ISSN No:-2456-2165

The Paradox on Long Term Digital Preservation and the Optical Storage

Ahmed Ibrahim Safana[1] ; Muhammad Abubakar Ibrahim[1]
Computer Department School of Science, Federal College of Education, Katsina

Abstract:- In the field of data storage, long-term digital preservation is a crucial concern. Due to their great capacity and low price, magnetic media-based solutions, such as tape and hard disk drive (HDD)-based archive systems, have long held a monopoly on the data archiving industry. The volume, velocity, and variety of data sets, which are quickly growing in the big data era, provide a number of problems to archive systems in a range of areas, including capacity, cost, performance, reliability, power consumption, and so on. With the resurgence of optical storage in recent years, large capacity optical media, like blu-ray discs (BDs) and holographic discs, have emerged. The archive systems based on those optical media, for example, the BD library, exhibit appealing qualities, such as cost per bit, reliability, power consumption, and so on, and as a result, become viable solutions in long-term digital preservation. This is because optical media have a naturally simple design. The research examined and compared magnetic and optical media-based solutions for long-term digital preservation, and then summarized ways to enhance the long term optical media-based archive system.

I. INTRODUCTION

Everyone is being able to capture, access, and use the world's accumulated digital knowledge as a result of accessible Internet access. For everyone's immediate access, people, organizations, enterprises, and governments spend time and resources creating and capturing digital information. A third of the expansion in US economic activity since 1992 has come from companies that manufacture computers, semiconductors, software, and communications equipment. The goal of researchers, archives and historians is to make this knowledge accessible to all communities. Unfortunately, given the rapid advancement of technology, the stated goal is not for granted due the nature of the ongoing preservation and accessibility of digital material.

Despite numerous efforts in information technology, the information infrastructure has a serious cumulative flaw. Short media life, outdated gear and software, poor read speeds of old media, and inactive Web sites are problems for the long-term preservation of digital information. In fact, the vast majority of goods and services available today were unimaginable just five years ago (Barrueco & Termens, 2022). More significantly, the system lack tested ways to guarantee that the digital information will continue to exist, that the public will be able to access it using the technical tools at our disposal, or that any information that is accessible is accurate and trustworthy (Antoniazzi, 2021). In the modern world, the long-term preservation of digital information is a crucial concern. There are always certain data that should be maintained, from the entire civilization to a single person. Additionally, a portion of the preservation is required by law; for instance, the majority of account information must be kept for at least seven years. It is recommended to keep the design documentation for more than 15 years. Additionally, vital medical records should be kept for at least 30 years. Particularly, the majority of those data, often referred to as cold data, are infrequently viewed (Masenya & Ngulube, 2020). Due to the enormous capacity and low cost of tape and hard disk drive (HDD)-based systems, they have long dominated the data archiving market. However, recently, things have changed. The total amount of data in the globe in 2021 was approximately 79 ZB (1 ZB = 106 PB = 1018 Bytes), and it would expand at a rate of over 50% per year, according to a report published by the International Data Corporation (IDC).

Instead of increasing at the same rate as linear tape open (LTO) tapes and magnetic tapes, which was predicted by the well-known Kryder's law, the growth ratio of storage density of magnetic tapes and disks slows down (Cherubini, 2022). For example, after the year 2010, HDD areal density increased at a rate of 20% per year, while that of LTO tapes increased at a rate of about 40% per year. The discrepancy between rising data ratios and storage densities will have a significant impact on the archive system and have broad implications for the storage sector (Cherubini, 2022). The archive system must also contend with the difficulty of power consumption as the data set expands. With the resurgence of optical storage in recent years, high capacity optical media have emerged. For example, the capacity of a BD ranges from 100 to 128 GB in the current generation (BDXL, high capacity recordable and rewritable discs), and that of the next generation is up to 500 GB. Additionally, compared to the tape (15–30 years), HDD (3–5 years), and optical medium (e.g., 50–100 years for BDXL), the optical medium has a far longer lifetime. Because there will be less migration, the optical media-based archive system will also have low power consumption in addition to excellent reliability (Chatzieleftheriou, Stefanovici, Narayanan, Thomsen, & Rowstron, 2020). The optical media-based archive system develops into a desirable alternative for long-term digital preservation as a result of all the aforementioned criteria, including capacity, dependability, and power consumption. Prototypes of BD-based archiving systems have been created, developed, and displayed over the last one to two years. This paper first reviewed the

ISSN No:-2456-2165

background of long-term digital preservation in the big data era, then compared current magnetic and optical media-based archive systems, proposed solutions to the optical media-based archive system's shortcomings, and made conclusions.

Take a look at this historical snapshot of irretrievable data from the previous 50 years to get an idea of what is at stake: The majority of TV interviews, the first e-mail written in 1964, 50% (or around 25,000) of the films made in the 1940s, as well as many pieces of intellectual and cultural legacy. We should legitimately be receiving long-term benefits as a return on our professional, personal, and financial investments in information technology. However, if we don't address the issues associated with digital information preservation, we risk creating cultural and intellectual poverty and wasting potential long-term gains. Even in the short term, this failure might result in significant recovery costs and unforeseen catastrophes.

II. INCONSISTENCIES IN DIGITAL PRESERVATION

In the past, maintaining something meant keeping it unchanged, but the digital world has fundamentally altered how we think about preservation needs. Digital information will become increasingly difficult to access, if not impossible, if we keep it around without making any changes. The formats for storing the information would change, and the gear and software required to extract the information would become dated even if we could develop a physical medium to permanently store unaltered digital data (Moles, 2022). For instance, if retrieving historical data from the Vietnam War required us to master 30-year-old technology, very few of us would ever contemplate doing so. A basic dilemma for digital preservation is created by this circumstance: We want to access digital material dynamically and with the most cutting-edge tools, while still preserving it exactly as it was when it was first created (El-Fakdi, & de la Rosa, 2022).

An important research challenge is figuring out how to alleviate the conflict between the contexts of creation and consumption. The research must examine needs for content, formats, and styles, context, storage media, systems technology, work-flow processes, and metadata policies in order to assess the problems with preserving digital information (Larson, 2020). The research can only focus on the fundamental concerns because of the complex and unexplored interdependencies between these topics.

III. REQUIREMENTS FOR DIGITAL PRESERVATION

Information that is "born digital" and information that already exists in another form but is converted or captured in digital form are both referred to in the creative context as digital information production. The creation document establishes the fundamental conditions for protecting digital information in three different ways. The information content that needs to be kept is first provided by the creation context. Second, it records this information in particular styles and forms that are crucial for preserving the accuracy of the data. Third, understanding digital information frequently requires knowledge of the context in which it was produced, namely who produced it, for what objectives, and how they arranged and processed it (Ocen, Mutua, Mugeni, Karume, & Matovu, 2019).

- *Content*:

The term content itself is the first issue with digital preservation. We deal with static items when maintaining analog objects (like books and music), making it easy to take their entire contents and store them in some manner. We must reevaluate the significance of object preservation in order to deal with digital content. For instance, links on websites can both alter and connect to other websites that are dynamically updating. New concerns about what it means to preserve a digital thing arise as it develops and evolves through time.

- *Data Format and Style*:

The ability to create a growing variety of recorded information types is made possible by computer technology. Digital versions of traditional forms including books, maps, photos, and sound recordings are still produced. There are now new formats, including interactive video, dynamic pages, multi-media, geographic information systems, and hypertext. For digital preservation, each format or style has unique obstacles (such as encoding and compression).

- *Context:*

To further grasp the contextual factors, we must conduct basic study. This research is required to help define, develop, and evaluate techniques to preserve the content, formats, and styles, as well as the context of digital information across information-technology generations; develop criteria and methods for evaluating and demonstrating the authenticity of the preserved information; and articulate techniques for evaluating the long-term value of digital information, including techniques for articulating what it says about the past.

- *Storage Media*:

The short shelf life of media, out-of-date gear and software, and sluggish read speeds of ancient media are all problems for digital preservation. Rapid technological advancements do not provide a solution; rather, we must periodically shift digital content from one technology generation to another. The preservation concerns for digital documents go beyond media life considerations. Because of the different formats for digital documents and photos, reading devices for these media quickly become obsolete. Accurate conceptions and fundamental characteristics of preservation in the life cycle of digital information are provided by research-based development of policies, procedures, standards, and protocols.

- *Systems Technology:*

The landscape of digital information preservation has evolved as a result of the existing communication infrastructure, which includes the Internet, personal computers, cable television, and data storage. It is crucial to take preservation difficulties into account at every step of the life cycle. Although there won't be a solely technological answer, it is crucial to comprehend the long-term research problems surrounding digital information preservation. The four main areas of information technology research software, scalable information infrastructure, high-end computing, and socio-economic impact. These areas of study are all pertinent to problems with data archiving and information preservation. Software needs to be reusable and information-preserving. To maintain digital information, the information infrastructure must be scalable throughout time. High-end computing produces scientific and critically important data for the country, which needs to be stored for the long term. Preservation-related concerns are crucial to the "socio-economic" research field.

- *Workflow Process:*

The widespread use of computers facilitates the creation and dissemination of information. However, it also increases the bar for how well we can use the information. Through the Internet, we may access information that is located anywhere in the globe, but we frequently have no idea who is providing it, if the information is accurate, and whether it truly appears as intended, let alone what it actually means. Although these issues are still in their infancy in the digital age, the transmission of documentary materials through generations will increase their significance and impact. To distinguish between information values, operational systems and archives, working documents and archived records, and other information access difficulties, it will be necessary to employ the workflow process, a concept used in office automation and electronic commerce. The ability to retrieve information stored anywhere in the world quickly and the creation of more robust, adaptable retrieval technologies have increased both the demand for access and the standard of service required in response. The established idea of archives and libraries as locations where people go to receive information is invalidated by rising demand. The repositories must transmit the saved information to users anywhere in the world in the digital environment. The solutions must be able to interface with the digital access systems depicted in Figure 1 in order to allow the migration of digital records between generations of technology and humans. The integrity of the recorded information given through interfaces, such as those between access systems and end users and between access systems and preservation systems, must always be protected while maintaining interoperability.

- *Polices for Metadata:*

There must be an increasing amount of metadata attached to save digital recordings due to the discrepancy between the contexts of creation and future use. Some metadata, such copy-rights and process information, should be updated on a regular basis. Despite the viability of data storage, it will be expensive to provide and manage necessary metadata. How much metadata is required to adequately preserve the unknown category? To transform policies into metadata decisions, we need both policies and automated technologies. Standards, languages, data structures, links to conceptually or physically remote ancillaries, presenters' and interpreters' software, terminology (document type definition, reference, and ontology), and contexts are all represented by metadata for each stored record. Although we have little control over short media lives and out-of-date equipment and software, sensible metadata regulations can help to address these serious problems. The criteria used to make archive decisions (such as metadata) are influenced by economic and social models, storage and software expenses, and costs associated with human resources.

Although adding more semantics to metadata may increase expenses, it will reduce the need for human interaction when accessing data and improve lifetime maintenance, transition of stewardship, and seamless support. Preservation is a probabilistic art, not an exact science. Around 1999, the information content provider Corbus noticed a loss rate of 150 out of every million photos in storage systems. What maximum permissible loss rates will there be? In particular domains, cost/benefit analyses should be part of quantitative measurements.

Initiating particular future requirements based on known and recognized requirements is one option available to domain specialists. The system design for preservation and archival purposes will be impacted by these needs. On the other hand, as digital technologies become more prevalent, people have personal archiving and preservation demands (for instance, for photographs taken with digital cameras). Decentralized versus centralized systems, multilevel control, scalability, dependability, redundancy, access links, and standards enforcement methods are a few examples of research topics. Workflow procedures continuously produce documents and records in businesses. We should link the workflow process with the preservation process appraisal, verification, maintenance, and eventually retirement—starting with creation and ingestion.

By enhancing digital technology to create automated tools, archivists can access information at many service levels (professionals, the general public, and individuals), choose and attach the appropriate metadata to a digital record, and exchange documents in ways that go beyond what is now possible. 1. The container concept popularized by the software industry (for instance, Apple's Bento) is helpful for combining metadata with records for aggregation, exchange, and emulation. 2. The NASA standard for digital preservation is the Reference Model for an Open Archival Information System. 3. A good example of container standards for multimedia content, including text, images, sound, video, and their interlinkages, is the UPF (Universal Preservation Format) project.

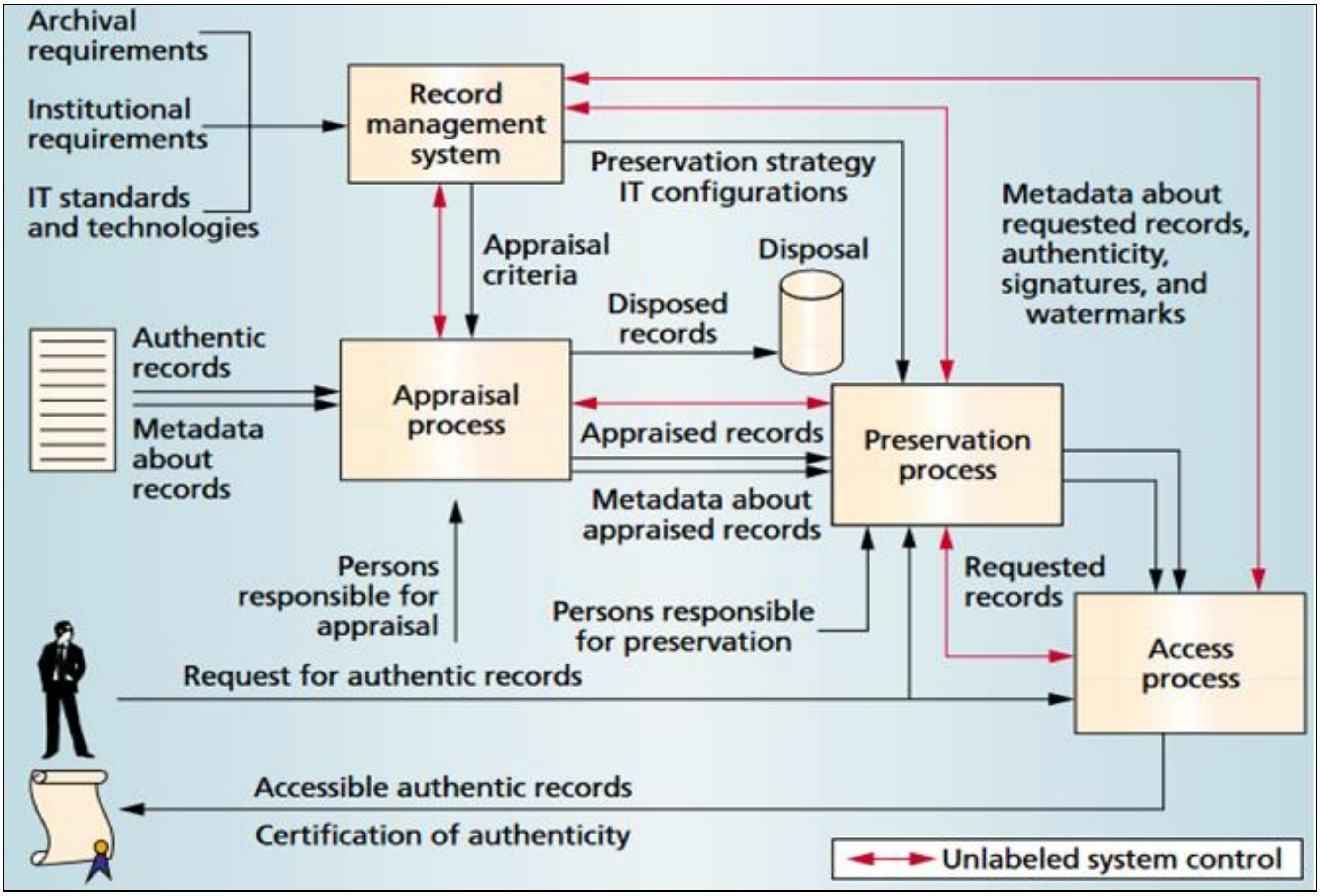

Fig 1 Report on the Survey on Data Retention by the SNIA

IV. LONG-TERM DIGITAL PRESERVATION VERSUS BIG DATA

The research will review the history of long-term digital preservation in the big data era in this section. As it was mentioned in above, a report by the IDC claims that the total amount of data in the world in 2022 was approximately 79 ZB and would increase at a speed of over 50% per year, i.e., it will be approximately 99 ZB in the year 2023 as shown in Fig. 1. However, even this prediction about data increasing ratio is thought to be conservative. The NSA Utah Data Center, whose construction began in 2011, will allegedly be the "first facility in the world expected to gather and house a yottabyte" (1 yottabyte = 1 YB = 1024 Bytes), according to Utah Governor Gary R. Herbert. Additionally, it is stated that the BRAIN research will produce approximately 1 YB of data in total.

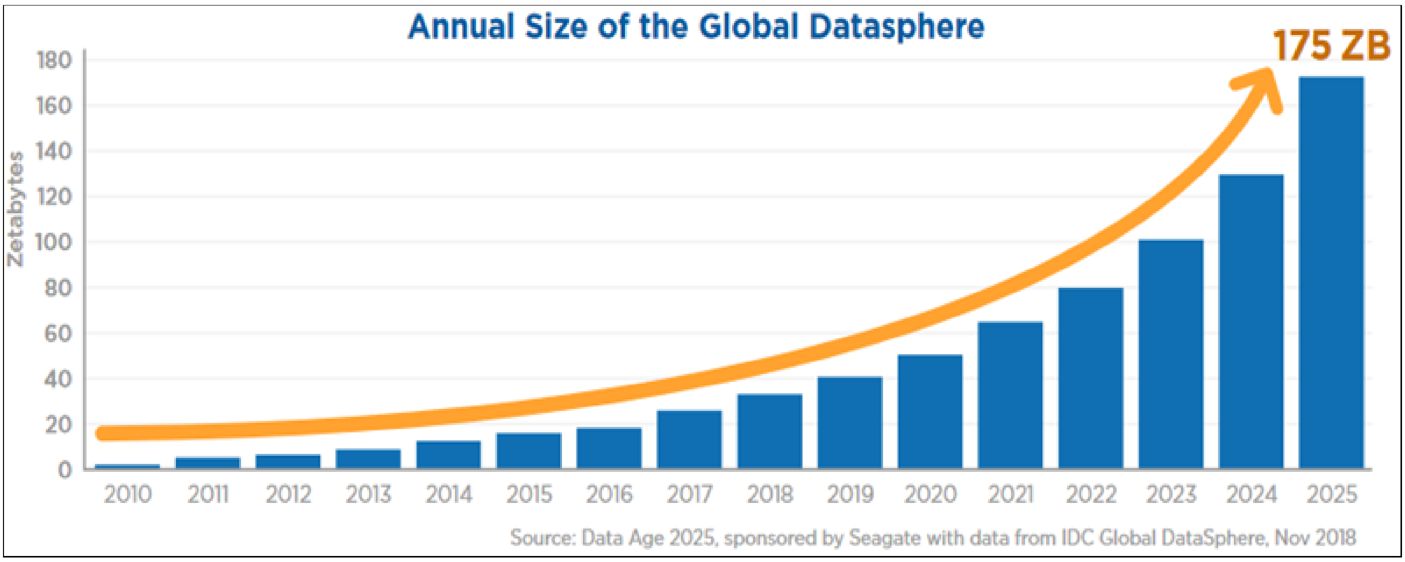

Fig 2 Annual Size of the Global Data

In addition to scientific data, there are vast amounts of commercial and personal data that should be preserved. Storage Networking Industry Association (SNIA) conducted a survey of 276 firms in 2019 about long-term digital preservation. Some survey findings are shown in Figure 2. Ninety percent of those businesses must keep their digital data for more than ten years. Even more than 50 years of data preservation is required, according to 81% of respondents. Additionally, 18% of the respondents stated that they need to save more data than 100 TB. It's also important to maintain the enormous data set produced by large-scale web services, like social networking. According to reports, the majority of photos posted to Facebook are never or very never seen; for example, 97% of content only receives 29% of requests. However, the corporation chooses to keep those cold data rather than delete it because of the enormous prospective value of those photographs. Facebook is currently building a data center to maintain those cold data. One option available to them is the BD-based archival system.

The act of maintaining information over a period of decades or longer in a correct and independently understandable form is referred to as "long-term preservation" in the most recent definition provided by SNIA, and "digital preservation" is defined as "Ensuring continued access to, and usability of, digital information and records, especially over long periods of time." Their concept states that the long-term digital preservation is not just constrained by capacity and price (or cost per bit). Many other aspects, such as the dependability, power consumption, and so on, should also be taken into account in order to maintain data at ZB or even YB level. Although the latter two criteria can generally be translated into monetary costs, we do not do so in this study because it is difficult to calculate the loss of a valuable piece of data or the emission of a ton of CO2. In order to examine the practical long-term digital preservation systems, we looked at the following five factors: storage density, storage cost, performance, dependability, and power consumption.

In addition to the aforementioned elements, other concerns, such as security, semantics, and other related matters, are also crucial to archive systems. For instance, Front Poach Digital suggests the Archive eXchange standard (AXF), an open standard for the long-term storage, transfer, and preservation of all types of content. Since the previous five metrics that we chose for comparison are more pertinent to the storage media, we do not address these issues in this study.

➢ *Comparisons of Some Feasible Solution*

In this part, the research first discussed potential long-term digital storage media for the big data era, and then we compared archive systems based on those media from the viewpoints of the storage medium (including the drive) and the entire system, respectively.

- *Non-Volatile Storage Media*

Kryder and Kim examined 13 different innovative nonvolatile memory technology types that are potential HDD replacements. This paper uses the well-known Kryder's law, "The 30-year history of the cost of digital storage media dropping exponentially," to forecast the characteristics of those memory technologies in the future. The economics of long-term digital storage were thoroughly examined by Rosenthal et al.

In their work, disk, tape, and NAND-flash storage methods were contrasted. They came to the conclusion that, among the three storage media mentioned above, tape is most likely to be the best option for long-term digital preservation. This is because tapes are relatively inexpensive and have the ability to increase storage density. Fontana also contrasted the development of tape, HDD, and NAND-flash technologies. Only the tape exhibits a possible large (over 40% per year) increasing ratio of storage density, according to their conclusions. Due to their relatively long lifespans and low cost per bit, the tape, HDD, and BD based archive systems were taken into consideration in this work. The NAND-flash-based archiving systems are not taken into consideration for the following two reasons: 1) the cost of NAND-flash is substantially higher per bit than other media; for example, it costs around 10 times as much per bit as an HDD. 2) The NAND-flash stores digital data by introducing electronics into field effect transistors. But this recording system is unstable.

Particularly in a multiple level cell (MLC, a cell records few bits) based NAND-flash memory, those circuitry will slowly evaporate and cause loss of stored information. This would result in numerous additional migration and refresh processes.

- *Perspective from the Storage Media*

The research compared the quantitative performance of HDD, tape, and BD. Table 1 presents the outcomes. Because they are more affordable per bit than 2.5 inch SAS HDDs, we decided to compare 2 TB 3.5 inch SATA HDDs. The LTO6 tape, the most recent version of LTO tapes, was chosen for the comparison because LTO tapes dominate the magnetic tapes market. Because the former is the standard production and the latter has previously been utilized in the first generation of BD-based archive systems, the 25 GB BD-R disc and 100 BDXL disc were selected for comparison.

✓ *Storage Density*:

Although the BD currently has the lowest storage density compared to the other storage mediums, this is expected to change significantly in the near future. For instance, the capacity of the BD's upcoming generation can reach 500 G.

✓ *Cost Per Bit*:

The tape and BD-R (25 GB) have the lowest costs per bit, while the BDXL has the highest. The cost per bit of an HDD can be two to three times higher than that of a tape. BD-R (25 GB). Because BDXL has not been made in large quantities, its cost per bit is extremely expensive. To put it another way, the price may decrease even more if the optical storage media manages to increase its market share.

✓ *Recording Bandwidth*:

When compared to other storage technologies, the recording bandwidth of optical storage technologies now in use is significantly less. For instance, the 16X BDR can record at up to 72 MB/s (1X = 4.5 MB/s) and often does so at 45 MB/s. The BDXL version's is much less, at only 18 MB/s. additionally, the increase of the recording layer is what drives the improvement of BD's storage density, not the growth of the recording bandwidth.

✓ *Reliability*:

Essentially, one key parameter to consider when assessing systemic reliability is longevity. The lifetime of HDD (also known as MTTF, or Mean Time to Failure of HDD) is substantially shorter than anticipated even though the magnetic media is relatively stable. The lifespan of tape is not influenced by the magnetic media, like it is with HDDs. According to promises made by Fujifilm and HP, the lifespan of their most recent products is roughly 30 years. But because the tape is folded up, it is simple to stick everything together. As a result, the tape's content needs to be frequently updated by being copied from one cassette to another. This refresh operation adds a lot more management work. The recording medium used in an optical medium typically determines how long it will last. The lifetime of BDXL normally varies from 50 to 100 years, according to the HITACHI research. Additionally, the M- DISC, a RITEK device, has a lifespan of roughly 1000 years.

✓ *Power Usage*:

The ratio of recording bandwidth to power is used to normalize power usage.

The power consumption of the BD-based archive system is significantly lower overall than that of the HDD- and tape-based archive systems for long-term digital preservation, despite the fact that the BD drive suffers from the worst power consumption during the recording process due to its low recording bandwidth. Essentially, the power consumption of migrations determines the total power consumption of the archive system because accessing alternatives are scarce in long-term digital preservation. Due to the lengthy lifespan of BD, as we discussed in Section 1, the BD-based archive system has significantly fewer migration delays.

Table 1 Quantitative Comparisons Among HDD, Tape, and BD

	HDD	**Tape**		**BD**	
Type	3.5" SATA	LTO6	BD-R		BDXL
Capacity/TB	2	2.5	0.025		0.1
Storage/Volume Density/(GB Cm³)	5.1	10.8	0.85		3.4
Cost Per Bit/(S-GB')	0.5-0.10	0.026-0.036	0..24-0.035		0.45-0.68
Latency/Ms	~10	-45000		~1.00	
Recording Bandwidth/(MB-S')	-100	-160	-45		-18
Lifetime/Year	3-5	15-30		50-100	
Recording Power Consumption/(MB-J')	8.3-12.5	6.7-13.4	1.5-3.0		0.6-1.2

• *View from the Achieve System*

According to the various organizations, archive systems with robot arms can typically be divided into two categories. The other is rack-style, while the first is rack-server style. The BD-based archive systems arranged in the aforementioned two methods are shown in Figure 2. In the old design, a single server houses both the access components (BD drives) and the storage medium (BD). To increase capacity and recording bandwidth, many servers can be placed in racks. In the latter design, a sizable rack houses both the storage devices and the Blu-ray drives. In the rack style, there is only one delivery component (robot). As a result, the capacity and storage density (spatial density) can both be increased.

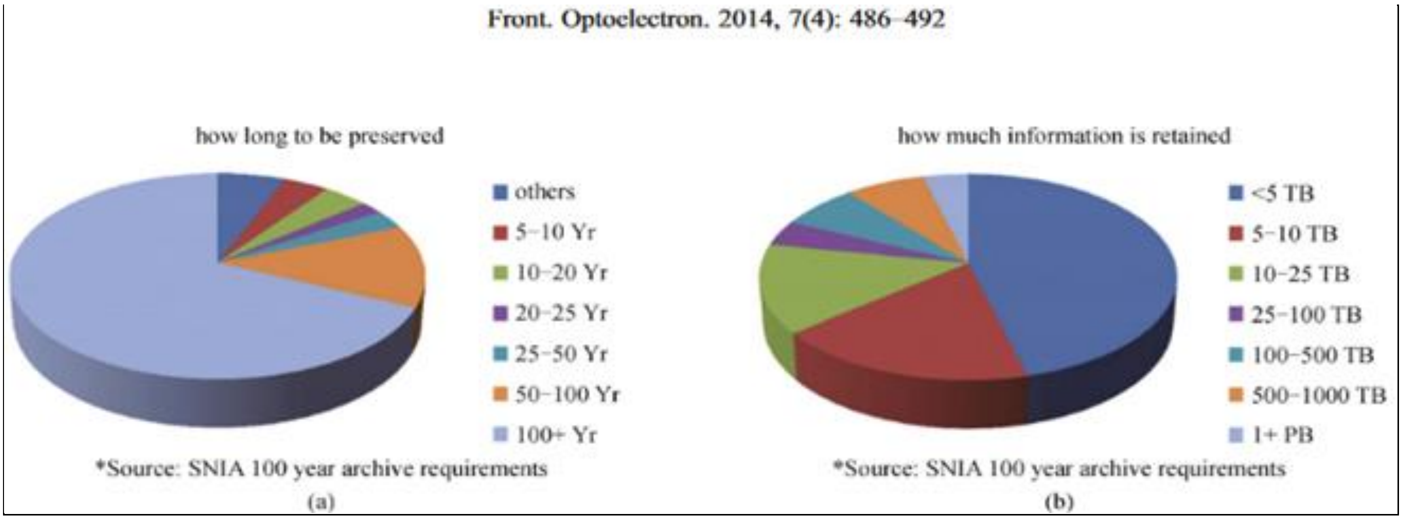

Fig 2 Result of Survey on Data Retention by SNIA (a) Requirements on Retention Time; (b) Amount of Data to Preserve

The rack-server style organization provides better performance, reliability, and scalability than the rack style organization, but it has a greater maintenance cost due to the additional CPU, memory, mainboard, and other components. All archive systems, including those that are HDD, tape, and BD based, can also adopt the rack-server architecture. However, to our knowledge, only tape- and BD-based archive systems can accommodate the rack layout because the storage media and the access components (drives) are physically separated.

Since the expenses of the additional component can vary greatly depending on the organizations, we did not conduct quantitative comparisons at the systemic level. Additionally, analyses of prices under specific configurations show that the costs of HDD, tape, and BD-based systems are 7.7, 2.9, and 1.6 million US dollars, respectively, to archive 100 TB of data for 20 years. The power consumption of HDD-based archive systems is regarded as the poorest among all aforementioned archive systems, despite the fact that HDDs have the best power efficiency during the recording process. According to reports, Facebook's prototype BD-based archive system can save costs by further 50% and energy use by 80% when compared to an HDD-based archive system that uses shingled recording. Because HDDs have a limited lifespan, data migration might result in significant additional power usage. And with tape-based archival systems, the scenario is similar. The BD-based archive system achieves the lowest power consumption thanks to the comparatively long lifetime.

V. IMPROVED THE BD BASED ARCHIVE SYSTEM

The BD-based archiving system is appealing, but it still has a number of problems. In this section, we talk about and provide possible solutions to such problems.

- *Recording Bandwidth*

Without a doubt, the BD-based archiving system's biggest shortcoming is its inadequate recording bandwidth. The bandwidth can be increased using parallel approaches to meet the demand for the velocity. First, discrete chunks of the desired data are created. After that, each of those portions is recorded using a different drive on a different disc. Furthermore, a separate write buffer should be set up to handle minor write requests and prevent data from being recorded on several tracks on a single disc.

- *Access Latency*

Regardless of whether the archive system is set up in the rack-server style or the rack style, the majority of those discs should be gathered and stored outside of the drives in order to achieve high recording bandwidth and PetaByte or ExaByte level capacity. Obviously, this will result in a large lag when accessing data saved on discs not connected to the drive. Since the discs must be put into the drives one at a time, the delay in the rack-server type BD based archive system, for instance, is roughly 65 s. By hiding or reducing that delay, it will thereby broaden the scope of applications for the BD-based archive system. The following methods could be applied to reduce the poor latency:

✓ *Buffer Catch*:

Buffer cache: To reduce the typical access latency, buffer cache is frequently used in storage systems. Performance can be improved by hiding the high latencies associated with accessing slow back end devices by collecting re-accesses to hot spots utilizing memory with low latency. As a result, a buffer cache may be used as the front end of the archive system to decrease the average access latency in such cases.

✓ *Priority based Schedules*:

Priority based schedulers are also frequently employed in computer architectures to lower access latency, especially in those latency-sensitive situations. The access time can be reduced compared to requesting the entire data block sequentially by precisely locating and sending the necessary data in a data block (the fundamental I/O unit). As a result, a priority based scheduler may significantly reduce the latency of accessing data in back end devices to improve the access latency in an archive system that must access the storage media one by one.

✓ *Other Approaches*:

In addition to the approaches mentioned above, new mechanical components that may insert all discs simultaneously may successfully reduce the latency.

- *Reliability*

Although it is believed that the BD is significantly more dependable than disks and tapes, the BD-based archive system still struggles with data dependability, especially in large-scale systems that manage millions of discs. The next two methods could be applied to solve this issue.

✓ *Data Redundancy*:

Keeping the storage system reliable often involves using data redundancy. It can prevent data loss due to media failures by creating redundant information from the original data and utilising that redundant information for recovery. Typically, replicating and erasure coding are two additional categories of data redundancy. The spatial utility (which can be estimated as: *amount of data/ amount (data + redundancy)*) and recovery speed (which determines an important parameter of dependability in storage systems, mean time to repair (MTTR)) may be the main differences between those two techniques. A storage system using two failure-tolerant erasure codes is RAID-6, for instance.

- *Sampling:*

In optical disc-based storage systems, sampling is typically used to identify and anticipate faults. Instead of continuously monitoring all discs, it can achieve great efficiency and accuracy by intermittently inspecting the sample discs.

- *Other Issues*

The BD-based archive system needs to handle a number of other challenges in addition to performance and reliability, including manageability, quick lookup, spatial efficiency, scalability, semantics, security, and others.

✓ *Manageability*:

The conventional optical disc file systems, including universal disc format (UDF), are built for a single disc and are not suitable for volumes spread across numerous discs. Therefore, new file systems should be created for the aforementioned scenario.

✓ *Fast Look Up*:

Finding data on CDs that have been stored off the drive at the Petabyte or Exabyte level is difficult. To solve this issue, two methods can be used. One is taking in the fact that the text is recognized and retrieved from graphic data as the lookup index. Another method for lowering the amount of I/Os required to access a single data chunk is to use multiple level meta-data.

✓ *Spatial Efficiency*:

Deduplication can significantly minimize the quantity of data that needs to be preserved, which boosts spatial efficiency. Deduplication is challenging to implement in optical disc libraries, as it is in tape-based archives. We suggest performing localized deduplication rather than the conventional global deduplication as a result. It can lower both the total quantity of fingerprints and the total amount of data by performing deduplication between two continuous copies of the data.

✓ *Scalability, Semantics, and Security*:

Are just a few of the issues that need to be discussed and resolved. For instance, the archive system needs to be scaled up to meet the dramatically rising demand for capacity, the archived information still needs to be recognized after 100 years or more, and the security and privacy of the archived information needs to be maintained. Only a list of the issues are presented here, leaving room for discussion.

VI. CONCLUSION

Long-term digital preservation is a crucial topic in the storage field, especially the preservation of cold data. Storage technologies, which are crucial to digital preservation, must overcome difficulties with regard to price, capacity, availability, performance, power consumption, and other factors. In this study, three commercially available archive technologies are examined and compared from the perspective of long-term digital preservation. This is followed by a discussion of ways to enhance optical media-based (in particular, BD-based) archive systems. The research draw the following conclusions from our comparison and debate.

- Due to the low cost per bit, extended lifespan, and low power consumption of the storage media, the optical disc-based archive system becomes a desirable option for long-term digital preservation. Due to the extremely low market share, the cost per bit of the BDXL now appears to be too costly to be able to afford. As a result of the cost per bit of BD-R (25 GB), the mainstream production, it also alludes to the enormous cut space in the future.
- The performance of the BD-based archive system needs to be substantially enhanced, especially the recording performance, in order to catch up with the HDD and tape-based archive systems. By enhancing the recording performance per drive, it will also simultaneously reduce the power consumption of the BD-based archiving system.
- The most appealing aspect of the BD-based archiving system is its lengthy lifespan. This lengthy lifespan, however, is simply a theoretical value derived from the lab. It still has to be tested in actual-world settings.
- To keep the optical-based archive system viable, high capacity optical discs, such as those with 1 to 10 TB/disc, must be produced. Fortunately, new optical storage methods with better storage capacity than the BD are still being researched, including holographic and multi-dimensional optical recording. The optical media-based archive system may be one of the finest options for long-term digital preservation, we conclude.

ACKNOWLEDGEMENT

This research work is support by Tertiary Education Trust Fund (tedfund) and part of the institutional based research (IBR) grand.

REFERENCES

[1]. Antoniazzi, L. (2021). Digital preservation and the sustainability of film heritage. *Information, Communication & Society*, *24*(11), 1658-1673.

[2]. Barrueco, J. M., & Termens, M. (2022). Digital preservation in institutional repositories: a systematic literature review. *Digital Library Perspectives*, *38*(2), 161-174.

[3]. Chatzieleftheriou, A., Stefanovici, I., Narayanan, D., Thomsen, B., & Rowstron, A. (2020, July). Could cloud storage be disrupted in the next decade?. In *Proceedings of the 12th USENIX Conference on Hot Topics in Storage and File Systems* (pp. 25-25).

[4]. Cherubini, G. (2022). Advanced control systems for data storage on magnetic tape: A long-lasting success story. *IEEE Control Syst. Mag.*, *42*(4), 8-11.

[5]. Daniel, R. (2019). Digital disruption in the music industry: The case of the compact disc. *Creative Industries Journal*, *12*(2), 159-166.

[6]. Dubbe, R., & Richards, D. (2019). The Impact of Memory Technology Advances on Microcomputers. In *Cutting Edge Technologies and Microcomputer Applications for Developing Countries* (pp. 41-60). Routledge.

[7]. El-Fakdi, A., & de la Rosa, J. L. (2022). Evaluating auction mechanisms for the preservation of cost-aware digital objects under constrained digital preservation budgets. *Mathematics*, *10*(1), 92.

ISSN No:-2456-2165

[8]. Furrer, S., Ebermann, P., Lantz, M. A., Rothuizen, H., Haeberle, W., Cherubini, G., ... & Doshita, H. (2021). 317 Gb/in 2 recording areal density on strontium ferrite tape. *IEEE Transactions on Magnetics*, *57*(7), 1-11.
[9]. Lantz, M., Furrer, S., Ebermann, P., Rothuizen, H., Haeberle, W., Cherubini, G., ... & Imaoka, N. (2020). 317 Gb/in2 recording areal density on strontium ferrite tape.
[10]. Larson, E. (2020). Big questions: Digital preservation of big data in government. *The American Archivist*, *83*(1), 5-20.
[11]. Masenya, T. M., & Ngulube, P. (2020). Factors that influence digital preservation sustainability in academic libraries in South Africa. *South African Journal of Libraries and Information Science*, *86*(1), 52-63. model. *International Journal of Advance Research, Ideas and Innovations in Technology*, *5*(6).
[12]. Moles, N. (2022). Preservation for diverse users: digital preservation and the “Designated Community” at the Ontario Jewish Archives. *Journal of Documentation*, *78*(3), 613-630.
[13]. Ocen, G. G., Mutua, S., Mugeni, G. B., Karume, S., & Matovu, D. (2019). Evaluating factors responsible for inconsistencies in mobile devices digital forensic evidence extraction process
[14]. Wan, S., Cao, Q., & Xie, C. (2014). Optical storage: an emerging option in long-term digital preservation. *Frontiers of Optoelectronics*, *7*, 486-492.
[15]. Shatskaya, M. V., Abramov, A. A., Fedorov, N. A., Kostenko, V. I., Likhachev, S. F., Seliverstov, S. I., & Sychev, D. A. (2020). Data processing center of RadioAstron space VLBI project. *Advances in Space Research*, *65*(2), 813-820.