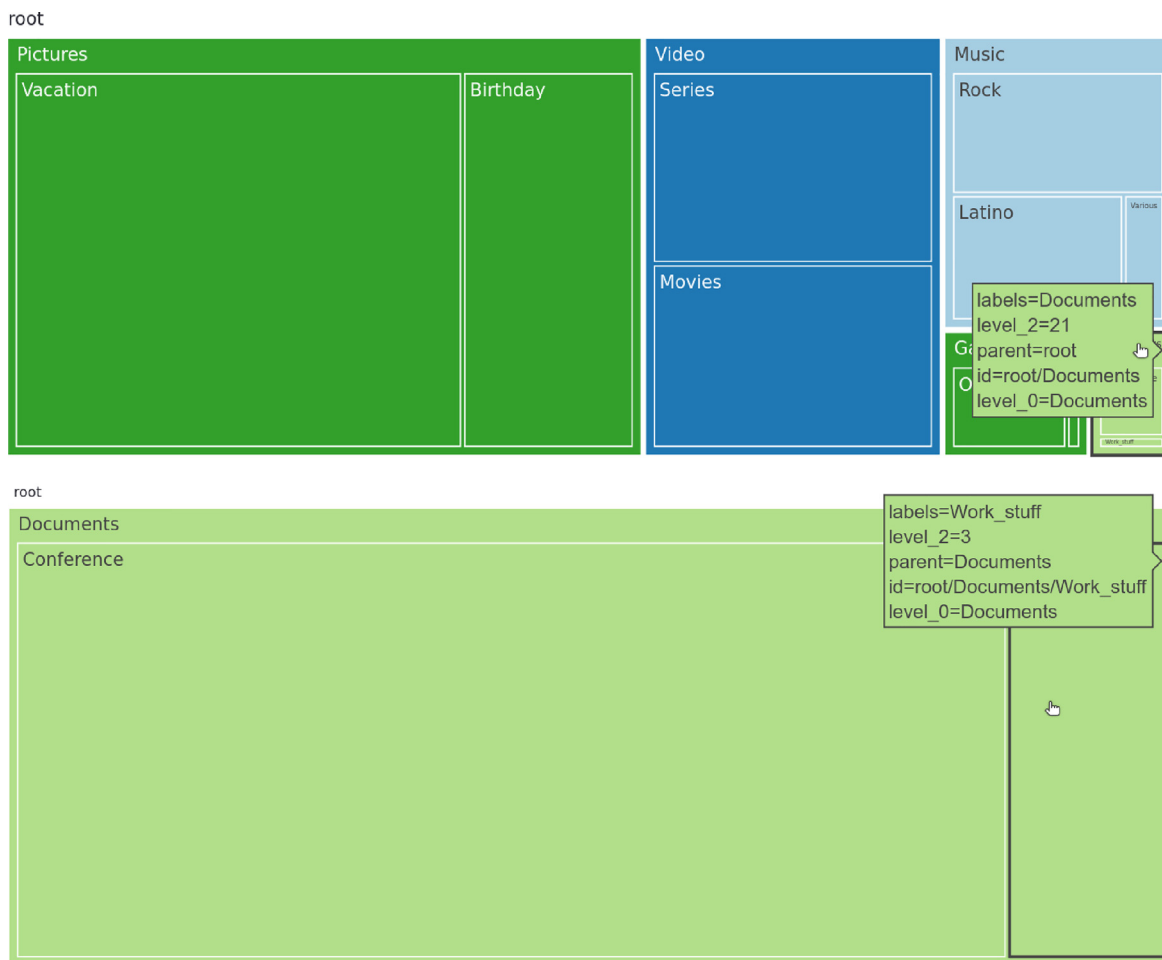


Fig. 10. Directory hierarchy visualized with a tree map chart. Each rectangle is colored according to the file type category present in the corresponding directory. In the image below, the *Documents* folder is now a top-level directory, with *root* as its parent directory. Supplementary information is shown while mouse-hovering over rectangles.

(bar) and the position (line). Since the lifespan of a medium largely depends on various external factors, the whiskers represent the confidence interval of this estimate, while the 95% interval is represented as a gray area. The lifespan property is also reported in a textual form as seen in Fig. 9: the estimated lifespan is 42 years.

Third and last, and as integrated into different operating systems, we implemented a directory hierarchy visualization by using the tree map visualization. The rationale of this visualization method is to maximize the pixel space at our disposal. It is highly efficient since it uses a space-filling technique and relies on creating multiple rectangular areas [97]. Each directory represents a rectangle, where hierarchies between different directories are encoded with containments to create a nested layout. A rectangle is encoded by its area and color channels. The larger the area, the larger the directory. The color channel follows the aforementioned nominal encoding in the usage/capacity chart with 4 file type categories. The name of a directory is encoded in textual form, with each name contained in a rectangle representing that directory, as seen in Fig. 10. Additionally, the accessibility property is not visualized but reported in a textual manner. This can be seen in Fig. 7. The implementation is done using Python 3.9, Altair 4.1, Plotly 4.14.3, and Streamlit 0.82.0 [98–101].

4. Conclusion

Existing storage technologies are insufficient for the long-term storage of the large amounts of data generated in all areas of life.

Novel and alternative media propose to extend the existing storage capacity with molecular storage media and devices. Hence, also reducing the risks for information loss in long-term data storage. Although technological advances are being made for molecular storage media types, there has been a lack of efforts for standardization. This is especially relevant before the adoption of upcoming storage media, such as synthetic DNA. Indeed, standards not only improve ways to standardize information storage for long-term archiving (i.e., lifespan), but also help researchers meet certain criteria when developing novel storage media. It is also important to consider the ways we digest the information that is conveyed by storage media properties. By means of ranking, our survey permitted us to identify relevant properties for domain experts and members of the public. Moreover, thanks to an analysis of the industry-based UI and visualizations, we were able to observe and converge to specific design choices. For the bar chart, the use of the color channel was mapped to file type categories (audio, video, documents, and other), while the area channel depicts the used amount of the available capacity (often in percentage or in GB). On one hand, the horizontal stacked bar chart had been a preponderant choice for the display of the usage of a medium’s capacity. On the other hand, the tree map had also been a common choice to display file directories and file hierarchies. For the tree map chart, the color channel encodes file type categories, and the area channel encodes the amount used by said files. In addition, the containment provided by this space-filling visualization method helped lay out hierarchies among different directories. Owing to these industry-based and widely accepted charts, we developed a user-

settable approach to toggle between an overview and details. This corresponded to the basic vs. advanced view. The latter provided additional data, including further annotations, and unraveled further details about the storage medium.

5. Discussion

First, and thanks to our survey, we were able to separate the different storage properties by relevance. Besides this, we found that the TOP 3 properties converged for the general public and when the expert participants considered themselves as members of the general public.

Second, by having the survey results, we were able to propose a new UI that is generalizable and could also be specifically used for molecular storage media or upcoming novel media. Even though accessibility was chosen as the most important data storage property, it might be reasonable to shift its position to second place. The rationale is that visual information is more relevant and has greater power. From a historical point of view, it is more reasonable to present information rather visually than textually. That is to say, a user should be first confronted with some type of graphical or visual representation. That is why we first present a user with the capacity/usage visualization first, then we present the other properties as ranked by the survey results.

Third, the need for standards extends beyond the currently provided example. Indeed, as seen in the proposed UI and visualization, certain properties have an intrinsic uncertainty. This may either be due to the fact that the property relies on estimation, or that a property is measured. In the general case of the lifespan property, estimates vary to include multiple storage and usage conditions such as temperature, number of read/write, recycling of the media, mechanical movements, etc. We argue that estimates could benefit from standardization so that precise values with certain confidence intervals may be reported. In this regard, the literature lacks evidence-based estimates. In the example case of the lifespan property of synthetic DNA, the theoretical limit is supported by the oldest known preserved DNA in existence, aging approximately one million years [102]. Moreover, the experimentally validated capacity for synthetic DNA seems small in relation to the theoretical limit. This limit is estimated to reach multiple orders of magnitude higher than the presented capacity value. However, the current costs of storing data inside a DNA molecule greatly limits the experimental validation of this maximum. Furthermore, methods and algorithms developed in the field of genomics may benefit current data storage approaches using DNA. This includes, but is not limited to, data compression and indexing [103–105]. In the example case of a synthetic metabolome, the capacity property depends on the number of metabolites present in a metabolome. That is to say, a measurement is made. Such aspects of uncertainty and the way such information is presented for general consumption remain an open question.

Fourth, a noteworthy example of a novel and experimental storage medium is the Data Sticky. Although very promising we excluded it from our findings of the literature search. Data stickies are inspired by sticky notes, such as Post-Its®. A realistic implementation exists, namely Post-Bit, where a small e-paper device stores multimedia contents and allows for paper-like manipulations [106]. They combine the affordability of physical tiny sticky memos, the digital handling, and the display of information using electron ink or e-ink [107]. Once this ink settles into an image, the display reflects light just like ordinary paper; as a non-volatile medium. Recent advancements in nanotechnology proposed a larger storage capacity using graphene paper and e-ink [108]. Estimates put such upgraded data stickies between 4 and 32 GB. There have also been considerable efforts to create nanos-

cale data storage using graphene. A promising strategy addressed high precision writing and drawing on graphene nanosheets by manipulating electrons with a one nanometer-based probe [109].

Fifth and last, even though our UI proposal is applicable to molecular data storage media, more visualization research is warranted for medium-specific properties. Our work herein reported properties that are in the broadest sense shared and relevant. Since molecular data storage media is in active development and research, we can expect some new storage medium-specific properties that could be more important than the properties we reported. That means that UIs and visualizations should evolve and adapt to the new storage media types and all of the important, specific properties those media types may have.

Author contributions statement

A.A. curated the data, implemented the survey, validated the results, and created the visualizations. A.A., D.H., and G.H. wrote the original draft. G.H. supervised the work. All authors reviewed the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

All authors are members of the MOSLA consortium, which has received funding from the Hessian Ministry for Science and the Arts (LOEWE). Specials thanks to Stefanie Dehnen and Bernhard Seeger for their support and expertise.

References

- Jayagopal V, Bassler K. Data management and big data analytics: Data management in digital economy. In: *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques*. IGI Global; 2019. p. 1–23..
- Zwolenski M, Weatherill L. The digital universe: Rich data and the increasing value of the internet of things. *J Telecommun Digital Econ* 2014.
- Giné J. Quantum fluctuations and the slow accelerating expansion of the universe. *EPL (Europhysics Letters)* 2019;125:50002.
- Roser M. Future population growth. *Our World in Data*; 2013. <https://ourworldindata.org/future-population-growth..>
- Bhushan B. *Tribology and mechanics of magnetic storage devices*. (Springer Science & Business Media; 2012.
- Robert W. Phonographic recording; 1962. US Patent 3,023,011..
- Doi, T., Sakata, S., Inoue, T. & Kuse, S. Magnetic tape (2005). US Patent App. 11/052,952..
- Keele RV, et al. Virtual addressing of optical storage media as magnetic tape equivalents; 1995. US Patent 5,455,926..
- Comstock R. Review modern magnetic materials in data storage. *J Mater Sci Mater Electr* 2002;13:509–23.
- Reiss G, Hütten A. Applications beyond data storage. *Nat Mater* 2005;4:725–6.
- Carter JB, et al. System and method for providing highly available data storage using globally addressable memory; 1999. US Patent 5,909,540..
- Bhat WA. Bridging data-capacity gap in big data storage. *Future Gen Comput Syst* 2018;87:538–48.
- Slik D. Estimating data storage device lifespan; 2016. US Patent 9,436,571..
- Lai D-W, Liao J-H, Chang H-T. Methods for measuring usable lifespan and replacing an in-system programming code of a memory device, and data storage system using the same; 2013. US Patent 8,402,204..
- Zenkin N. Ancient rna stems that terminate transcription. *RNA Biol* 2014;11:295–7.
- Huang Y et al. Ancient origin and recent innovations of rna polymerase iv and v. *Mol Biol Evol* 2015;32:1788–99.
- Son A, Horowitz S, Seong BL. Chaperna: linking the ancient rna and protein worlds. *RNA Biol* 2020;1–8.
- Adleman LM. Computing with DNA. *Sci Am* 1998;279:54–61.
- Feringa BL, Jager WF, de Lange B. Organic materials for reversible optical data storage. *Tetrahedron* 1993;49:8267–310.
- Hu B et al. Inorganic-organic hybrid polymer with multiple redox for high-density data storage. *Chem Sci* 2014;5:3404–8.

- [21] Micheloni R et al. Non-volatile memories for removable media. *Proc IEEE* 2009;97:148–60.
- [22] Prüfer K et al. The complete genome sequence of a neanderthal from the altai mountains. *Nature* 2014;505:43–9.
- [23] Orlando L et al. Recalibrating equus evolution using the genome sequence of an early middle pleistocene horse. *Nature* 2013;499:74–8.
- [24] Malthaner W, Vaughan H. An automatic telephone system employing magnetic drum memory. *Proc IRE* 1953;41:1341–7.
- [25] Hoagland AS. Magnetic drum recording of digital data. *Trans Am Inst Electr Eng Part I Commun Electr* 1954;73:381–5.
- [26] Klein D. The history of semiconductor memory: From magnetic tape to nand flash memory. *IEEE Solid-State Circuits Mag* 2016;8:16–22.
- [27] Copeland B. The manchester computer: a revised history part 1: The memory. *IEEE Ann History Comput* 2010;33:4–21.
- [28] Huskey HD. Williams tube memory. In: *Encyclopedia of Computer Science*. GBR: (John Wiley and Sons Ltd.; 2003. p. 1851–3.
- [29] Foss E, Partridge, R.S.A 32,000-word magnetic-core memory. *IBM J Res Develop* 1957;1:102–9.
- [30] Papian WN. The mit magnetic-core memory. In *Papers and discussions presented at the Dec. 8–10, 1953, eastern joint AIEEE-IRE computer conference: information processing systems—reliability and requirements; 1953:37–42.*
- [31] Dee RH. Magnetic tape for data storage: an enduring technology. *Proc IEEE* 2008;96:1775–85.
- [32] Athey SW. Magnetic tape recording, vol. 5038. US Government Printing Office; 1966.
- [33] Al Mamun A, Guo G, Bi C. Hard disk drive: mechatronics and control. CRC Press; 2017.
- [34] Dwivedi N et al. Graphene overcoats for ultra-high storage density magnetic media. *Nat Commun* 2021;12:1–13.
- [35] Rankl W, Effing W. Smart card handbook. John Wiley & Sons; 2004.
- [36] Svigals J. The long life and imminent death of the mag-stripe card. *IEEE Spectrum* 2012;49:72–6.
- [37] Amankwah-Amoah J. Competing technologies, competing forces: The rise and fall of the floppy disk, 1971–2010. *Technol Forecast Social Change* 2016;107:121–9.
- [38] Trikha B. A journey from floppy disk to cloud storage. *Int J Comput Sci Eng* 2010;2:1449–52.
- [39] Bonyhard P et al. Magnetic bubble memory chip design. *IEEE Trans Magn* 1973;9:433–6.
- [40] Chen TC, Chang H. Magnetic bubble memory and logic. *Advances in computers*, vol. 17. (Elsevier; 1978. p. 223–82.
- [41] Chang H. Magnetic-bubble memory technology. New York 1978;6.
- [42] Smoke JP. Video laserdiscs. *Commun Technol Update* 1994;248.
- [43] Kasavajhala V. Solid state drive vs. hard disk drive price and performance study. Proc. Dell Tech. White Paper 2011:8–9.
- [44] Cornwell M. Anatomy of a solid-state drive. *Commun ACM* 2012;55:59–63.
- [45] Chen PP-S. The compact disk rom: How it works: An offshoot of the compact digital audio disc, this multimegabyte storage technique is revolutionizing database technology. *IEEE Spectrum* 1986;23:43–9.
- [46] Baert L. Digital audio technology: a guide to CD, MiniDisc, SACD, DVD (A), MP3 and DAT. Taylor & Francis; 2001.
- [47] Tan E, Vermuelen B. Digital audio tape for data storage. *IEEE Spectrum* 1989;26:34–8.
- [48] Yoshida T. The rewritable minidisc system. *Proc IEEE* 1994;82:1492–500.
- [49] Hoogendoorn A. Digital compact cassette. *Proc IEEE* 1994;82:1479–89.
- [50] Lokhoff GC. Dcc-digital compact cassette. *IEEE Trans Consum Electr* 1991;37:702–6.
- [51] Ernst M. Digital storage. *Interface: J Edu Community Values* 2003;3.
- [52] Leventhal A. Flash storage memory. *Commun ACM* 2008;51:47–51.
- [53] Yamada H. Dvd overview [removable storage media]. In: *Proceedings IEEE COMPCON 97. Digest of Papers*. IEEE; 1997. p. 287–90..
- [54] Scheible JP. A survey of storage options. *Computer* 2002;35:42–6.
- [55] Ajwani D, Malingier I, Meyer U, Toledo S. Characterizing the performance of flash memory storage devices and its impact on algorithm design. In: *International Workshop on Experimental and Efficient Algorithms*. Springer; 2008. p. 208–19..
- [56] Mordvinova O, Kunkel JM, Baun C, Ludwig T, Kunze M. Usb flash drives as an energy efficient storage alternative. In: *2009 10th IEEE/ACM International Conference on Grid Computing*. IEEE; 2009. p. 175–82..
- [57] Gay WW. Sd card storage. In: *Raspberry Pi Hardware Reference*. Springer; 2014. p. 81–8..
- [58] Fontana Jr R, Decad G, Hetzler S. Volumetric density trends (tb/in. 3) for storage components: Tape, hard disk drives, nand, and blu-ray. *J Appl Phys* 2015;117:17E301.
- [59] Shida N et al. Multilayer optical read-only-memory disk applicable to blu-ray disc standard using a photopolymer sheet with a recording capacity of 100 gb. *Japanese J Appl Phys* 2004;43:4983.
- [60] Horimai H, Tan X. Holographic information storage system: today and future. *IEEE Trans Magn* 2007;43:943–7.
- [61] Horimai H, Tan X. Holographic versatile disc system. *Organic Holographic Materials and Applications III*, vol. 5939. (International Society for Optics and Photonics; 2005. p. 593901.
- [62] Deepika G. Holographic versatile disc. In: *2011 National Conference on Innovations in Emerging Technology*. IEEE; 2011. p. 145–6..
- [63] Lunt BM, et al. Toward permanence in digital data storage. In: *Archiving Conference*, vol. 2013. Society for Imaging Science and Technology; 2013. p. 132–6..
- [64] Cox JP. Long-term data storage in DNA. *TRENDS Biotechnol* 2001;19:247–50.
- [65] Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. *Science* 2012;337:1628.
- [66] Goldman N et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 2013;494:77–80.
- [67] Grass RN, Heckel R, Puddu M, Paunescu D, Stark WJ. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew Chem Int Ed* 2015;54:2552–5.
- [68] Bornholt J et al. A DNA-based archival storage system. In: *Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems*. p. 637–49.
- [69] Erlich Y, Zielinski D. DNA Fountain enables a robust and efficient storage architecture. *Science* 2017;355:950–4.
- [70] Organick L et al. Random access in large-scale DNA data storage. *Nat Biotechnol* 2018;36:242.
- [71] Yim AK-Y et al. The essential component in DNA-based information storage system: robust error-tolerating module. *Front Bioeng Biotechnol* 2014;2:49.
- [72] Sun J et al. Digital information storage on DNA in living organisms. *Med Res Arch* 2019;7.
- [73] Hao M et al. A mixed culture of bacterial cells enables an economic DNA storage on a large scale. *Commun Biol* 2020;3:1–9.
- [74] Phillips CL et al. Digital colloids: reconfigurable clusters as high information density elements. *Soft Matter* 2014;10:7468–79.
- [75] Kalf F et al. A kilobyte rewritable atomic memory. *Nat Nanotechnol* 2016;11:926–9.
- [76] Kennedy E et al. Encoding information in synthetic metabolomes. *PLoS One* 2019;14:e0217364.
- [77] An W. Pulse transfer controlling device; 1955. US Patent 2,708,722..
- [78] IBM. IBM 350 disk storage unit; 2021. URL: https://www.ibm.com/ibm/history/exhibits/storage/storage_350.html..
- [79] Pinheiro E, Weber W-D, Barroso LA. Failure trends in a large disk drive population. In: *Proceedings of the 5th USENIX Conference on File and Storage Technologies, FAST '07*, 2. USA: USENIX Association; 2007.
- [80] Schroeder B, Lagisetty R, Merchant A. Flash reliability in production: The expected and the unexpected. In: *14th USENIX Conference on File and Storage Technologies (FAST 16)*. Santa Clara, CA: (USENIX Association; 2016. p. 67–80.
- [81] ECMA. ECMA-367: Eiffel analysis, design and programming language (ECMA 2006). Geneva, Switzerland: European Association for Standardizing Information and Communication Systems; 2006.
- [82] Vitter JS. External memory algorithms and data structures: Dealing with massive data. *ACM Comput Surv (CSUR)* 2001;33:209–71.
- [83] Blawat M, et al. Forward error correction for DNA data storage. *Procedia Comput Sci* 2016;80:1011–22. International Conference on Computational Science 2016, ICCS 2016, 6–8 June 2016, San Diego, California, USA..
- [84] Liva G, Paolini E, Chiani M. Performance versus overhead for fountain codes over f q. *IEEE Commun Lett* 2010;14:178–80.
- [85] Kalcher S, Lindenstruth V. Accelerating galois field arithmetic for reed-solomon erasure codes in storage applications. In: *2011 IEEE International Conference on Cluster Computing*. IEEE, 2011. p. 290–8..
- [86] Dong Y, Sun F, Ping Z, Ouyang Q, Qian L. DNA storage: research landscape and future prospects. *Natl Sci Rev* 2020;7:1092–107.
- [87] Rosemann NW et al. A highly efficient directional molecular white-light emitter driven by a continuous-wave laser diode. *Science* 2016;352:1301–4.
- [88] Rosemann NW, Eufner JP, Dornsiepen E, Chatterjee S, Dehnen S. Organotetrel chalcogenide clusters: Between strong second-harmonic and white-light continuum generation. *J Am Chem Soc* 2016;138:16224–7.
- [89] Dornsiepen E et al. White-light generation upon in-situ amorphization of single crystals of $\{[(\text{Me}_3\text{P})_3\text{AuSn}](\text{phsn})_3\text{s}_6\}$ and $\{[(\text{Et}_3\text{P})_3\text{AgSn}](\text{phsn})_3\text{s}_6\}$. *Adv Opt Mater* 2019;7:1801793.
- [90] Wei P, Li B, Pentzer E, et al. Beyond binary: optical data storage with 0, 1, 2, and 3 in polymer films. *J Mater Chem C* 2017;5:5780–6.
- [91] Green JE et al. A 160-kilobit molecular electronic memory patterned at 10 11 bits per square centimetre. *Nature* 2007;445:414–7.
- [92] Kazansky P et al. Eternal 5d data storage via ultrafast-laser writing in glass. *SPIE Optoelectr Commun* 2016.
- [93] Unden T et al. Coherent control of solid state nuclear spin nano-ensembles. *npj Quant Inf* 2018;4:1–6.
- [94] Staley DJ. *Computers, visualization, and history: How new technology will transform our understanding of the past*. Routledge; 2015.
- [95] Munzner T. *Visualization analysis and design*. CRC Press; 2015.
- [96] Hattab G, Rhyne T-M, Heider D. Ten simple rules to colorize biological data visualization; 2020..
- [97] Bederson BB, Shneiderman B, Wattenberg M. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *AcM Trans Graph (TOG)* 2002;21:833–54.
- [98] Van Rossum G, Drake Jr FL. *Python tutorial*. The Netherlands: Centrum voor Wiskunde en Informatica Amsterdam; 1995.
- [99] Satyanarayan A, Moritz D, Wongsuphasawat K, Heer J. Vega-lite: A grammar of interactive graphics. *IEEE Trans Visual Comput Graph* 2017;23:341–50.
- [100] VanderPlas J et al. Altair: Interactive statistical visualizations for python. *J Open Sour Softw* 2018;3:1057. <https://doi.org/10.21105/joss.01057>.

- [101] Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC; 2020.
- [102] van der Valk T et al. Million-year-old dna sheds light on the genomic history of mammoths. *Nature* 2021;591:265–9.
- [103] Chandak S, Tatwawadi K, Ochoa I, Hernaez M, Weissman T. SPRING: a next-generation compressor for FASTQ data. *Bioinformatics* 2018;35:2674–6. <https://doi.org/10.1093/bioinformatics/bty1015>. https://academic.oup.com/bioinformatics/article-pdf/35/15/2674/29012442/bty1015_supplementary_data.pdf.
- [104] Deorowicz S. Fqsqueezzer: k-mer-based compression of sequencing data. *Sci Rep* 2020;10:1–9.
- [105] Silva M, Pratas D, Pinho AJ. Efficient DNA sequence compression with neural networks. *GigaScience* 2020;9. <https://doi.org/10.1093/gigascience/giaa119>.
Giaa119. https://academic.oup.com/gigascience/article-pdf/9/11/giaa119/34251854/giaa119_supplemental_file.pdf.
- [106] Matsumoto T, Dunnigan T, Back M. Post-bit: embodied video contents on tiny stickies. In: *Proceedings of the 13th annual ACM international conference on Multimedia*. p. 263–4.
- [107] Matsumoto T, Back, M. Post-bit: multimedia epaper stickies; 2007. US Patent 7,195,170.
- [108] Polat EO et al. Graphene-enabled optoelectronics on paper. *Acs Photon* 2016;3:964–71.
- [109] Zhang W, Zhang Q, Zhao M-Q, Kuhn LT. Direct writing on graphene 'paper' by manipulating electrons as 'invisible ink'. *Nanotechnology* 2013;24:275301.