DNA Data Storage

Rev. 1/08/2019

ENGM 2210: Technology Strategy

Section 1

Engineering Management Program

Vanderbilt University School of Engineering

December 9th, 2024

(Luke Golumbic, Ethan Kur, Carter Buell, Dilara Isik)

Executive Summary

Final Report Outline

Our DNA data storage solution, utilizing enzymatic synthesis, offers a high-capacity, long-term, and energy-efficient alternative for data management, specifically designed to meet the growing demands of the BFSI sector. By collaborating with IBM, a global leader in cloud infrastructure and technology, we combine the remarkable data density and durability of DNA with IBM's proven cloud and security capabilities. This synergy creates a scalable solution that addresses critical challenges in the BFSI industry, such as exponential data growth, secure archiving, and rising storage costs.

Traditional data storage methods struggle to keep pace with the increasing volume of data in the BFSI sector, often requiring substantial physical space and energy. Our DNA data storage system, however, offers an efficient, compact, and sustainable alternative that significantly reduces space and energy consumption, while ensuring data integrity for decades. IBM's cloud platform further enhances the system's security, scalability, and seamless integration across BFSI ecosystems, ensuring a robust solution to meet regulatory and disaster recovery needs.

The product's commercialization will progress through key milestones, including the successful development and pilot testing of a scalable DNA data storage system, integration with IBM's cloud infrastructure, and establishing strategic partnerships with major BFSI players. Six years after launch we estimate to obtain a gross margin of 32%. Initial adoption will focus on large BFSI enterprises, positioning our DNA storage solution as a next-generation technology aligned with the sector's evolving data management and sustainability goals. Once validated, the product will expand globally, driving a competitive advantage for early adopters and delivering an environmentally sustainable data storage solution.

ENGM 2210: Technology Strategy

Final Report Outline

Table of Contents

Table of Figures

1 Macroeconomic/Societal Environment

1.1 Crisis identification

Global data storage capacity is currently reaching a critical point. The IDC projects a 27% growth rate in necessary storage space by 2025, demonstrating an "urgent need for storage space within a few years.". Because of this, companies are bolstering their manufacturing of current data storage technology which is an inefficient and unsustainable practice. There is a clear need and market for data storage innovation which absolutely includes the field of DNA data storage.

Additionally, current data storage centers are highly inefficient in their energy usage at an estimated "50 Megawatts of power per center." This energy comes in many forms. For example, a center in Arizona uses over one million gallons of water a day for server cooling. Only 11% of power used in these facilities is used directly towards data storage while the remaining energy is spent on necessary auxiliary needs. This model is wasteful and expensive and demonstrates a need for change as expressed by the relatively high importance and advantage of Crisis Intensity (Figure 1).

1.2 Identification of innovation enablers or inhibitors (Fisher)

Societal trends such as the demand for data storage and digital transformation drive technological advancements. Traditions and customs, however, may hinder innovation if they promote resistance to change or favor conventional practices over novel approaches. Unspoken rules, such as norms in academic or corporate environments, may either foster creativity through collaboration and interdisciplinary work or inhibit progress through rigid hierarchies and a focus on short-term goals.

DNA data storage involves five main stages: encoding, writing, storing, reading, and decoding. The first part in the encoding step is to get the data or information you want and convert it to a binary file – 0s and 1s. Now, knowing the sequence of 0s and 1s, it is possible to convert them into DNA bases, via a mapping code that can divide the data into writable chunks, and to determine which sequence of As, Cs, Ts, and Gs will represent the binary data. In the end, this will generate a set of DNA sequences that should be synthesized to store that binary file into DNA.

In writing and storing stages, DNA can be synthesized chemically usually using the phosphoramidite route, which has been used for decades. It is necessary to store DNA in a way that it will remain available for years to millennia, either by keeping it in a very cold environment or by freeze drying it, maintaining the DNA in a chamber with an inert atmosphere. These processes will ensure that data will not be corrupted via DNA degradation. Storing particles or even silicon wafers are current technological alternatives.

In the reading stage, it is necessary to access the DNA that was carefully saved and to discover its base sequence. At present, the most common method to unveil that mystery is by sequencing the DNA using one of the available sequencing technologies [1].

Finally in decoding stage, as the DNA sequence is available, use the decoding algorithm associated with that particular encoding method to retrieve the binary file. During this step, the decoding not only maps As, Cs, Ts, and Gs into 0s and 1s, it also makes sure to arrange the strands of DNA in the correct order and checks for errors and corrects them.

Tech companies like Microsoft are pushing the boundaries by developing DNA drives capable of both writing and reading DNA data, and cloud storage providers are actively exploring how they can incorporate DNA data storage to meet the demands of future storage needs [2]. This innovation reflects the dynamic interplay of societal enablers, such as technological advancements and corporate investment, alongside potential inhibitors like the challenges of large-scale implementation and traditional resistance to disruptive technologies.

1.3 Timing

The world is not fully ready for DNA Data Storage. There are many logistical and societal progressions that need to occur before this product can be utilized by our society. Currently, synthesizing DNA using a chemical-based approach is both costly and unsustainable (Synbio Beta). Companies would require an expensive recycling system to recycle the excess solvent produced by chemical synthesis. For this product to become cost-effective and sustainable it would need to take an enzyme-based approach (Wynn Institute). Commercial enzymatic synthesis only became feasible a few years ago, so this approach is relatively new and is not as effective as DNA created by chemical synthesis yet (IEEE). Enzymatic synthesis will likely be the adopted approach for DNA Data Storage, since this approach is currently less effective than our current data storage units companies will likely be hesitant to adopt it. In a larger societal scale, many people may be nervous to adopt "alive" data storage centers. Considering the hesitation by many Americans in regards to the covid vaccinations as 36% of Americans believed that the risks outweighed the benefits(PEW Research); as the vaccines were MRNA many Americans may be skeptical of DNA Data Storage as they are unsure of how a biological system will function with our technology. Figure 1 supports this notion as its current score demonstrates that at least part of our current society would be hesitant to adopt DNA data storage as a common practice.

Another piece to consider is the sustainability factor of the DNA data storage units. According to Sachin Rawat from Synbio Beta, companies are currently using chemical-based approaches to synthesize DNA which are not sustainable. According to a poll done by the Energy Policy Institute at University of Chicago, 82% of Americans believe climate change is happening. Given this poll, many Americans would likely stray away from the current unsustainable chemical synthesis. Research needs to be done to further the development of the sustainable enzymatic synthesis as it would garner support from those Americans who believe in climate change.

1.4 Strategic implications

It's clear from our research that our product is still extremely far from production, but we believe that it is still possible to be pushed into production within the 5 to 10-year window that we need to follow. The vision for the product is clear and there are some very specific research goals that need to be met in order to make it a reality. This is shown in Figure 1 by the high relative importance score (Figure 1).

Currently the alternatives to our technology are good enough that our technology wouldn't need to be used, but as technology advanced, we rapidly increase the amount of data that needs to be stored to make it all work. In Figure 1 you can see that the crisis intensity relative importance score is a above a 4, which shows that the crisis becoming more dire is important to the future of our product (Figure 1). Within years it could be possible that our current methods of data storage will fall short, and we need a new technology like ours to keep up with the high demand for data storage.

It is clear from Figure 1 that we still have lots of areas that need to be improved and lots of things that need to happen to make our product a success. This is show by looking at Figure 1 and seeing that the relative importance of most of the factors is high while the relative advantage is low (Figure 1). Even though all these things need to happen, the probable importance of our product in the future makes it very possible that everything that needs to happen for our product to reach market, will happen.

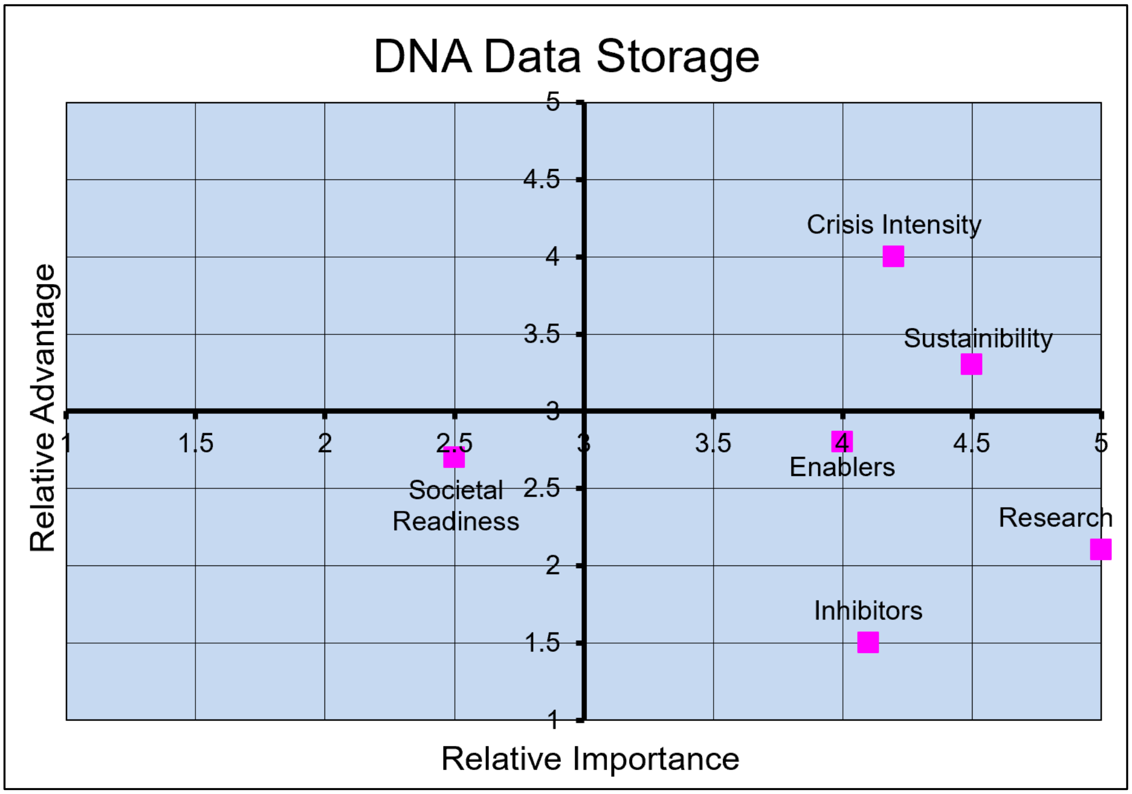


Figure 1. Importance-Advantage Matrix for Macro-Societal Innovation Success Factors.

2 Market/Demand Environment

2.1 Ideal market for your technology

Currently valued at 186.75 billion USD and projected to grow to 774 billion by 2032 (Fortune), the Data Storage market presents a well of opportunity that could be tapped with the innovative DNA Data Storage technology. This growth rate was bolstered by the COVID-19 pandemic which increased demand for remote Cloud storage. This technology would be sold in the greatest capacity to large tech companies whose long-term success relies heavily on storing customer data. The value of *cheap* data storage is exorbitant because, as mentioned in Section 1.1, the monetary and energy cost of maintaining conventional data storage centers consistently reaches new heights. For these reasons, this technology would be a mutually beneficial addition to the Data Storage Market as it improves the capacity and efficiency of a resource that is essential to the continued success of the tech sector.

There are several industries within the Data Storage Market that would benefit from DNA data storage technology, the largest of which being BFSI (Banking, Financial Services, and Insurance) which alone makes up about a quarter of the market share. Data analytics play an unprecedently large role in the financial sector, but there is a limit to its lucrative potential because of the plateauing data storage technology. (Experion) The immense capital within the BFSI industry presents a great market opportunity, but there are many other potential industries within the current data storage market including, but not limited to, IT/Telecom, Manufacturing, Retail, and Entertainment.

2.2 Candidate markets

Identify three markets for your technology that come closest to this ideal? Select the one you consider the most attractive (provide a rationale for your selection) and respond to the following questions, using one or more published market research studies.

According to Fortune Business Insights the three main markets for DNA data storage include BFSI(Banking, Financial Services, Insurance), IT and telecom, and healthcare and life sciences.

The first potential market is BFSI was valued at 60.5 billion US dollars in 2023 and is expected to grow to 131.4 billion US dollars by 2032 (IMARC). Banking, financial services, and insurance (BFSI) security involves using various services and solutions to prevent fraud and mitigate security breaches. This market is of high value to us as DNA data storage would address the needs for high-capacity and secure storage for those whom are processing vast amounts of data and capital.

The second potential market is IT and telecom. According to Statista in 2021 7.1 billion people were mobile users. Considering a majority of the planet uses mobile devices DNA data storage would offer compelling benefits. High data density and reduced physical space requirements, which can help manage the vast amounts of data generated daily. Its durability and compact nature provide an efficient solution for long-term data preservation. However, the technology faces challenges, including integration complexities to the already developed IT and telecom market. This can hinder its adoption in the telecom industry, where immediate data access and cost-efficiency are crucial for handling large-scale, real-time operations.

The third potential is the healthcare industry. According to Fortune Business Insights in this industry storage plays a crucial role and this market is expected to grow at the highest rate. DNA data storage would play a pivotal role in safely archiving and accessing patient records, medical images, and research data, facilitating effective healthcare services. This market is expected to have the highest growth rate, however, integrating DNA storage with existing healthcare IT systems and workflows can be complex and disruptive, requiring significant adjustments and potentially causing operational challenges.

Considering all the three markets BFSI is the ideal market for DNA data storage. This market is already the largest for data storage and is the easiest to scale. According to Fortune Business Insider this market is expected to maintain the largest share as the rise in demand for safe and secure storage mediums continues to grow.

2.3 *Market size*

According to Grand View Research by 2025 the BFSI securities market is valued at 74.3 billion US dollars with an annual growth rate of 13.1%. These numbers indicate a flourishing market with lots of potential for our product to grow.

DNA data storage will be narrowed into specifically the commercial banking and asset management portion of the market as this industry in need of a high-performance and secure form of storage and security to manage their large funds. According to Freedonia, the commercial banking market is forecasted to 1.1 trillion US dollars in 2026 representing annual growth of 5.8% from $865 billion in 2021. In 2021, U.S. commercial banks generated $865 billion in revenue, with an average annual growth of 2.8% from 2011 to 2021. Interest revenue fluctuated primarily due to shifts in interest rates and the total value of outstanding loans, both influenced by broader macroeconomic conditions. Noninterest revenue varied based on bank fee structures, regulatory factors, and the volume of saving and spending activity, which determines the base amount on which fees are applied. From 2011 to 2021, the sharpest revenue decline occurred in 2013, with a 1.4% drop, as many loans held by banks were originated or refinanced after 2008 at low interest rates. This trend continued into 2014, when revenues hit a decade low of $720 billion due to historically low interest rates following the recession. As interest rates began to rise in 2015, revenues increased, peaking in 2019. Banks also sought to boost revenues through other means, such as fees. In 2018, rising loan volumes and interest rates led to a decade-high 12% revenue growth. However, interest rates fell to pre-2015 levels during the COVID-19 pandemic, as the Federal Reserve enacted expansionary policies, resulting in reduced bank revenues in 2020 and 2021 (Freedonia).

The DNA data storage market was valued at USD 70 million in 2023 and is projected to grow at a compound annual growth rate (CAGR) of over 80% from 2024 to 2032(Global Market Insights). According to Tech Radar it currently costs about 1,000 US dollars per kilobyte of DNA storage. We would like to get this cost down by half. Through research and development we would like to target our launch for the product in the year 2032 as that maximizes the commercial banking market as well as the current DNA data storage market. We would need to reduce the cost of our product and target the commercial banking industry as the demand for data storage continues to rise.

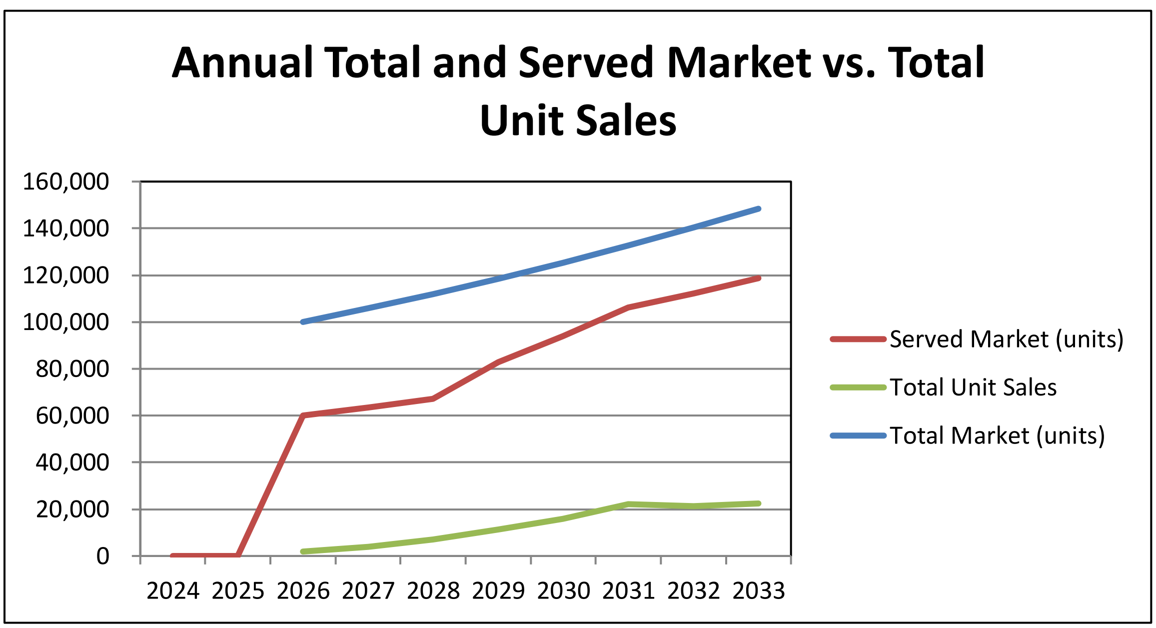


Figure 2. Annual Total and Served Market vs. Total Unit Sales

2.4 Market's Needs/Expectations from the Technology

In the BFSI (Banking, Financial Services, and Insurance) sector, data plays a central role in operations, decision-making, customer service, and risk management. Predictive analytics are crucial in this sector, helping forecast trends such as market movements, loan defaults, or insurance claims. As a result, storing historical data becomes vital. However, the data storage industry is facing capacity challenges. There are currently 11,000 data centers worldwide, and in 2023, 3,077.8 MW of data center space was under construction in primary markets. Additionally, the need for quick data storage and retrieval is paramount in the BFSI sector, where millions of transactions occur every second. Efficient processing of these transactions in real-time is essential to ensure smooth operations.

DNA data storage promises to deliver cheaper, higher-capacity, and longer-lasting solutions to meet the growing demands of big-data operations in both the government and private sectors. In 2012, 1 gram of DNA could store around 700 terabytes of data, but by 2017, advancements allowed 1 gram of DNA to store an astonishing 215,000 terabytes—equivalent to about 100,000 2TB hard drives. This represents a vast improvement in data density, requiring far less space and materials than conventional storage solutions. The short-term goal is to write one terabyte of data and read 10 terabytes within 24 hours at a cost of $1,000.

More BFSI companies are increasingly shifting to cloud-based solutions, and many are heavily investing in cloud infrastructure. For instance, JPMorgan Chase reportedly planned to spend $12 billion on technology in 2023, with a portion dedicated to cloud migration and data analytics. Despite the growing popularity of cloud solutions, DNA data storage stands out as an ideal option for long-term archival storage due to its compactness, durability, and low energy requirements—making it suitable for preserving large datasets over millennia.

Estimates show that large banks and financial institutions allocate 15-20% of their total IT budget to data management and analytics. In 2022, global spending on AI in the BFSI sector was around $14 billion, and this number is expected to grow substantially in the coming years. This further highlights the importance of exploring efficient and scalable storage solutions like DNA data storage to meet the sector's evolving needs.

2.5 Pricing – customer willingness to pay

When it comes to a new state of the art technology, it's essential to gauge the potential market price of this product by considering the current expenditures associated with constructing data centers using existing technologies. The current cost provides a tangible starting point, a baseline from which we can estimate the value that customers might place on the enhancements our technology promises.

According to a detailed analysis by the real estate services firm Jones Lang LaSalle, the investment required for establishing a state-of-the-art data center range between 1 billion and 1.5 billion (Hawkins). It's reasonable to infer that companies would be willing to invest even more than this for technology that offers significant improvements in performance and efficiency.

Our technology stands out by offering major advancements that could justify a higher price point. However, without concrete data on the exact value these benefits provide, we remain cautious and use the current market cost as a reference. It's important to note that the costs mentioned pertain to the construction of a single data center and not the total cost for a company. For large corporations, especially those in the BFSI industry, the cumulative cost of building and maintaining a network of data centers could reach hundreds of billions of dollars.

The lifecycle of data centers is such that they require regular updates and overhauls to stay up to date with technological advancements and operational demands. This ongoing need for modernization presents an opportunity for our technology to serve as a cost-effective and more efficient solution. Instead of pouring resources into new construction or extensive renovations, companies could integrate our technology.

2.6 Triggers and barriers to adopting this technology

One of the biggest challenges for DNA data storage is **compatibility**. BFSI companies primarily rely on cloud and physical data centers, which means they need new systems to manage DNA encoded data. Therefore, compatibility with existing infrastructure can be time and money consuming. The **trialability** of DNA storage will be gradually adopted, starting with non-critical, long-term data storage before expanding to more critical areas. This keeps the trialability at high performance (Figure 3). Additionally, even though DNA storage requires minimal maintenance, due to the resource-intensive need for initial setup, it is rated as low in **service-intensity** (Figure 3). It has a high **product life**, where data lasts for thousands of years, making it ideal for long-term archiving purposes, which is important for BFSI companies. DNA data storage is also extremely **energy efficient** and does not require power once data is stored, unlike traditional data storage such as cloud-based (as it is discussed in 1.1), reducing both costs and environmental impact. Finally, its compactness (**size limit**) makes it a cost-effective solution for large-scale data storage, which is a crucial factor for BFSI companies. While initial compatibility and service-intensity are the most important challenges to overcome, DNA's product life, energy efficiency and compactness offer high potential for the future data needs of the BFSI sector.

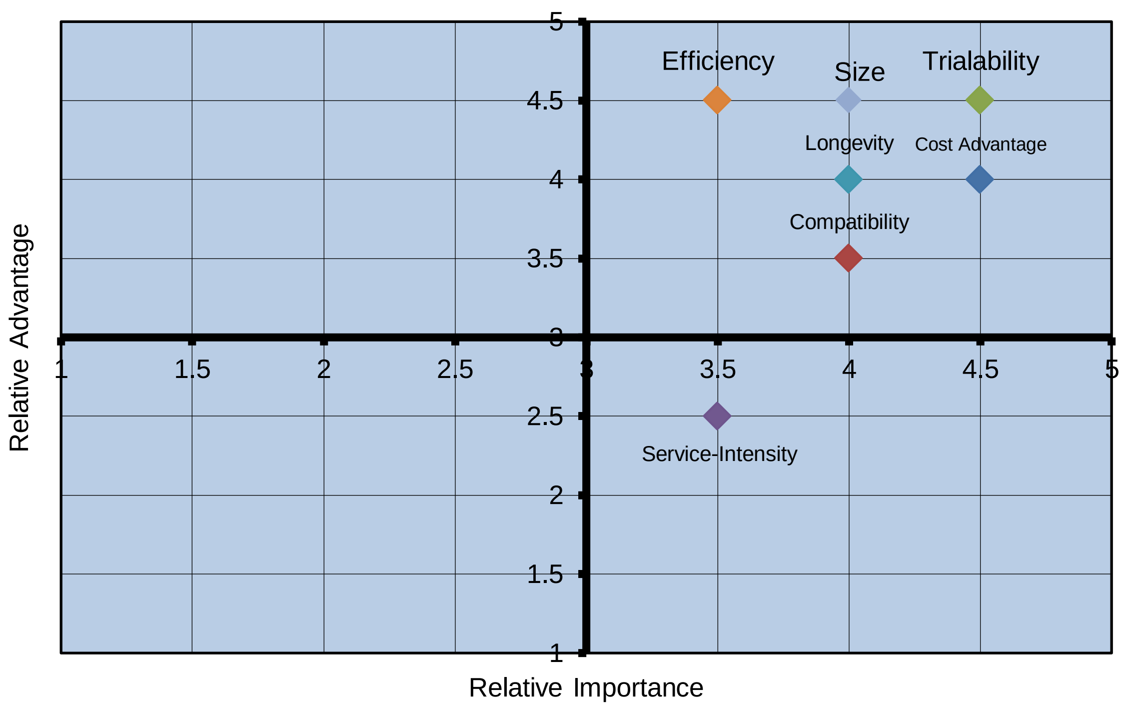


Figure 3. Market ISF's for DNA Data Storage

2.7 Strategic implications

The greatest market opportunity we have for DNA data storage lies within the rapidly increasing needs for high-capacity data storage in the BSFI industry. More and more data is being collected in this industry and our current methods of high capacity data storage will eventually not be enough. DNA data storage will be more sustainable, more space efficient and have the possibility for much higher data capacity.

Our biggest challenge lies in our lack of research. The current forms of data storage are currently enough to meet the needs of most industries but worries over sustainability and space usage will eventually lead to many industries wanting a better method of data storage. Movements to decrease the massive amount of space needed for traditional data storage centers are popping up in many places in the United States. In an article posted on the website *Data Center Knowledge,* we can see that people are protesting the large data centers being built in their cities. They have even given the movement a name, NIMBYism, which means 'not in my backyard' ism (Tozzi). These protests will likely lead people to demand industries to reconsider their data storage needs, leaving our technology as a viable alternative. This is supported in our market ISF graph by the high relative performance and relative importance scores for the 'size' ISF (Figure 3).

Looking at the ISF's for our market environment, our product fits the needs of the market extremely well. The relative importance and relative performance for many of the factors are similar, meaning we perform very well in the places that are the most important (Figure 3).

3 Technological Environment

3.1 History of the technology

The first historical method of data storage was used before “computers” were invented. The punch card was developed in 1725 by Basile Bouchon to control looms by giving them a series of instructions that were inscribed on the punch card in the form of holes (Foote). Over 100 years later, similar punch cards were used on the first computing machines. They were able to store not only instructions for the computer, but data that the computer could read and interpret. Punch cards were so successful that they were used widely all the way until the mid-1980s. They remained even longer in a few places, including voting machines.

Punch cards were gradually replaced by magnetic storage alternatives starting in the 1960’s. Magnetic tape was invented in 1928, by Fritz Pfleumer (Foote). This technology was used to make cassette tapes, which were used widely for personal computers.

The floppy disk was another type of magnetic storage, but it used magnetic disks rather than magnetic tape. The floppy disk was portable, easy to insert and remove, and inexpensive, but it was also easy to damage. The floppy disk got smaller over time, starting at 8 inches and eventually becoming 3.5 inches. The trend of storage becoming smaller in form but larger in the amount that it could store would continue until today.

Optical disks came about in the 1960s when an inventor named James T. Russel used light as a mechanism to record and play music (Foote). CDs (Compact Discs) were used to store sound while DVDs (Digital Video Recordings) were used to store video. These were, and still are, popular in the movie and music industry, but typically are not used for high-capacity data storage.

Today, many different types of data storage are still used and for different reasons. Hard disk drives are still widely used in nearly every type of computer. They have an extremely large storage capacity, are relatively cheap, and are decently fast, yet they do have the downside of having a limited number of times they can be written to and read from. This number is very large, but after many years of use, a hard drive may need to be replaced. Flash drives and SD cards are also still very popular and known for their extremely small size, decently large capacity, and ease of use. Both of these data storage devices use chips and transistors to store data and thus have no moving parts. They can be plugged into computers and removed very easily and can be written to and read from nearly limitless times. The main drawback of this type of storage is it’s limited capacity compared to something like a hard drive. Another popular data storage device is the solid-state drive or SSD. An SSD is a nonvolatile storage device that has a similar function to a hard drive but with some notable differences. They store data on interlinked flash memory chips. These are similar to the chips used in flash drives but are faster and reliable. Solid-state drives have high data capacity and very fast speeds compared to hard disk drives but are much more expensive.

In large industries such as the one our product would be released into, small forms of data storage such as flash drive and SD cards are irrelevant, and instead we need to focus on storing massive amounts of data. The need to store massive amounts of data has led to the creation of huge data systems originally called data silos, and eventually called data lakes (Foote). These data systems are what you would find in a modern-day data center. A data center is typically a massive warehouse full of data storage devices. Even though data centers have data storage needs higher than any single data storage drive could hold, hard disk drives and solid-state drives are still used in these data centers. The data centers hold massive systems that can hold thousands or even hundreds of thousands of drives. Hard disk drives are the type of drive used most often due to their high data capacities, but their longevity and speed is not as good as an SSD which are used in data centers, just not as frequently.

Our technology would be able to solve many of the issues that currently face the data systems that use hard disk drives and solid-state drives that we have today. DNA data storage would require much less space due to its extremely high data capacity. It would also be able to be written to and read from far more times than a hard disk drive would. Our technology would also likely mesh well with the current technology because of the way our current data storage systems work. Computers are typically not wired directly to storage drives in these massive data centers, and instead connect to a system that manages all the drives and relays with the computer. If we created our technology so that it worked with the existing systems, the computers would be able to connect just like they do now. Currently our product is far from development, but within 5 to 10 years, could become a very viable alternative to the current methods of data storage.

There hasn’t been a massive change in data storage technology in quite a while but instead, a gradual creep of improvement. If we look to the past, we can see that in the 1960’s to 2000’s, data storage changed so frequently, and yet technology was able to adapt. This gives good evidence to support the idea that people could adapt to our technology.

3.2 Physical architecture

The DNA data storage includes four main stages: write, store, retrieve, read. The first stage, writing, starts with encoding: a computer algorithm maps strings of bits, contains 0s and 1s, into DNA sequences. This amping help us to store massive amounts of information in compact and stable ways. The resulting DNA sequences are then synthesized, which generates many physical copies of each sequence. DNA sequences are arbitrary but finite length. This means that the long bit strings needed to be broken down into smaller as well as manageable chunks of DNA, which can be synthesized and stored. To enable reassembly, it is necessary to either include an index in each chunk or store overlapping chunks in different DNA sequences. It is proven that using simple index-based coding approach is the most optimal. For large amount of information, a large number of different DNA sequences needs to be synthesized, making it attractive to use array-based synthesis, which enables the synthesis of many unique sequences in parallel.

After the synthetization is done, the resulting DNA needs to be stored. Organick et al.3 estimate that a single physically isolated DNA pool can store on the order of 10^12 bytes. DNA pool is a small, physically isolated collection of synthesized DNA molecules that stores a substantial amount of data. To store large-scale data, many such pools are organized into a library, which allows DNA storage systems to scale up to enormous storage capacities.

In DNA data storage, when a specific data item is requested, the corresponding DNA pool containing the data must be physically retrieved and sampled. To avoid reading the entire pool, a system must support random access, which allows for the selection of a specific data item from the larger set. While random access is straightforward in conventional digital storage systems, it's more challenging in molecular storage due to the lack of physical organization among the data items in the same pool. In DNA storage, random access can be achieved through selective processes, such as magnetic bead extraction or PCR (polymerase chain reaction). Magnetic bead extraction works by using probes that bind to specific DNA sequences, allowing targeted extraction of the desired data item. PCR can also be employed, where unique primers (short DNA sequences) are associated with each data item during encoding, enabling the amplification of only the specific DNA sequence of interest. These selective methods provide a way to efficiently retrieve individual pieces of data from a vast DNA pool without the need to sequence all the stored data.

Once a DNA sample is selected from the pool, the next step is to sequence it and create a set of reads. These are DNA fragments that represent the molecules detected by the sequencing machine. These reads are then encoded back into the original digital data. The accuracy of this decoding process depends largely on two factors: sequencing coverage and the error rate encountered during the process. Sequencing coverage refers to the number of times a particular DNA sequence is read, and higher coverage increases the likelihood of accurately reproducing the original data by reducing errors or gaps. The error rate represents the frequency of errors introduced during sequencing; for example, incorrect nucleotide identification that can affect the accuracy of the data retrieval process. By balancing high sequencing coverage and minimizing errors, DNA storage systems can achieve a high probability of success in accurately recovering the original digital information.

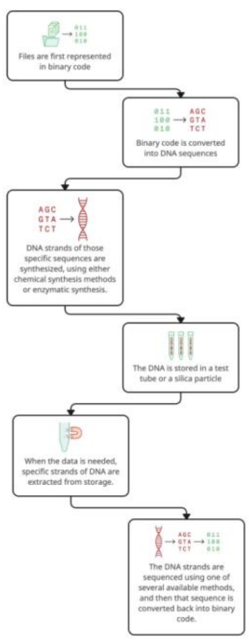


Figure 4. Hierarchical Block Diagram for the Physical Architecture of DNA Data Storage.

3.3 Technical Specification

There are several requirements for data storage technology in the BFSI market for it to be viable. These include data security, consistent availability/reliability, and cost efficiency. Any technology must meet a certain standard in these categories to be profitable in any market, and while DNA Data storage is not ready for industrial use yet, it is projected to exceed in all of them.

The integration of biology into modern technology has already led to the rise of Cyberbiosecurity, “a collective mixture of multiple disciplines such as cyber security, bio-security and cyber physical security…” (Arshad et al). Already, there have been hacking attacks targeting biologically based digital systems, but in contrast, there have also been some methods implemented to protect data specifically in the healthcare industry. These methods must be expanded upon and improved for BFSI companies to consider DNA data storage as privacy is among the most valuable commodity in the finance world, but the cyberbiosecurity technology is certainly heading in the right direction.

Additionally, DNA data storage must meet the industry standard in terms of easy accessibility and functionality. The potential for the technology to surpass the incumbent tech is clear, but it will not be profitable until companies know for a fact that it is a long term solution to the inefficiencies of hard drives and SSD’s. One method for ensuring that the technology lasts for a long time is by encapsulating the data in a silica matrix. However, when extracting data from the encapsulation, users run the risk of damaging the stored information, so as of now this strategy is only significantly viable for minimally accessed data. (Matange et al).

Finally, companies in the BFSI industry will prioritize, above all else, the cost benefit of implementing DNA Data Storage technology. As of 2022, Doricchi et al estimated that DNA data storage cost 800 million USD in stark contrast with the 15 USD per terabyte of traditional storage methods (ACS Nano). This is clearly not viable, but there is much optimism that this number will shrink drastically. The DNA Data Storage Alliance predicts that the cost will decrease from 800 million to just 1 USD per terabyte. If this prediction comes to fruition it is safe to say that business will, at the very least, strongly consider implementing innovative DNA technology into their data storage infrastructure.

3.4 Comparison with alternative technological approaches

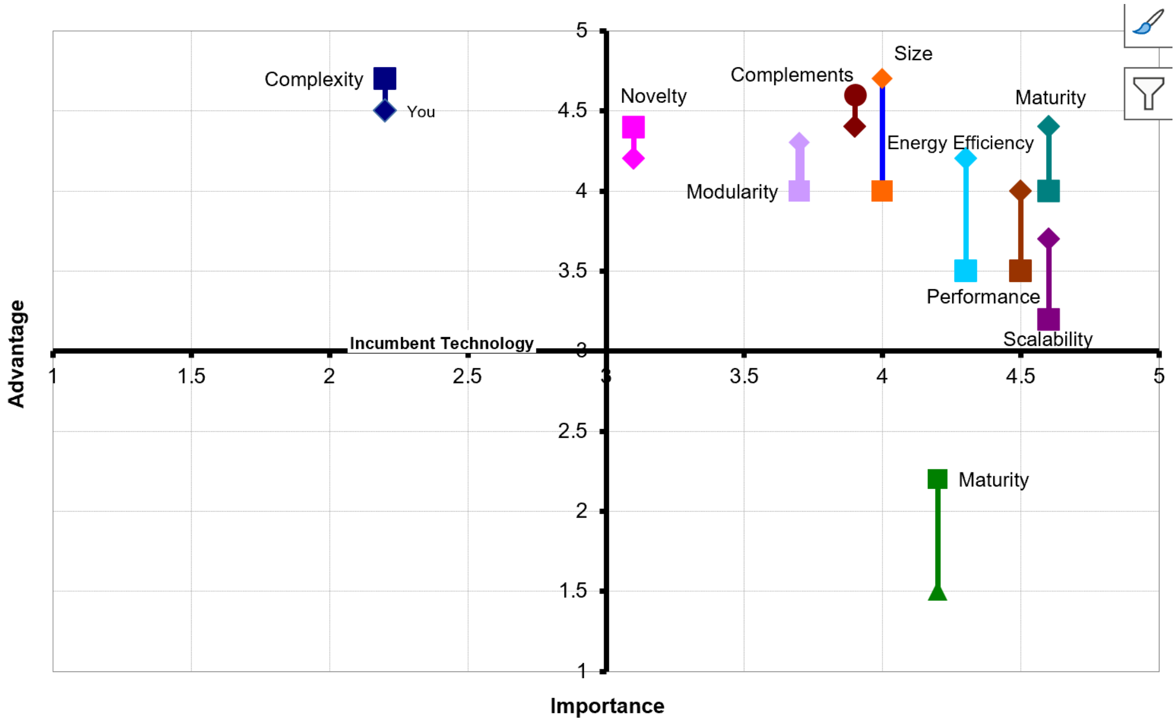


Figure 5. Importance-Advantage Matrix for DNA Storage vs Leading Alternative vs Incumbent Innovation Success Factors.

The incumbent technology alternative to our technology when it comes to high-capacity data storage is hard disk drives. Hard disk drives use magnetic disks to store data. They have very high capacity and are relatively inexpensive but have limited reusability due to the limited number of times they can be written to and read from. Roughly 90% of the data in a typical data center is currently stored on hard disk drives (Seagate).

Another alternative to our technology is protein-based data storage. Protein-based data storage is an approach to data storage where you record digital data within the sequences of amino acids in proteins. It promises high-density data storage but faces challenges in encoding, retrieval, and stability (Ng). These challenges are similar to the ones that affect our technology, making this a very similar alternative. The main benefits to our technology include higher data density, and that it is currently further along and has more progress towards the supporting technology needed to make it happen. This ties into our three most critical ISF's, which are energy efficiency, scalability, and maturity.

The Energy Efficiency ISF is very important to our technology because it relates directly to one of the crises that our technology aims to solve, a lack of sustainability that is leading to global warming. The current incumbent technology requires a lot more energy and is thus less sustainable. Protein based data storage and DNA based data storage would both be an improvement compared to this and thus would rate highly for this ISF. This gives us leverage over the incumbent technology and provides a valid reason for people to switch to our technology. This, however, does not give us a leg up on protein-based data storage, and for that, we must look at our other ISF's.

The Scalability ISF was chosen due to the other crisis that drives our technology, the ever-growing need for more and more data storage. Industries are needing to store more and more data, and the incumbent solution is taking up too much space. This makes it much less scalable than the newer solutions. Due to the ultra-high capacity of both DNA and protein-based data storage, these can be scaled larger, without taking up nearly as much space, and thus they both rank higher than the incumbent for this ISF. These first two ISF's explain why the newer solutions have a clear benefit to the incumbent as well as how they are going to help solve the various crises that surround data storge.

The Maturity ISF was chosen due to its relevance to both of the newer alternatives. This is one of the places where the incumbent technology has a very clear leg up on both the alternatives. Both DNA and protein-based data storage are still far from being ready to release to a market, while Hard Disk storage is a well-known technology that has matured to what is likely its full potential. Because of this, both alternatives will score lower on the ISF of Maturity, however, DNA data storage is further along than protein-based data storage, and thus ranks higher than it. This ISF shows us that our technology has a clear advantage over the alternative and will most likely be able to beat it to market. When combined with the previous two ISF's, we can show that our technology has an advantage over both the incumbent, as well as the other alternative.

3.5 Pipeline (follow-on) technologies

Currently DNA is synthesized using a chemical-based approach(Wynn Institute). This process is both extensive and expensive. For DNA data storage to be expanded on a large scale, the best approach will likely be to pursue enzymatic DNA synthesis. Since the demand for synthetic DNA is growing rapidly across various research and commercial sectors. Fields such as engineering biology, therapy, data storage, and nanotechnology stand to advance significantly if DNA can be produced at scale and low cost. However as of today, producing sequences longer than 200 base pairs remains prohibitively expensive. Our goal to further develop this enzymatic approach would be to accompany the current joint venture of Codexis and Molecular Assemblies which have begun to develop promising technologies in that realm (Nature Reviews Chemistry). Enzymatic approaches are a key factor of producing longer and more affordable strands. These approaches are most attractive in this regard and are also scalable, stereospecific and environmentally friendly. Enzymes can mediate mismatch recognition enabling the selective annealing of complementary strands, reduce the number of steps in each elongation cycle by eliminating the need for coupling reagents and decrease the dependence on organic solvents. Enzymes can promote synthesis with or without DNA templates, through amplification or in the synthesis of de novo sequences (Nature Review Chemistry). Enzymatic approaches are highly advantageous for DNA data storage due to their precision, scalability, and environmental benefits. Enzymes ensure accurate DNA sequence synthesis by selectively managing nucleotide additions, which is crucial for maintaining data fidelity. These methods are scalable, making them suitable for large-scale DNA synthesis required for extensive data storage, and they can lower costs by eliminating the need for coupling reagents and reducing reliance on organic solvents. Additionally, enzymatic processes are more eco-friendly compared to traditional chemical methods, aligning with sustainable practices. Currently this approach is still in the developmental stage, so further research and development will need to be done to scale the enzymatic synthesis.

3.6 Implications for your technology strategy

Although DNA data storage is a bright path for the current data storage industry it comes with its fair share of challenges. For example, the de novo synthesis of long DNA sequences remains difficult, necessitating the fragmentation of these sequences into smaller segments (approximately 200 bases). This approach requires generating a large number of unique DNA sequences. Data retrieval also poses challenges: although it is conceptually similar to the magnetic readout of a hard disk drive, DNA sequencing is needed to decode the information stored in individual oligonucleotides. Sequencing typically relies on fluorescence, which demands costly fluorophores, specialized optical equipment, skilled personnel, significant quantities of DNA, and extended reading times(ACS Nano). As of today DNA data storage is estimated to cost 800 million US dollars per terabyte compared to tape storage which costs 15 US dollars per terabyte. The solution to the current high-cost DNA synthesis is thought to be an enzyme-based approach. However, since this approach is still in its infancy, it is not readily scalable(Science Direct). As of now, Enzyme kinetics and the inherent characteristics of TdT offer a promising foundation for developing practical enzymatic DNA synthesis techniques (Science Direct).

4 Competitive Environment

4.1 Top competitors and basis of competition

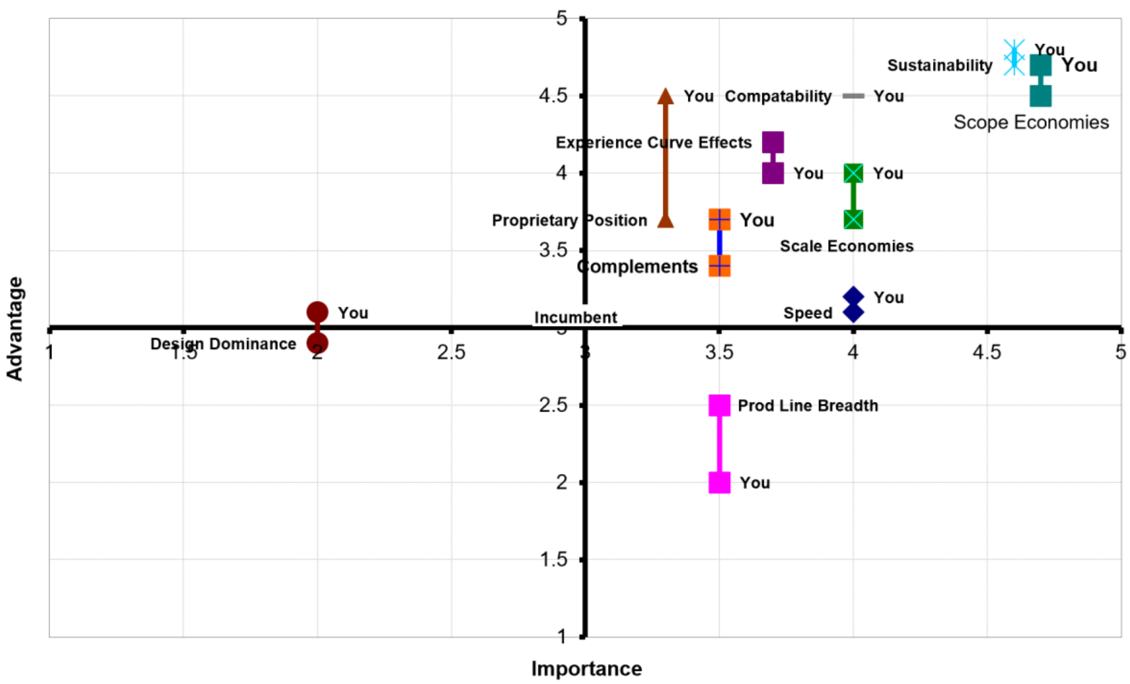


Figure 6. Top Competitors and basis of competition ISF chart

Our main incumbent is IBM. IBM is a global technology and consulting company. IBM is a leader in cloud computing**,** artificial intelligence (AI)**,** quantum computing**,** and enterprise solutions**.** Many banks depend on IBM Z® enterprise servers to manage their core transactional processes due to the high performance, security, and scalability these systems offer. These servers not only handle massive volumes of financial data but also seamlessly integrate with new banking applications that require real-time data for customer interactions, analytics, and fraud detection. Currently 92 out of the top 100 banks use IBM mainframes (Tozzi, Christopher). IBM is a well-known and well established company in both the commercial banking industry and the data storage industry as a whole. Convincing these financial institutions to transition from IBM to DNA data storage is a tall task, but can be done with a competitive price and simple integration. In 2023, IBM had a revenue of 61.9 billion dollars (Krishna, Arvind). Given this large some of revenue IBM has shown to be an industry leader for decades and is very compatible numerous uses for commercial banks. Our ISF score demonstrates our product is compatible, so we must demonstrate this compatibility to commercial banks and demonstrate the superiority and scalability of our product to encourage them to switch from the industry standard.

Catalog is emerging as a strong competitor to enzyme-based DNA data storage due to its innovative approach to using DNA for digital data storage and computation. Catalog has developed a platform that can store and process data using DNA molecules, which they claim can be written at high speeds and low costs, making it more scalable compared to traditional chemical synthesis methods. Currently Catalog has raised 50 million dollars and has less than 25 employees (ZoomInfo). They currently use a DNA synthesis approach instead of the enzymatic approach that we plan on using. Their company is currently more established than us and has had success encoding 17, 000 words on its DNA encoder (Mellor, Chris). Both of our solutions aim to make DNA data storage a viable option for the next generation of data storage, but Catalog’s current position in the market is further along than us meaning we would need to rapidly develop our product to demonstrate to the industry the capabilities of our product.

4.2 Competitive Landscape

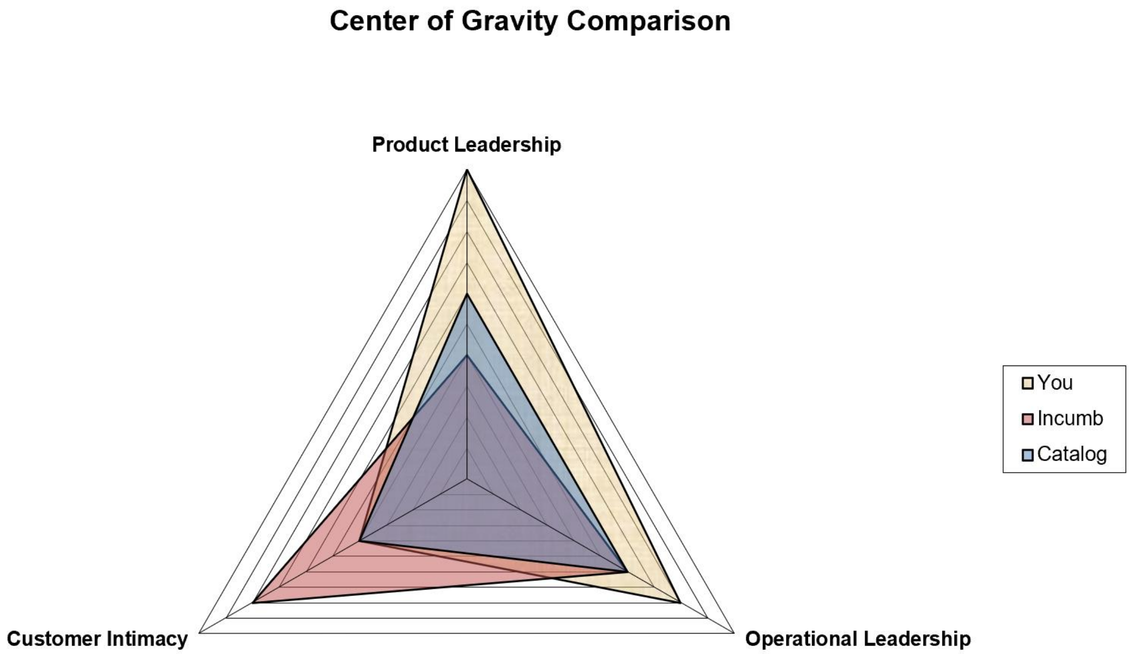


Figure 7. Center of Gravity Comparison

When situating DNA data storage within the broader competitive landscape, key factors to consider include product leadership, customer intimacy, operational leadership, and scope economies. DNA data storage has the potential to be a leader due to its exceptional data density, durability, and advantage in how data is stored. While current technologies like hard disk drives (HDDs) and solid-state drives (SSDs) are well-established, they are increasingly unable to keep up with the rapidly expanding data generation (IBM). DNA data storage offers a leap forward with the ability to store large amounts of data in an extremely compact format. Though the initial research and development costs are high, the long-term potential to store more data at a fraction of the space and infrastructure costs suggests that DNA data storage could become the preferred option for industries demanding efficient, large-scale storage solutions, like how cloud storage transformed the data market.

Scope economics is another area where DNA data storage is good at since it could excel across a variety of industries. For example, in the healthcare and life sciences sector, DNA data storage can offer a highly efficient solution for archiving large volumes of genetic, medical imaging, and patient records, ensuring long-term preservation without the risk of data degradation. In the entertainment and media industry, where there is a constant need to store high-resolution video and audio content, DNA data storage provides an innovative way to handle massive data volumes with minimal physical space requirements. Technology can also be attractive to sectors like IT and telecom, which require secure and scalable data solutions that can adapt to growing data needs (MIT News).

When it comes to operational leadership, DNA data storage currently faces hurdles, as conventional data storage technologies have the advantage of decades of mass production, mature supply chains, and cost efficiency. However, as DNA synthesis and sequencing technologies become more advanced and cheaper, DNA data storage is expected to offer significant cost advantages in the long run. With its ability to store large volumes of data in minimal physical space and without the need for constant maintenance or replacement, DNA data storage could achieve operational efficiencies that HDDs and SSDs can't match over time.

Another significant advantage lies in scale economies, where DNA data storage's high compatibility with existing digital data formats and its potential to be cheap at scale provide it with a competitive edge. As the technology matures, DNA data storage can be adapted to various industries and applications, ranging from archival storage to real-time data access, without requiring a complete overhaul of existing infrastructure. This adaptability, paired with the declining costs associated with DNA synthesis, means that DNA data storage can eventually become a highly cost-effective solution that integrates seamlessly into a wide range of data storage needs. Unlike other emerging technologies, which often face integration challenges, DNA data storage's compatibility with current data encoding and retrieval processes ensures it can be adopted without significant disruption. In essence, as DNA data storage technology continues to develop, it has the potential to offer a cheap, adaptable solution across multiple markets, allowing it to achieve economies of scope that few competitors can match.

In summary, while DNA data storage still has some distance to cover in operational leadership, its high compatibility, cost-effectiveness in terms of scope economies, and potential for product leadership make it a highly advantageous technology in the evolving data storage landscape.

4.3 Customer Value Proposition

For businesses in data-heavy industries like healthcare and IT that face rising costs and inefficiencies with traditional storage solutions, DNA data storage offers a transformative technology with increased data density and durability. Unlike hard disk drives and solid-state drives, which have physical limitations and data degradation, DNA data storage allows vast amounts of information to be stored in a compact, long-lasting format, ideal for industries needing reliable, long-term archiving. Moreover, DNA data storage integrates seamlessly with current digital infrastructures, requiring no significant overhauls. As DNA synthesis costs decrease, the technology becomes even more cost-effective, providing a future-proof solution that minimizes maintenance and physical space requirements, making it a superior choice compared to traditional storage methods.

4.4 Projected market share

To determine the expected market share for DNA data storage, we must first consider the total addressable market, which includes the entire BFSI market valued at $74.3 billion as of 2025 (Grand View Research). A smaller market within this market that is relevant to our technology is commercial banking. It is forecasted to grow to $1.1 trillion by 2026 (Freedonia).

Given that we are centering on commercial banking, our serviceable available market is a subset of the broader financial market of BFSI. For an emerging technology such as DNA data storage, capturing even a conservative percentage of the serviceable obtainable market within the first few years after launching would be a difficult target but we aim to slowly increase our share as the technology gains more market presence. Our company's position will be determined by our ability to leverage the competitive advantages of DNA data storage technology, notably in the areas of efficiency and sustainability.

We acknowledge that early market entry is typically accompanied by a minuscule market share due to the lack of known product reliability and brand presence. As evidence of positive reception and product efficiency emerge, we can expect some market share growth Beyond the initial years, as our technology matures and production costs diminish, our market share ought to experience more significant and sustained growth, with the possibility and hope that it may one day lead the market due to its clear benefits over the incumbent technology.

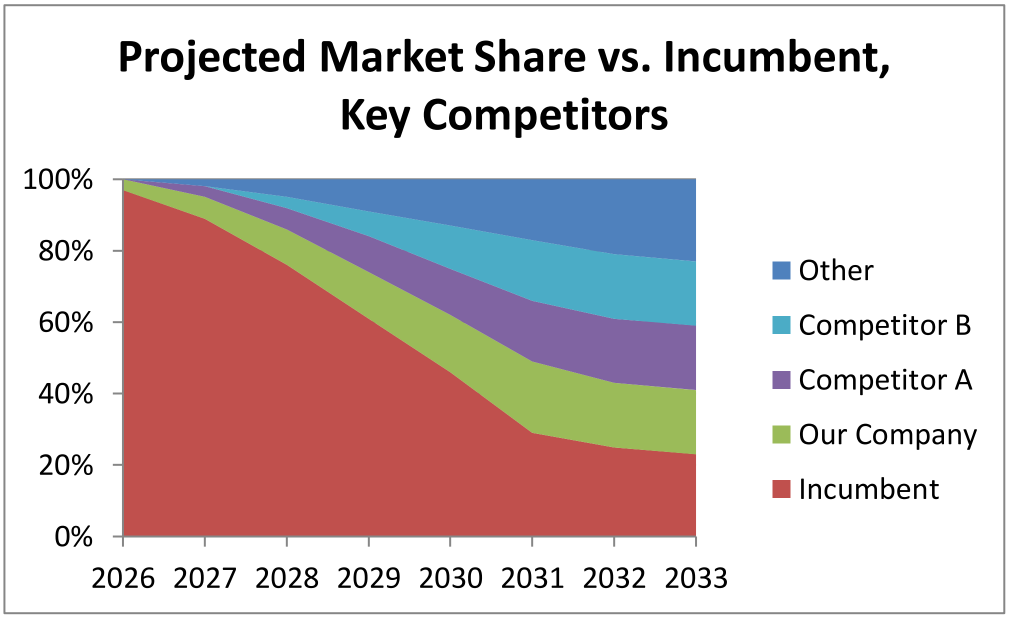


Figure 8. Projected Market Share vs. Incumbent Technology and Key Competitors.

5 Host Company Identification and Assessment

5.1 Capabilities required

Identifying a suitable host company for the DNA data storage technology requires a set of distinct capabilities that align with the innovation's nature and lifecycle. The company should possess a strong base in technology with the ability to foster cutting-edge research in bioinformatics and molecular biology. An established infrastructure capable of steering robust research and development efforts is essential given the technology's infancy and the need for innovation.

Moreover, the company must demonstrate the ability to scale operations efficiently. This includes the capacity for large-scale investments, expansions, and a track record of environmental stewardship, considering the sustainable advantage of DNA data storage. Our host company also needs to have a strong familiarity with the market and a strong customer base that will place its trust in a new technology.

The host company should have an innate culture of innovation, actively seeking to deliver breakthrough solutions and maintaining the agility to adapt to evolving customer needs and market demands. Strategic collaborations will be pivotal, enabling access to new research, shared knowledge, and resources. Crucially, the host should embrace risks associated with pioneering technologies while remaining resilient and responsive. Their mission must embody a visionary commitment to transforming the data storage landscape with solutions that are not just technologically advanced but also economically viable and environmentally sound. This may include a willingness to invest funds into the large amount of research that will be needed to take this product to market.

The ideal host will also be at the forefront of sustainable innovation, deeply attuned to customer needs, and dedicated to the long-term evolution of data storage, ensuring that the DNA data storage technology is nurtured to its full market potential.

5.2 Candidate host companies

Based on the criteria we looked at in 5.1, three companies that stand out in the technology and research sectors which could potentially host our DNA Data Storage technology are IBM, Illumina, and Microsoft.

Capability	Weight	IBM		Illumina		Microsoft	
		Score	Weighted Score	Score	Weighted Score	Score	Weighted Score
Scalability	7	9	63	7	49	8	56
Culture	4	7	28	9	36	7	28
Commitment to Sustainability	5	8	40	7	35	8	40
Market Presense	9	9	81	6	54	8	72
Willingness to accept risk	6	9	54	8	48	8	48
Technicogical Knowledege	8	8	64	8	64	9	72
Totals:			330		286		316

Figure 9. Host Company Comparison

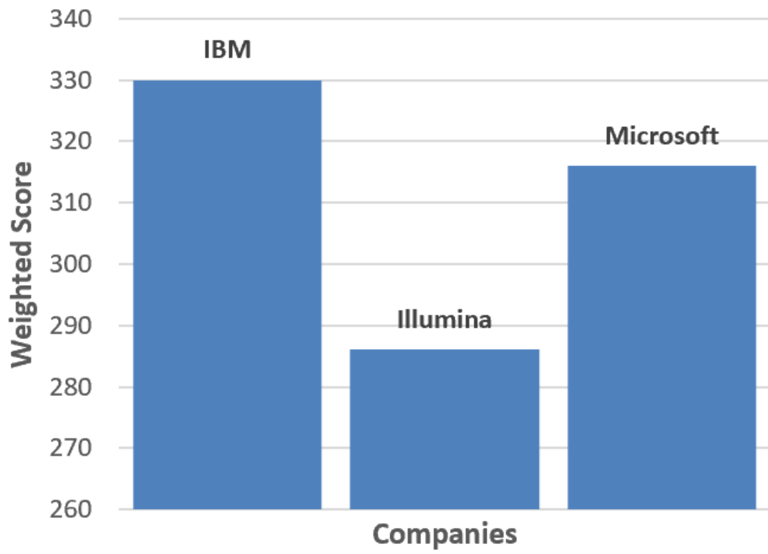


Figure 10. Host Company Comparison Weighted Score Bar Graph

IBM has a strong legacy in pioneering computing and data storage technologies. With substantial investment in research and development and a dedicated focus on innovation, particularly in areas such as quantum computing and artificial intelligence, IBM stands out as a potential host. Their cloud platform and business services are well-integrated into the BFSI sector, providing a synergistic environment for DNA data storage solutions (McDowell).

Illumina is a global leader in genomics and DNA sequencing, allowing them to leverage their technological abilities and comprehensive understanding of DNA science to advance the vital research that needs to be done to make DNA Data Storage a reality (Philippidis). The company's commitment to innovation and its existing infrastructure for high-throughput DNA synthesis and analysis positions it as a strong candidate for hosting our technology.

Microsoft is another good candidate due to their extensive cloud infrastructure, Microsoft Azure, and an already existing investment in research involving DNA data storage. Microsoft Research has already demonstrated interest in the field, making significant strides with its "Project Silica," aiming to store data in quartz glass, showing their commitment to alternative data storage solutions (Philippidis).

Based on the weighted scores given by our analysis in Figure 10, IBM is the correct choice for our host company. This makes a lot of sense due to their strong presence in our target market as well as their scalability and willingness to accept risk. IBM has always innovated with new data storage technology ever since the technology first came around. While they don't have current branches that are looking into data storage already, this just means that we will have to be their first partner to do so. This gives us a great position in the market due to the large amount of trust their customers have in them.

5.3 Company's business, key customers or markets.

Based on IBM's 10k filings IBM is a business that combines business insight with technological innovation, helping clients transform digitally through its hybrid cloud platform and AI technology. By leveraging expertise in software**,** consulting services, and infrastructure, IBM enhances operational agility and delivers impactful business results, backed by one of the world's leading research organizations (SEC). As stated in the 10k filings IBM is positioning itself for the next business opportunity, such as quantum computing (SEC). DNA data storage may not be a direct relationship to quantum computing, but it could serve a purpose to complement IBM's development of quantum computing. According to the Broad Institute of MIT and Harvard, DNA data storage could pose a solution to long term storage as DNA can store forever (Trafton, Anne). DNA data storage could also provide a solution for quantum computing's large data sets that need to be stored as it can store larges amounts of data forever. The current data IBM acquiring our company would allow it to have a strong foothold in the quantum computing and data storage industry. In 2023, IBM spent about $7 billion dollars on R&D and over $5 billion dollars to acquire nine companies (Krishna, Arvind). This demonstrates that IBM is a large company committed to innovation and willing to spend the money necessary to acquire our company and aid in the development of our technology.

5.4 Analysts' views of company's strengths, weaknesses, and future prospects

The first investment report analyzing IBM's prospects is Seeking Alpha which holds an overall positive tone for the future of IBM stock. This is surprising given the discussion of IBM's debt position of $30 billion which is described as "particularly significant…when IBM's market cap is approximately $175 billion" (De Oliveira). The report still considers that IBM's 14x Forward Free Cash Flows demonstrate so much potential as to overpower the debt issue. The report concludes that this puts IBM in a great position for the near future as the business model has succeeded beyond expectations.

The second report, by Simply Wall Street, also views IBM with a highly positive outlook with a single main downside. The report describes IBM as being sold significantly below fair value with positive price-to-earnings ratios but without considerably clear analyst forecasting. It also mentions the issue of high debt but argues that the increasing earnings and reliable dividends of the company are evidence enough to demonstrate strong future stock potential.

The third and final source, Zacks Equity Research, also holds an overall positive outlook for IBM, but with a bit more tentativeness than the other sources. Recently, IBM stock has outperformed Zacks S&P 500 composite, but this is not conclusive evidence that it will continue to do so in the near future. The report concurs with Seeking Alpha concerning the many surprisingly positive results as IBM "beat consensus EPS estimates in each of the trailing four quarters" (Zacks). The fear, however, lies in the valuation of IBM stock which, while far from struggling, does not align with the recent success of the stock and could potentially indicate a regression to the mean. All in all, IBM has demonstrated highly positive results with potential, if not certainty, to continue to perform.

IBM itself holds an optimistic, results-oriented outlook based on its positive showings in recent quarters. Revenue, profit margins, and cash flow all exceeded expectations through the second quarter of this fiscal year with the plurality of the success attributed to the Software segment of the company. Still, IBM recognizes the uncertainty of any forward-looking expectations for similar reasons as outlines by the three reports discussed above. Overall, the internal and external stock analyses are consistent with one another.

5.5 Company's technical strengths

IBM demonstrates strong competencies in customer intimacy and operational leadership, while maintaining a moderate standing in product leadership. The company excels in customer intimacy, possessing a profound understanding of client needs and tailoring its services to meet them effectively. Its adoption of the hybrid IBM Design Thinking approach places the user at the center of the product development process. This method demonstrates IBM's commitment to deeply understanding and addressing customer needs by involving users throughout the software development lifecycle. For instance, the use of "sponsor users" allows IBM to engage directly with the individuals who will be using their software, ensuring that the products are tailored to real-world requirements. By implementing practices such as ethnography, anthropology, and other user-research techniques, IBM gains profound insights into how customers interact with their products, resulting in solutions that are not only technically robust but also aligned with users' daily workflows. Additionally, the "playback" sessions ensure that user feedback is consistently integrated into product iterations, leading to continuous improvement based on actual customer experiences (Harvard Business Review). With over 3,196 companies using IBM's cloud services (IBM Cloud), the company's consistent engagement reflects IBM's dedication to building long-term relationships (6sense).

In terms of operational leadership, IBM leverages its core technologies, such as AI, cloud computing, and data analytics, efficiently delivering market-relevant solutions. This is evident in their success with hybrid cloud solutions (2023 IBM Annual Report). By offering solutions like process mining to understand workflows, digital workers to handle repetitive tasks, and Robotic Process Automation (RPA) to reduce errors, IBM enables streamlined, efficient processes. Additionally, smarter resource management, observability for application performance, and API management further enhance operational efficiency. This integration of automation tools demonstrates IBM's ability to implement scalable, adaptable workflows, solidifying its role in driving operational excellence (IBM). While IBM has a strong grasp of the technology behind its product line, it isn't the main driver of the company's competitive edge, leading to a moderate product leadership rating. With over 3,000 patents, IBM demonstrates innovation, but its focus tends to be on established technologies rather than consistently pursuing advancements, which explains the more average rating. To address this, IBM strategically enhances its offerings by leveraging partnerships and acquisitions. For instance, IBM's acquisition of Red Hat allowed the company to expand its cloud capabilities and integrate more innovative, open-source solutions into its product line. Overall, IBM's core competencies lie in its deep customer knowledge and operational leadership, enabling it to maintain a competitive edge in the consulting, IT services, and cloud-based solutions sectors.

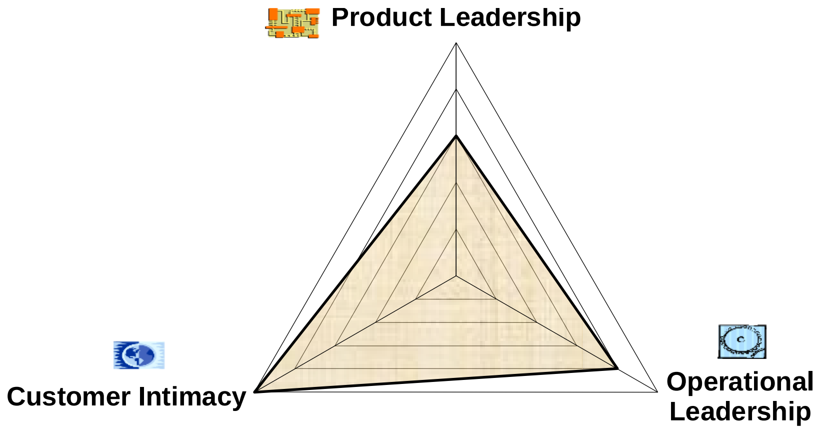


Figure 11. Center of Gravity for IBM.

5.6 Strategic implications

Based on the above analysis, IBM's key strengths lie in its deep understanding of customer needs, efficient operational capabilities, and strong market adaptability. These strengths suggest that IBM is well-positioned to continue delivering tailored solutions and maintain its competitive edge, particularly in cloud computing and intelligent automation. However, a challenge arises from IBM's moderate product leadership, as the company often relies on established technologies rather than driving innovation. This could limit IBM's ability to remain at the cutting edge of rapidly changing technology markets.

To address this, IBM could further invest in research and development or continue acquiring innovative companies, as it did with Red Hat, to enhance its product leadership, which is mentioned in 5.5. Additionally, IBM collaborates with different industries and invests in ecosystem partnerships to stay up-to-date with emerging technologies, thus ensuring it remains relevant and competitive even if product leadership isn't its primary strength. Although this weakness exists, IBM's strong customer relationships and operational efficiency can help offset it, allowing the company to integrate new technologies effectively and continue meeting market demands. By leveraging its existing strengths, IBM has the potential to navigate this challenge and strengthen its position in the technology sector.

6 Technology/Business Intelligence

6.1 Priority issues for intelligence collection

The current gaps in our information include scalable manufacturing, transitioning the data stored in the DNA to digital information, compatibility, efficiency of the enzymatic synthesis, and implementation in the commercial banking market.

Scalable manufacturing in DNA data storage faces a few obstacles primarily due to the complexity and cost of producing large quantities of synthetic DNA. While enzymatic synthesis offers an exciting alternative to traditional chemical methods, it still lacks the efficiency and automation needed to support high production. Additionally, ensuring accuracy and speed in synthesizing DNA strands at scale, while maintaining low error rates, remains a significant challenge for commercial viability. These limitations make it difficult to transition DNA data storage from experimental phases to widespread use in industries like commercial banking.

One of the main challenges in DNA data storage is efficiently retrieving and converting the information encoded in DNA back into a digital format. This process typically involves reading the sequence using DNA sequencing technologies, which can be slow and expensive. Additionally, the error rates during sequencing and data retrieval present issues in ensuring that the original data is accurately reconstructed. Advancements in sequencing technology and error correction methods are needed to make this transition more seamless and scalable for practical applications.

Compatibility is a significant issue for DNA data storage, particularly when integrating it with existing digital storage systems. There is currently no standardized method for converting digital data into DNA sequences and vice versa that is universally adopted across platforms. As industries like commercial banking rely on highly secure and standardized storage systems, ensuring that DNA data storage can be compatible with these systems while maintaining data integrity and security is a crucial challenge. Developing protocols and interfaces for smooth compatibility between DNA storage and digital systems is essential.

Although enzymatic synthesis offers an alternative to traditional chemical methods for DNA synthesis, its efficiency is still a major concern. Enzymatic methods need to achieve higher throughput, lower costs, and improved accuracy to be competitive in the DNA data storage realm. The current challenges include the time required to synthesize large volumes of DNA and maintaining high fidelity during the process. Further research into optimizing enzymatic synthesis, such as increasing speed without sacrificing accuracy, is necessary for its application in scalable DNA data storage.

Implementing DNA data storage in the commercial banking sector presents unique challenges. Banks require highly secure, reliable, and scalable data storage solutions, and while DNA storage offers immense potential in terms of density and durability, its practical implementation remains unanswered. Factors such as the speed of data retrieval, integration with existing financial systems, and ensuring compliance with industry regulations need to be addressed. Additionally, developing a business model that supports the transition to DNA storage within the banking industry is crucial for widespread adoption.

6.2 Interview questions

Scalable Manufacturing in DNA Data Storage (Lead Users in Biotech and Industry Experts):

- What are the main barriers to scaling the production of synthetic DNA for data storage at an industrial level?
- Is the current cost of manufacturing sustainable for large-scale implementation?
- How are advancements in enzymatic synthesis addressing these scalability issues, particularly in terms of reducing costs and improving automation?

Transitioning DNA Data to Digital Format (Experts in Data Storage)

- What are the main barriers to scaling the production of synthetic DNA for data storage at an industrial level?
- Is the current cost of manufacturing sustainable for large-scale implementation?
- How are advancements in enzymatic synthesis addressing these scalability issues, particularly in terms of reducing costs and improving automation?

Compatibility with Existing Digital Storage Systems (Potential Users):

- What obstacles are companies encountering when integrating DNA data storage with existing digital storage systems?
- Are there any successful implementations or ongoing standardization efforts that could improve compatibility?
- How important is compatibility with current storage systems for your business, and what are your concerns about transitioning to DNA storage?

Efficiency of Enzymatic Synthesis (Host Company):

- What specific advancements in enzymatic synthesis are being researched to make it a viable large-scale alternative to chemical DNA synthesis for data storage?
- How is your team improving throughput and reducing errors in the enzymatic synthesis process to make it more competitive for large-scale DNA data storage?

Implementation in the Commercial Banking Market (BFSI Industry):

- What are the key security, speed, and integration challenges for implementing DNA data storage in the commercial banking sector?
- How can DNA data storage be adapted to meet the scalability and compliance requirements of the banking industry?

6.3 Potential interview sources

Scalable Manufacturing in DNA Data Storage:

Dr. Nicholas Guise is a senior research scientist at Georgia Tech who studies the potential of DNA Data Storage technology. He has been working specifically on expanding the existing technology to a broader scale and is highly optimistic, saying DNA storage "has evolved over the ages as a very compact and reliable means of information storage" (gatech). If Dr. Guise is difficult to reach, there are many other experts involved in expanding the scope of DNA data storage researching at the Georgia Tech research institute, so it is a valuable source for interviewees.

Transitioning DNA Data to Digital Format:

Two of the leading experts on merging genetic and binary code are Dr. Albert Keung and Dr. Orlin Velev, both with the NC State Department of Chemical and Biomolecular Engineering. Along with researchers from Johns Hopkins, they have "demonstrated a technology capable of a suite of data storage and computing functions" (NC State). A couple of the head researchers from Johns Hopkins are BME professor Winston Timp and Post-doc researcher Paul Hook. This lab is on the cutting edge of this field of DNA Data Storage innovation, and it is worth reaching out to these professors or others involved with their work to learn more about transitioning data between genetic and digital formats

Compatibility with Existing Digital Storage Systems:

Muhammad Hassan Raza, Salil Desai, Shyam Aravamudhan, and Reza Zadegan published an article called "An Outlook on the current challenges and opportunities in DNA data storage" in 2023 which discuss the combined implementation of genetic-based data storage with silicon-based data storage (ScienceDirect). As accredited biotechnology experts, any of the four authors would offer valuable insight into this facet of DNA Data Storage technology.

Efficiency of Enzymatic Synthesis:

There are several companies in various industries that are beginning to view DNA Data Storage as a viable option considering the increased synthesis efficiency. Iridia CEO Murali Prahalad discussed the recent shift in public outlook from science fiction to reasonable possibility. It is for this reason that Iridia and similar companies such as Catalog and Twist have recently invested great sums of money into DNA Data Storage research. In addition to Prahalad, Twist CEO Emily Leproust and business development executive Steffen Hellmond should be contacted to discuss the reasoning behind the dramatic shift in perspective among the data storage industry (Tansey).

Implementation in the Commercial Banking Market:

Michael Wu, GM and President of Phison Technology Inc., discussed the vital importance of data storage to the financial industry as a whole. While he does not discuss DNA data storage explicitly, his clear views regarding data would still be valuable in terms of strategizing how to push the innovative tech into the BFSI industry. In his article for Fast Company, Wu describes a few of the reasons for data storage's importance to the industry including enhancing the customer experience, improving security/risk management, and increasing overall efficiency. Although his technical expertise is not directly aligned with genetically encoded data, Wu and other experts within Phison Tech should be contacted in order to discuss BFSI data storage as a general matter.

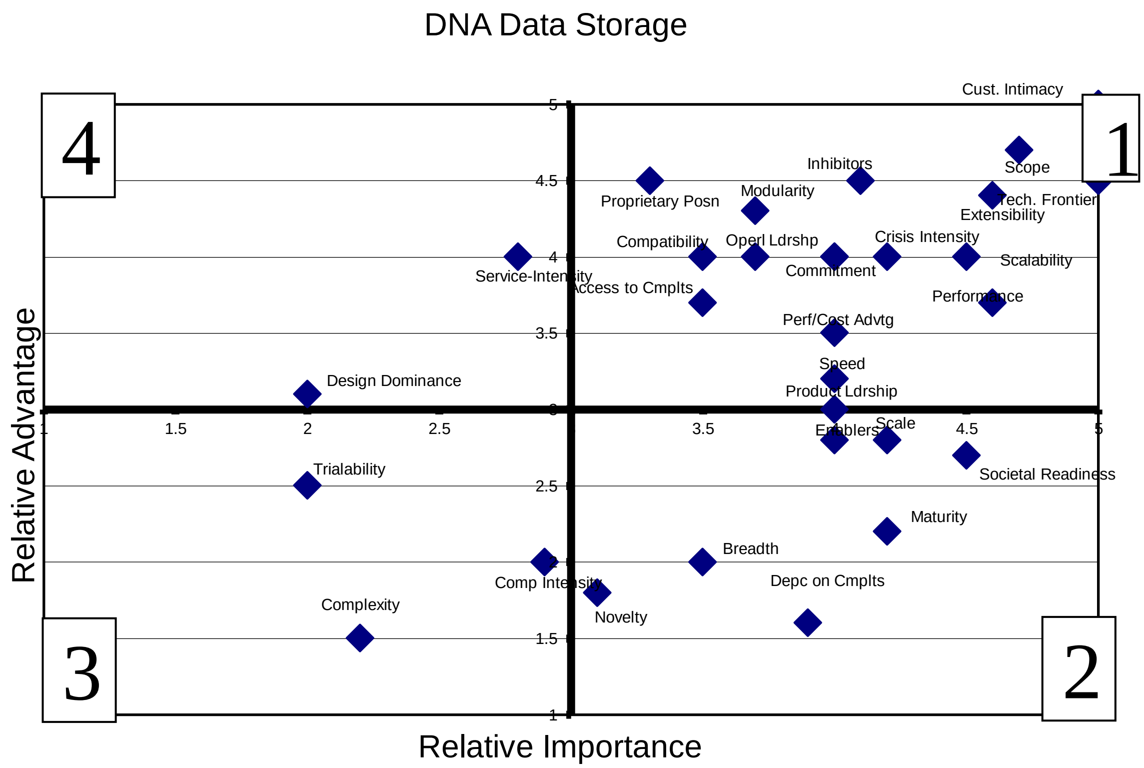


Figure 12. ISF Matrix for DNA Data Storage Portraying the Strengths, Weaknesses, Opportunities, and Threats.

6.4 Interview summaries

Full interview summaries can be seen in Section 14.2.

Summary of Strengths, Weaknesses Opportunities, and Threats (SWOTs Analysis)

6.5 Summary of Strengths, Weaknesses Opportunities, and Threats

This table (Figure 12) provides a visual representation of the factors that will influence the development and commercialization of DNA Data Storage. Critical success factors include cost/performance advantages as a strength, dependence on specialized complementary assets as a weakness, technological complexity as an opportunity, and service intensity as a threat.

DNA data storage offers a compelling cost/performance advantage due to its longevity, minimal energy requirements, and low maintenance capabilities. This positions it as a highly cost-effective solution compared to traditional storage methods such as cloud storage, which requires constant energy for operation, cooling, and regular hardware updates (Data Center Knowledge). The ability to store data for thousands of years without infrastructure maintenance provides a strong performance advantage, especially as global demand for sustainable, high-capacity storage increases (SEAGATE). Despite these strengths, initial setup and integration pose significant challenges, especially in industries such as Banking, Financial Services, and Insurance (BFSI), where seamless compatibility with existing IT infrastructures is important (IMARC). DNA data storage's dependence on specialized complementary assets such as advanced sequencing and enzymatic synthesis technologies also adds complexity. However, promising developments in these areas will improve the accessibility and scalability of the technology (ScienceDirect). Because the initial integration process is resource-intensive, service intensity is low until the system is operational, which is a barrier to adoption. However, as supporting technologies advance and storage solutions become easier to implement, DNA storage could revolutionize long-term, sustainable data storage, making it a very attractive option for industries looking to reduce operational costs and environmental impact. Despite competition, DNA storage offers a strategically positioned, future-proof alternative to meet the evolving needs of large-scale data storage.

7 Intellectual Property Strategy

7.1 Patent strategy

Even though the application for a patent is a long, tedious process, the DNA data storage market is becoming increasingly more competitive, deploying a patent would give us a strategic advantage over the rest of the market.

Currently, many companies are synthesizing DNA using chemical-synthesis. Our product differs in that instead of chemical synthesis we use enzymatic synthesis. This process is not used by many companies and is even in a premature stage at the companies who use it. Our goal would be to incorporate a novel aspect into our enzymatic synthesis to further differentiate our product, making it more easily patentable. Current efforts in enzymatic synthesis have shown promise but have faced challenges with efficiency (Carlson, 2024). By investing in research to improve enzyme-based synthesis, we can introduce a more cost-effective and sustainable solution to the market, positioning our product as both unique and highly competitive. We will also pursue a technique for enzymatic synthesis and digital coding that is based on template-independent polymerase TdT and nanopore reading. This method allows information to be stored in DNA without the need for single-base precision, reducing costs through miniaturization and enzyme recycling. Furthermore, the synthesis of 1000-nucleotide-long strands with homopolymeric sequences has significantly shortened synthesis times (Doricchi et al., 2022). If we are able to implement this technique and further research our solution, this could bring a unique novel aspect to our product positioning it for a patent.

Our approach to enzymatic DNA synthesis is non-obvious as it goes beyond merely applying existing technologies. While template-independent polymerase TdT and nanopore reading have been explored, our innovation lies in combining these with proprietary techniques for enzyme recycling and miniaturization, optimizing both cost and efficiency. The use of homopolymeric sequences to accelerate synthesis is a key differentiator, but our specific method for integrating enzyme-based synthesis with digital coding processes makes this solution non-obvious to those skilled in the field.

Our enzymatic DNA synthesis technology offers significant utility by providing a more efficient, cost-effective, and sustainable solution for DNA data storage. Unlike traditional chemical synthesis methods, our approach reduces production costs through enzyme recycling and miniaturization, while also accelerating the synthesis process. This makes DNA data storage more accessible for large-scale applications, such as long-term archival storage and data-intensive fields like commercial banking in which our host company already has a strong footprint in.

7.2 Other strategies for building your proprietary position

Due to the complex nature of our technology, if we are able to patent that technology, it will be very hard for other companies to replicated it and not break the terms of the patent. This increases our proprietary position because it will be difficult and uneconomical to imitate our technology.

Some of the other strategies that are often used to increase proprietary position don't work well for our technology. It would be very difficult for us to create a monopoly like the one Amazon has on one click online shopping, because we would have to scale up our production to cover the needs of every industry rather than just a few target industries that fit the use of our product better. This could eventually become a viable strategy, but for now, we will put it on the back burner and focus on other things.

The strategy of acquiring other intellectual property to enhance our overall competitiveness could also in theory work but may not be our best course of action. The money that we receive to fund our business would be much better spent on funding our research and creating a better product. Any similar technology that we would consider acquiring is most likely very important to other high density data storage, and this would only make sense to acquire if our technology was looking to replace every single other mode of data storage, rather than being used when a very high amount of storage is needed. We should, however, be looking at new innovations that could help our technology come to market. For example, if a company makes a technology that would assist with enzyme synthesis, we will want to try and acquire that technology as soon as possible, as it would greatly increase the viability of our own product.

One strategy that could be very promising to increase our proprietary position is the sale of patents and licensing our technology for other purposes. The research we are doing into DNA synthesis has many uses outside of DNA Data Storage. This means that we could license our technology to other companies for use in a variety of other fields. This would allow us to capitalize on markets that have nothing to do with data storage and will cost very little because we are already doing the research for use in our own product. It would be very possible to license our DNA synthesis for use in the medical field as this technology has major implications in medicine. In an article published by Vox news, it is said that mastering DNA synthesis could lead us to new vaccines that could help stop future pandemics and fight deadly pathogens (Piper). This shows that there are current and relevant uses for our technology in the medial field that we could capitalize on.

8 Product Strategy

8.1 Product-market scope

To expand the product-market scope, DNA data storage should initially focus on product development, creating a suite of complementary tools to facilitate the encoding and decoding processes required for data retrieval. These may include specialized data management software for seamless interaction with the DNA storage infrastructure, and modular sequencing tools that integrate with common enterprise storage solutions, making the system intuitive and understandable for users in industries such as Banking, Financial Services and Insurance (BFSI). Such complementary products are crucial to "complete" the DNA data storage solution, ensuring efficient and reliable data access while minimizing specialized knowledge requirements for the end user. This product development approach strengthens the core functionality of the technology while reducing the complexity barrier, making it easier for companies to adopt DNA storage for highly secure data with minimal training.

Market development should focus on adjacent industries with critical long-term data storage needs, such as government archives, healthcare, and scientific research organizations. These industries have similar requirements for data life, minimal maintenance, and security that align with the benefits offered by DNA data storage. These adjacent markets will require only modest adaptations to the core product, such as specialized compliance features or integration capabilities with existing data security protocols. Establishing a strong presence in these markets will allow DNA storage to gain reliability, leverage cross-industry applications, and build resilience, which could attract more industries as the technology matures.

Although diversification poses higher risks, it could be explored once DNA storage has proven successful in core markets. The most promising diversification path could involve scaling into consumer storage for privacy-conscious users looking for sustainable alternatives to cloud storage. This stepwise expansion approach—focusing primarily on product and market development in high-risk storage sectors—creates a solid foundation and allows for future diversification into consumer markets as DNA storage becomes more refined and manufacturing costs decrease. By following a sequential, expansion strategy, DNA storage technology can increasingly establish a broader market footprint, maximizing long-term adoption potential while reducing risks.

8.2 Product family

One way to expand the DNA data storage product family is by developing specialized add-on components for enhanced compatibility and integration with existing data infrastructures in the BFSI (Banking, Financial Services, and Insurance) sector. As DNA data storage technology offers unmatched density and longevity for secure, long-term archival storage, add-ons that facilitate seamless interfacing with existing digital storage platforms and cloud services would add significant value. An example of such a component could be an adapter that allows users to convert and transfer digital data to DNA sequences directly from their current systems, streamlining the encoding process. Given the BFSI sector's high security and compliance needs, another valuable enhancement would be a built-in encryption module to ensure that data integrity and confidentiality are maintained across storage and retrieval processes. By addressing industry-specific needs through complementary components, the DNA data storage product line would meet critical requirements and stand apart from traditional data storage options, ultimately maximizing adoption and customer value.

To focus on market development, DNA data storage products could be adapted for different applications and market scales. The core DNA storage unit could be optimized for high-priority, secure data archiving in large financial institutions, allowing these organizations to reliably store sensitive, historical data in a compact and sustainable format. In contrast, a more compact DNA storage unit could be designed for smaller-scale operations in healthcare and legal industries, where data security and long-term storage are equally critical, but data volumes may be lower. This smaller unit would offer organizations a highly secure, scalable option for archiving confidential data while reducing their physical storage footprint. Additionally, partnerships with data service providers, such as IBM could help bring DNA data storage to broader enterprise audiences, leveraging these providers' established networks to ensure smooth integration across industries and enhance the product’s credibility and reach.

8.3 Product specifications

To ensure DNA data storage technology is effective within the BFSI market, there are several critical technical specifications that must be met in our initial minimum viable product. There are also some specifications that can be implemented later in the product's developmental life cycle to further improve the product. These specifications will ensure that DNA data storage maintains a high level of performance and ensure a successful adoption of this technology within the BFSI sector.

1. The DNA data storage system shall provide data security and privacy protection that meets or exceeds industry standards for the BFSI sector, which prioritizes the privacy of customer information.
2. The DNA data storage system shall have a consistent availability and reliability to ensure uninterrupted access to critical data required for business operations.
3. The DNA data storage system shall be compatible with existing BFSI infrastructure and data management workflows, to facilitate easy integration and adoption within the industry.
4. The DNA data storage system should reduce the total cost of ownership for data storage compared to traditional storage solutions.
5. The DNA data storage system should allow for data retrieval speeds within a timeframe that exceeds the BFSI sector's need for real-time data processing and analytics.
6. The DNA data storage system should demonstrate a level of scalability that can accommodate the projected growth in data demands of the BFSI industry while also allowing for less overall space requirements.

8.4 Potential lead users and minimum viable product

A potential lead user in the BFSI sector could be any of the large financial institutions that are heavily investing in cloud infrastructure and data management. These organizations are likely facing intense pressure to manage growing data demands, while also prioritizing data security and cost efficiency. More specifically, it would be advantageous for us to choose an organization that is growing at a fast pace and may have an urgent need for our technology. For our DNA data storage technology, an initial minimum viable product could be a small-scale pilot deployment to store and retrieve a subset of the organization's less frequently accessed data. This could serve to demonstrate the technology's capabilities in terms of data density, security, and cost effectiveness.

A good option for our lead user could be the company Robinhood. Robinhood provides online stock brokerage solutions that allows users to invest in publicly traded companies and exchange traded funds worldwide. Robinhood's business model relies on rapidly processing and analyzing large volumes of financial data, making DNA data storage an attractive solution to manage its growing data needs (Satia). Additionally, as a forward-thinking technology company with a tech-savvy user base, Robinhood has demonstrated a willingness to adopt cutting-edge innovations. Robinhood's position as a relatively young and agile player in the BFSI industry means it likely has fewer legacy systems, making it more likely to integrate new data storage technologies

The key pieces of information to capture from the lead user would include the lead user's specific data storage requirements, including the volume, access patterns, and security needs for the subset of data they would like to store using the DNA data storage technology. We would also benefit from feedback on the initial product's performance, ease of integration, and overall user experience as well as insights into the lead user's decision-making process for adopting new data storage solutions, including any specific technical, operational, or financial criteria that they would need met. We can us these suggestions for future product enhancement and to create additional features that would make the DNA data storage technology more appealing and valuable to the lead user.

To capture this information, I would conduct interviews with key stakeholders at the lead user organization, such as the CIO (Chief Information Officer), CTO (Chief Technology Officer), or whoever else may be in charge of the companies' data management. We could also organize demo sessions to allow the lead user to actively engage with the initial product and provide real-time feedback. Additionally, we could implement feedback collection, such as a survey, to continuously gather insights and monitor the lead user's experience over the course of the products deployment.

8.5 Product-market penetration path

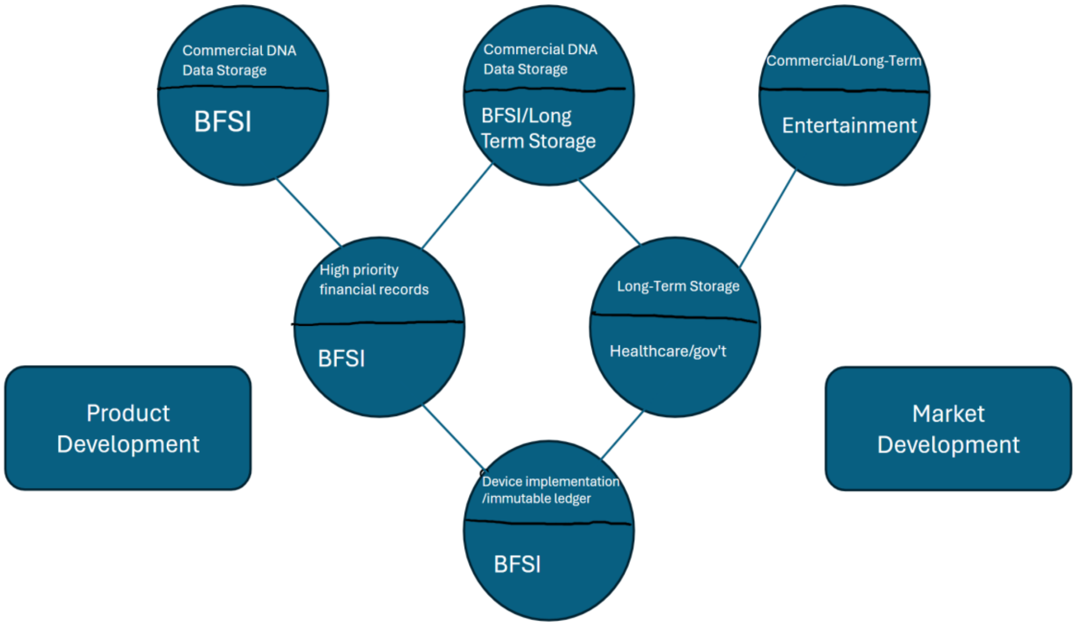


Figure 13. Application and Market Path Diagram for DNA Data Storage

After securing a spot for DNA Data Storage in the BFSI industry, there are multiple potential paths for it to pursue if the product or market is developed (Figure 13). The bowling pin diagram demonstrates that the next step in product development would be to use DNA data storage for secure backup of high-priority financial records. This application would allow financial institutions to use our product as a solution for disaster recovery and data integrity. A major advancement in DNA Data Storage would be device implementations and the use as an immutable ledger. Implementing DNA data storage into a device allows businesses to have local storage solutions that maintain data with integrity and for thousands of years. The other application would be as an immutable ledger for critical financial transactions and client interactions, aiding in fraud detection and regulatory compliance audits. These types of products have a lot of potential in the BFSI industry as they rely upon trustworthy and secure data storage solutions. If the market is expanded upon, the healthcare and government industries would be potential suitors for our DNA data storage technologies as they rely upon effective and long-lasting data storage technologies. Another broad industry to consider is within entertainment. Media and entertainment companies need scalable, cost-effective storage for massive archives, with minimal alteration to data formats. DNA storage's longevity and scalability are strong selling points for preserving valuable media assets, making this a logical next step without radically changing the product.

9 Operational Strategy

9.1 Operational Architecture of the Business

Suppliers	Inputs	Process	Outputs	Customers	Requirements
Suppliers of Raw Materials	Enzymes	Receive and process raw materials	DNA data storage unit	Large corporations	High reliability and data stability
Investors	Capital Biotech companies	Fund R&D and infrastructure development	Long-term storage solution	Governments, research institutions	High capacity and cost-effectiveness
Equipment Manufacturers	DNA sequencing and synthesis equipment	Assemble DNA storage units	DNA encoding and decoding service	Finance sectors	Affordable storage and retrieval
Labor Pool	Skilled bioinformatics and data management professionals	Develop encoding protocols, train personnel	Maintenance and support services	Data centers, BFSI	Accessible maintenance support
DNA Data Storage Researchers	Technology IP	Conduct R&D for data encoding/decoding	Scalable DNA storage solutions	Cloud providers, enterprise customers	Scalable storage capacity
Facility Supplier	Manufacturing and distribution facilities	Provide storage and maintenance infrastructure	Installed DNA data storage systems	BFSI	Long-term data integrity and security

Figure 14. High-level SIPOC Diagram

The high-level SIPOC diagram for DNA Data Storage is shown in Figure 14. The high-level SIPOC diagram for the DNA data storage business outlines the key components necessary for its operational architecture. The suppliers for DNA Data Storage include raw material suppliers who provide enzymes. Equipment manufacturers supply DNA sequencing and synthesis equipment. The bioinformatics R&D team for enzyme synthesis is responsible for developing property encoding algorithms and software which is the next and one of the most crucial steps after enzyme synthesis. The other suppliers are a customer support and training team and a maintenance and technical support team. The inputs required for business include synthesized DNA, enzymes, and sequencing kits as the foundational materials. DNA sequencing and synthesis equipment, proprietary encoding algorithms and software for optimized data storage, training materials and user guides, and tools and replacement parts are necessary inputs for maintenance. The main outputs of the business are the fully assembled and tested DNA storage units ready for data encoding, optimized encoding, optimize encoding and decoding protocols, and knowledgeable users who have been trained to operate and maintain the system. Additional outputs include restored and maintained DNA storage units, ensuring continuous functionality.

Within the chosen market, which includes industries with high-capacity and long-term data storage needs, the customers are data centers, government agencies, big-tech companies, and BFSI that prioritize data security, longevity, and reliability. These customers have specific requirements for the DNA data storage system, such as high data reliability, secure long-term storage, accessible maintenance support, and cost-effective operation. Reliability is crucial as these industries cannot afford data loss or downtime, and the system must provide consistent functionality over extended periods, especially BFSI sector, which is our target audience. Accessible maintenance services are essential to quickly address any issues that may arise, preventing interruptions in data storage. Additionally, the solution must be cost-effective to be competitive against traditional storage systems, encouraging adoption within sectors where storage budgets are tightly managed. Together, these components from a comprehensive operational structure that supports the business in delivering a highly-quality, innovative storage solution tailored to meet the unique demands of long-term, high-capacity data storage customers.

9.2 Key Processes

The three key processes from our High Level SIPOC diagram we chose to highlight are the Enzyme Synthesis process, installation process, and maintenance processes. These three processes cover a good portion of the timeline of our high-level process. Enzyme synthesis is a key part of the creation of our product and is one of the most technical parts of our product. The installation process shows the transition from our product being in our hands to the users, which is very important and needs to go well. The maintenance process shows what we will need to do in continuity even after our product is installed.

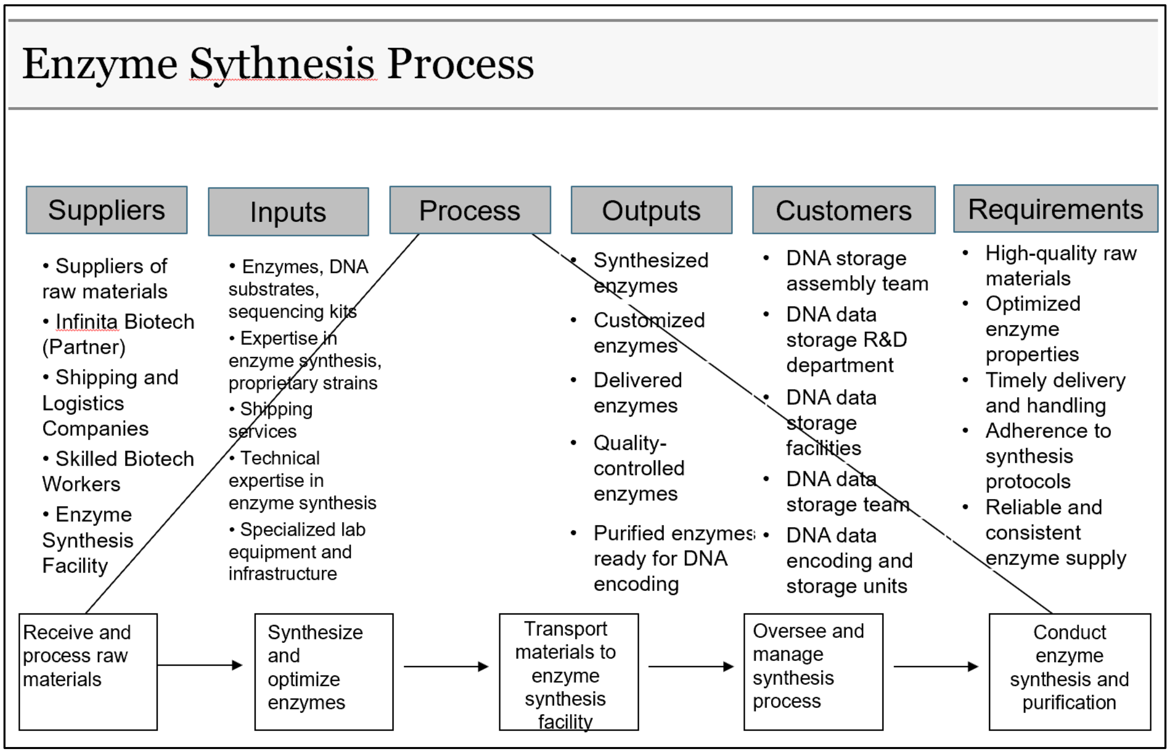


Figure 15. SIPOC Diagram for Enzyme Synthesis Process

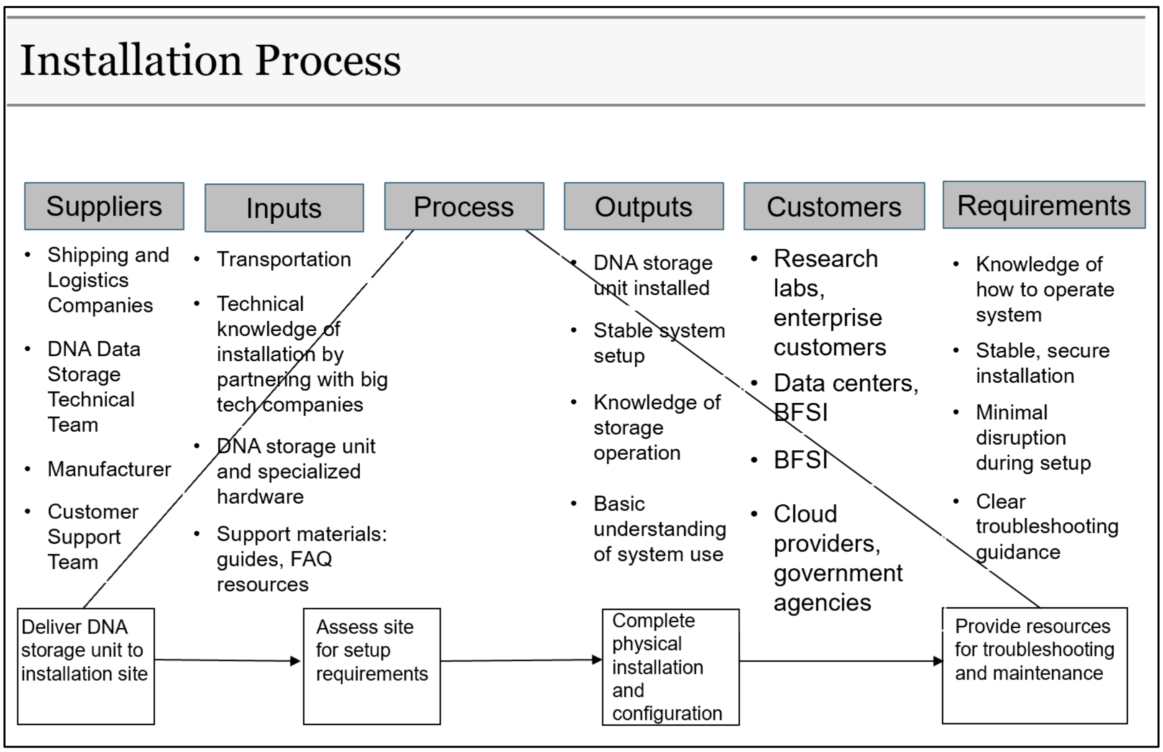


Figure 16. SIPOC Diagram for Installation Process

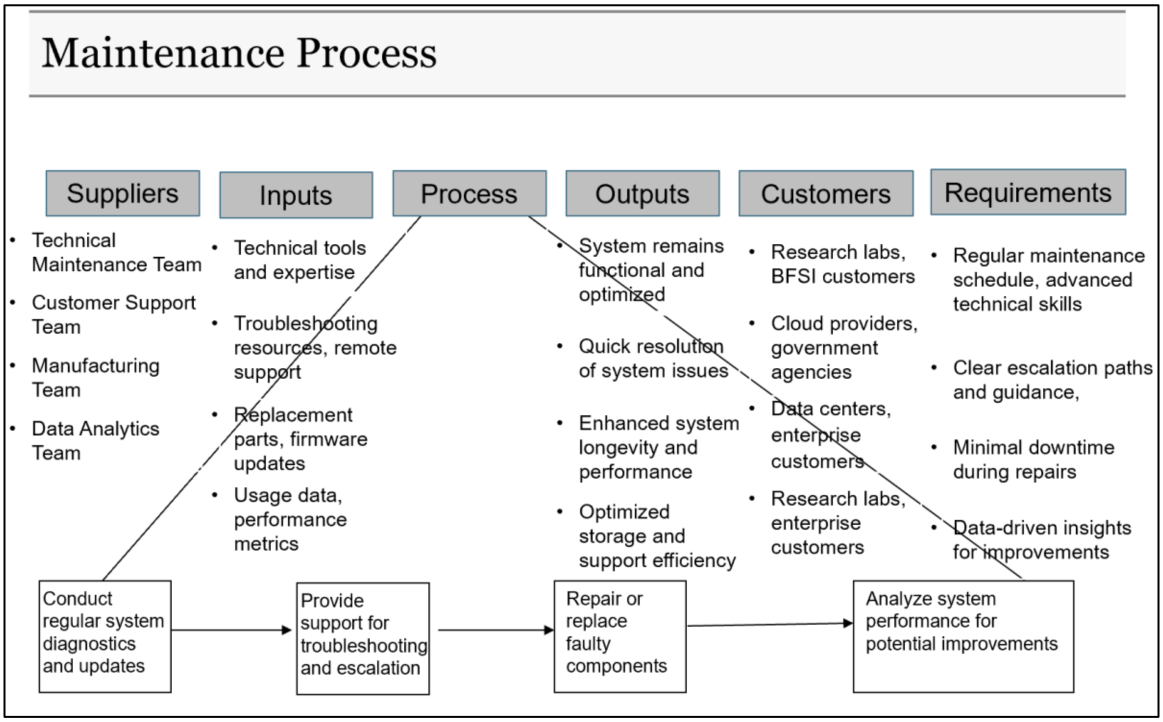


Figure 17. SIPOC Diagram for Maintenance

9.3 Sourcing

For our DNA data storage product it is very important for us to carry out our enzymatic synthesis internally as this is a novel part of our technology. We can utilize the infrastructure of our host company IBM to do this. IBM can establish a small in-house team for research and development in enzymatic synthesis, specifically in computational biology, to optimize DNA storage methods and retain proprietary innovations. Data encoding and decoding algorithms should be developed internally, drawing on IBM's expertise in encryption, error correction, and complex storage algorithms. This approach ensures that IBM can maintain proprietary control over these essential processes. Additionally, the development of data analysis and management platforms should be managed internally to create efficient systems for handling and retrieving DNA-stored data, potentially transforming this capability into a profit center by offering clients custom DNA data storage and management solutions.

Outsourcing will be strategically applied to areas where IBM does not specialize. For example, biological enzyme production could be contracted out to a biotech firm with expertise in large-scale enzyme production, as this ensures high-quality materials while allowing IBM to focus on data-driven processes. This outside firm could be Infinita Biotech which is a company based in India who is a major player in enzyme production as they currently predict 30% growth in the enzyme market (Times of India). In addition, IBM's data management and analysis platforms could be adjusted for DNA applications, providing a unique market offering that integrates its strengths in data analytics with DNA data storage. By building data-centric capabilities internally and outsourcing biological manufacturing, IBM can maximize its competitive edge while maintaining control over the proprietary aspects of DNA data management innovations.

10 Technology Commercialization/Collaboration Strategy

Stakeholder	Dependence	Interest	Posture	Approach
Infinita Biotech	High	Expanding applications of enzyme technology and entering new high-tech markets	Supportive in a way that complements our goals without direct competition	Persuade
IBM	Medium	Innovating in sustainable, high-capacity storage solutions	Interest in collaboration but may seek IP or revenue-sharing agreements	Collaborate
Illumina	High	Expanding DNA sequencing applications into the data storage industry	May want control over proprietary sequencing methods	Collaborate

Figure 18. Key Stakeholder Assessment Table

10.1 Capabilities sought

To successfully advance our DNA data storage product that leverages enzyme-based synthesis, we will seek strategic collaboration. We will seek partnerships with research institutions and biotech firms like Infinita for their expertise in enzymatic synthesis and reaction optimization, and with companies specializing in high-throughput DNA sequencing and synthesis technologies to enhance speed and cost-effectiveness. Collaborators with computational biology capabilities will aid in developing encoding and error correction algorithms, while regulatory experts will ensure our processes meet biosafety and data compliance standards. Additionally, partnerships with data centers and cloud providers will support hybrid storage integration, and legal firms specializing in biotechnology will protect our intellectual property. Finally, strategic collaborations with established technology companies will facilitate market access, co-development, and targeted marketing efforts to reach key industries such as finance, healthcare, and cloud computing.

10.2 Prospective collaborators

In the pursuit of advancing our enzyme-based DNA data storage technology, we are considering strategic partnerships with several key companies, each bringing unique expertise and resources to the table.

Infinita Biotech emerges as a leading prospective collaborator with their specialized knowledge in enzymatic DNA synthesis and reaction optimization. Infinita's research in enzyme engineering is directly applicable to our core technology, potentially enhancing the efficiency of our DNA data storage processes. This collaboration would directly contribute to the technical capabilities we are seeking to develop.

IBM is another valuable potential partner, particularly for their extensive experience in data storage technologies. IBM's research division has made significant strides in developing advanced algorithms for data encoding and error correction, which are crucial for the reliability and scalability of DNA-based data storage. By collaborating with IBM, we can leverage their computational expertise and integrate it with our biological systems, ensuring robust data integrity and retrieval. They will also be able to help integrate our technology with preexisting data storage methods due to their deep pool of knowledge when it comes to any type of data storage.

Lastly, another company that aligns well with our objectives is Illumina, a global leader in DNA sequencing technology. Their high-efficacy sequencing technology could be instrumental in reading the stored data accurately and quickly, a vital component for the practical application of DNA data storage. Partnering with Illumina would address the need for rapid and reliable data retrieval systems, which is a critical part of our technology’s market viability.

Each of these companies offers distinct advantages and resources that, when combined through strategic partnerships, would significantly increase the development and commercialization capabilities of our DNA data storage technology.

10.3 Collaborator assessment

We evaluated the capabilities, interests, and stances of three key companies: Infinita Biotech, IBM, and Illumina.

Infinita Biotech is a leading candidate for collaboration due to their expertise in enzyme engineering and synthesis, which directly aligns with our need for efficient enzyme-based DNA synthesis. Infinita’s expertise in reaction optimization and enzyme scalability can significantly increase the efficiency and cost-effectiveness of our data storage processes, a critical factor for commercial viability. This partnership will increase their market presence, particularly in high-tech industries outside of traditional biotechnology. Given that Infinita is not a direct competitor in data storage, their stance on our innovation is likely to be collaborative and supportive. They bring complementary technical expertise without competing in our primary market, making them an ideal partner to advance the enzyme synthesis aspect of our solution.

IBM offers a valuable partnership opportunity, particularly due to their deep expertise in data storage technologies, computational biology, and advanced algorithms for data encoding and error correction. IBM’s knowledge of high-throughput data management and encryption is essential to creating reliable and scalable encoding algorithms, while a hybrid storage infrastructure can facilitate the integration of DNA data storage with existing digital systems. IBM has a vested interest in pioneering new data storage methods, and DNA storage aligns with their vision for high-capacity, sustainable storage solutions. By supporting DNA data storage, IBM can strengthen its reputation as a leader in next-generation storage technology. However, IBM may seek intellectual property (IP) rights or revenue sharing agreements to secure returns on their significant contributions to algorithmic development. Given IBM’s significant experience and influence in data storage, this partnership could allow us to leverage their strengths without direct market competition.

Illumina, a global leader in DNA sequencing, is another potential partner for whom high-throughput sequencing technology is essential for rapid and accurate extraction of data from DNA. Illumina’s sequencing technology advances could play a vital role in making DNA data storage commercially viable by improving the speed and reliability of data readout processes, a key requirement for practical application. Illumina’s interest in expanding applications of DNA sequencing beyond the biomedical domain makes this partnership mutually beneficial as they can enter the data storage market and showcase the versatility of their technology. Given that DNA data storage represents a new application for sequencing technologies, their stance on our innovation is likely to be collaborative. However, Illumina may want to maintain some control over the use of their sequencing technology in order to protect their proprietary methods.

These collaborations address key technical, regulatory, and market needs, positioning us for success in sectors like finance, healthcare, and cloud computing.

10.4 Collaboration form

The optimal collaboration strategy with Infinita Biotech is *Partnering*. By forming a strategic partnership, we can leverage their specialized expertise in enzyme engineering and synthesis, which is crucial for optimizing our enzyme-based DNA synthesis processes. This partnership will be structured to mutually benefit both parties, allowing us to enhance the efficiency and cost-effectiveness of our data storage solutions while helping Infinita expand their presence in non-traditional, high-tech sectors. Given that Infinita does not compete directly in the data storage market, this collaborative approach will foster a supportive, non-competitive relationship that leverages their technical strengths without IP conflicts.

For IBM, we will pursue a combination of *Partnering* and *Licensing*. This hybrid strategy ensures we can access IBM\u2019s deep expertise in data storage technologies, computational biology, and advanced data encoding algorithms while formalizing an agreement that outlines the use of their proprietary technology. A partnership will allow us to co-develop scalable encoding and management systems, integrating DNA data storage with existing digital infrastructure. The licensing aspect will address IBM\u2019s potential requirement for IP rights or revenue sharing, ensuring a clear framework for contributions and returns. This collaboration aligns with IBM\u2019s strategic interest in sustainable, next-generation storage solutions and allows us to benefit from their technological prowess without direct market competition.

The strategy with Illumina will be *Partnering* with a consideration for *Licensing* if proprietary sequencing methods are involved. Partnering will allow us to incorporate Illumina\u2019s cutting-edge high-throughput sequencing technology to enhance the speed and accuracy of data readout processes, which is vital for commercial viability. Given Illumina\u2019s goal of diversifying the application of their sequencing technology, this partnership is likely to be highly collaborative, positioning them to gain a foothold in the emerging DNA data storage market. Licensing agreements may be necessary to address their need for control over proprietary technology and ensure compliance with their standards while allowing us to integrate their advancements effectively into our solutions.

11 Project Valuation & Financing

11.1 Profit Model

Considering the innovative nature of our DNA data storage technology and its application within data centers, particularly in the BFSI industry, adopting a traditional profit model is the most suitable strategy for maximizing profit. This model will involve selling our DNA data storage systems at a premium, while also offering service and maintenance contracts.

The initial setup of our DNA data storage systems will involve some costs and after initial setup, our technology can be expanded to fit the clients needs. These costs are mitigated by the savings in space, energy consumption, and the efficiency compared to traditional data storage. Our pricing strategy will reflect the value of these benefits, and we anticipate that clients within the BFSI sector will recognize the advantages and be willing to pay a premium. Service and maintenance contracts will provide continuous revenue and ensure that our systems remain operational and up to date.

The primary challenge of this profit model lies in the initial higher costs compared to other data storage solutions. However, as production scales and our processes become more optimized, we expect costs to decrease, allowing for the possibility of more competitive pricing in the future. Until then, our focus will be on showing clients that our DNA data storage technology will revolutionize data storage and be worth the investment.

11.2 Pro forma financial statement

The pro forma financial statement outlines the projected outcomes for sales and distribution of DNA data storage units, offering essential insights for company executives and potential investors in this cutting-edge technology. Based on the example of a comparable startup, Catalog, which was founded in 2016 and have since raised $32 million in funding but has yet to commercially launch its product, we have determined that an appropriate launch date for our product would be in 2032. Unlike Catalog, however, we benefit from the extensive resources of our host company, IBM, which places us in a significantly stronger position to accelerate development and achieve a commercial launch in a shorter timeframe. This advantage allows us to leverage IBM's established infrastructure, technical expertise, and market presence to potentially bring our DNA data storage units to market ahead of the estimated timeline.

The served market percentage for DNA data storage units is projected to begin at 2% and rise to 10% by 2039. This forecast is based on our current R&D budget, shown in Figure X, and influenced by broader macroeconomic factors, including Biden's executive order targeting a 65% reduction in emissions by 2030. If these environmental policies continue through subsequent administrations and the trend toward stricter environmental regulations persists, it is likely to create favorable conditions for sustainable technologies like DNA data storage. Moreover, Goldman Sachs Research anticipates a 160% increase in data center power demand by 2030 (Goldman). As the demand for data storage escalates, the environmental impact could become significant, with data center CO_2 emissions potentially more than doubling between 2022 and 2030 (Goldman). This heightened demand underscores the urgent need for sustainable storage solutions, positioning DNA data storage as an attractive, eco-friendly alternative poised to address both market growth and environmental concerns.

The industry estimation is based upon AWS who has about a quarter of the market (Research Nester). We're projecting up to 10% by 2039 as we are a smaller and less developed service compared to Amazon, but our novel technology will allow us to consume a new portion of the market, one that leans towards sustainability.

The anticipated price for DNA data storage units is projected to be around $470 per 20 TB, positioning it competitively within the current data storage market. This pricing is derived by comparing the current cost of a 20 TB hard disk drive (HDD), such as those offered by Western Digital, which stands at approximately $460. Although DNA data storage is a cutting-edge technology, our aim is to align its price with established storage solutions to make it accessible and attractive to early adopters and environmentally conscious organizations. As our production scales and technological advances drive costs down, we anticipate that DNA data storage could become an increasingly viable alternative for both enterprise and consumer markets, offering the dual benefits of high-density storage and reduced environmental impact.

The capital investment is based upon a data storage company in their first year after their IPO. Equinix produces data centers and in their first public year 2000 their capital investment was roughly $600,000 (Equinix Annual Report). Considering this cost was for the lease of the data center as well as the storage mediums, we expect a smaller capital investment of $130,000 in the first year as our product takes up less space and we are the storage medium not the center itself.

The initial cost per unit estimate for DNA data storage is derived from industry data and Equinix's financial records, factoring in production materials, labor, and operational needs. Based on Equinix's financial data, initial variable costs reflect supplies and equipment required for DNA synthesis and storage processes. In 2024, Equinix processed around 300 TB of data with net sales of approximately $1.5 million, resulting in an estimated cost of $5,000 per TB. Equinix's total operational costs were $700,000, averaging around $2,333 per TB or $2.33 per GB (Equinix 10-k).

Labor costs are projected at $250,000, assuming DNA data storage operations scale proportionally with anticipated larger setups than the startup. Equinix's total payroll in 2024 was $180,000, and scaling it up for larger operations reflects additional production and technical demands. Assuming an initial selling price of $470 per 20 TB, the profit per TB, after subtracting costs, would be approximately $1.12. For sustainability, 20% of this profit is allocated for research and development, equating to $0.22 per TB reinvested. Capital costs, estimated at $0.15 per GB, are based on Equinix's non-current assets, including machinery and fixed equipment costs. This results in a total initial cost of $3.13 per GB.

The experience curve factor for DNA data storage is estimated at 9%, chosen by averaging comparable tech, manufacturing, and BFSI industries, including biotech with an 8-15% curve and high-tech manufacturing at 10-20%.

The R&D costs for DNA data storage technology are projected to decrease over time, starting at a high of $5,000,000 in the initial years and gradually tapering down to $810,000. Early-stage development demands significant investment, with the first few years ranging from $4,500,000 to $2,500,000, primarily dedicated to advancing the technology and scaling up for production. As the technology matures, the required R&D expenditure decreases, with the later years demanding between $1,000,000 and $900,000, and eventually $810,000. This progressive reduction reflects the natural shift from foundational research toward more incremental improvements. Most R&D expenses will be incurred before full commercialization, covering technology maturation and readiness for large-scale manufacturing.

Other operating expenses for DNA data storage units utilizing enzymatic synthesis technology include marketing costs, employee salaries, facilities, utilities, insurance, and various general and administrative (G&A) costs not directly related to the production process. These expenses are estimated to start at $500,000 in the initial year, reflecting the higher complexity and marketing demands associated with introducing enzymatic DNA data storage technology. As the technology matures and production becomes more efficient, these costs are projected to decrease gradually to $150,000 per year. This reduction aligns with increased operational efficiencies and market familiarity with the product, reducing the need for high marketing expenditure. Enzymatic synthesis-based DNA storage does not require a specific service life assumption, as the technology will be continuously improved and optimized over time.

We predict our service life to be roughly one hundred years given the ability for DNA to be used as a long term storage source. DNA can hold data for thousands of years, but given our current technology we estimate that ours would last closer to 100 (Gervasio, Joao).

The discount rate is estimated to be 32% based on the overall cost of capital for DNA data storage technology using enzymatic synthesis. This rate accommodates the risk-free rate of return along with an assumed project risk premium. The risk premium is projected at 20%, reflecting the high-risk nature of this large-scale, innovative project. Given that DNA data storage technology is still in its early stages, substantial risks are involved. However, if successful, the project holds significant potential for profitability. While additional data on specific metrics may refine these estimates, a 32% discount rate effectively accounts for the project's cost of capital and inherent risks.

The risk-free rate of return represents the interest earned if funds were invested in a completely risk-free asset, like a government-issued bond. The Financial Pro Forma spreadsheet currently sets this rate at 1.04%, but it may be updated to align with the prevailing US Treasury yield rate.

The total BFSI Data Storage market projects to consist of 5,000,000,000 units of 20 TB each (roughly 100 ZB) (Fortune Business Insights).

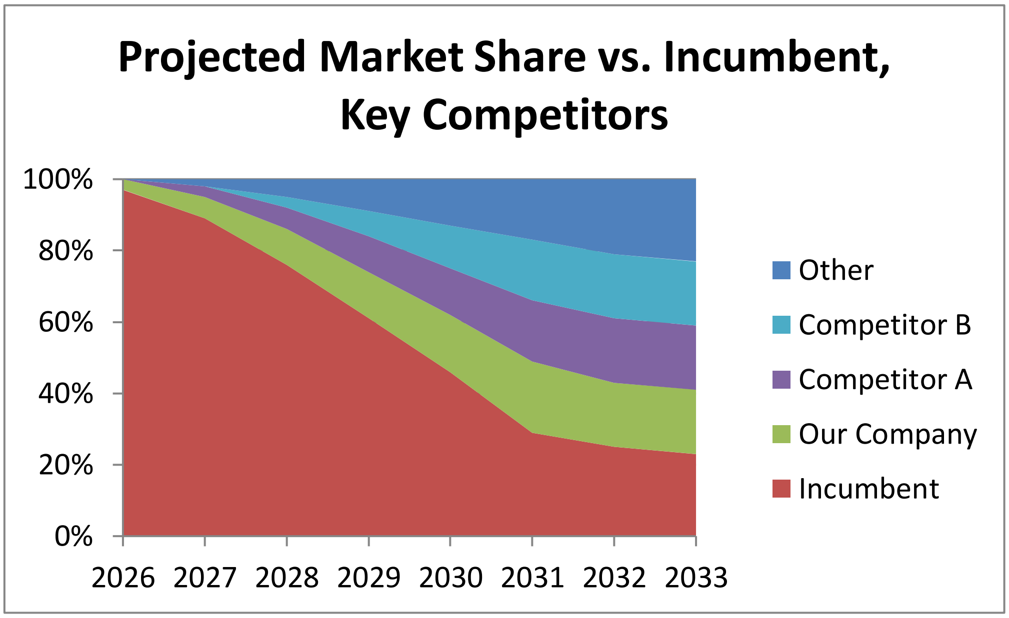


Figure 19. Projected Market Share vs. Incumbent Technology and Key Competitors.

11.3 Results

Based on projections from the pro forma financial statement, the gross margin for DNA Data Storage technology is forecasted to grow steadily from approximately $51,763 in 2032, the planned product launch year, to $3,723,668,478 by 2039. This growth reflects an increase in the gross margin percentage from 0% to 32% over this period, indicating a healthy scaling in profitability as the technology gains traction in the market. Figure 21 illustrates the steady increase in annual sales during the initial years after the technology's introduction, which supports the upward trend in gross margin.

Figure 20 further demonstrates improvement in net cash flow, with the business anticipated to achieve positive cash flow around year 2. The actual cash flow is projected to rise significantly from -$4,528,238 in 2032 to $1,650,471,469 by 2039, showing the technology's increasing economic viability. Similarly, the discounted cash flow is expected to grow from -$2,599,638 in 2032 to roughly $135,842,262 by 2039, underscoring the long-term financial potential of the technology. As highlighted in Figure 20, the cumulative net cash flow could reach as much as $3,506,318,091 by 2039.

The net present value (NPV) of the DNA Data Storage technology is estimated to be about $396,895,336, with a payback period of 3.29 years and a discounted payback period of 4.07 years, indicating a quick return on investment for early stakeholders. One area that could be optimized to maximize profitability is the price per unit, which begins at $470 per unit. Although this initial price point results in a narrow profit margin due to the close per unit cost of $468 per unit, the purchased input cost inflator will improve these margins over time, gradually enhancing the profitability of each unit. This incremental increase in margins will further strengthen the economic sustainability and attractiveness of DNA Data Storage technology as it matures in the market.

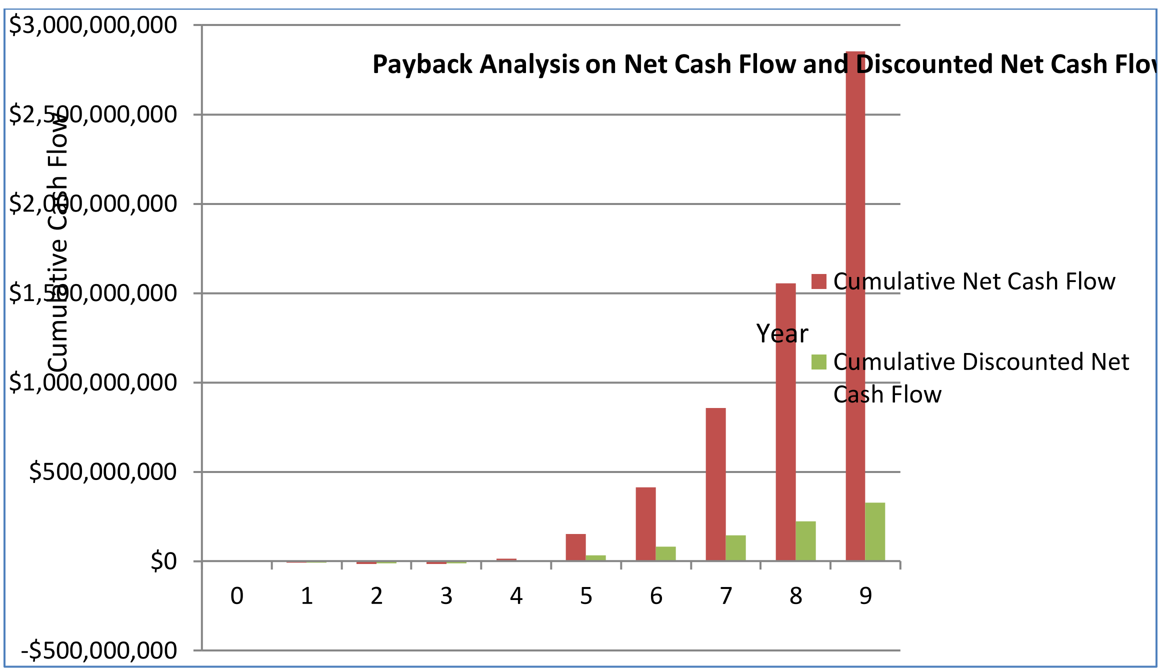


Figure 20. Payback Analysis on Net Cash Flow and Discounted Net Cash Flow

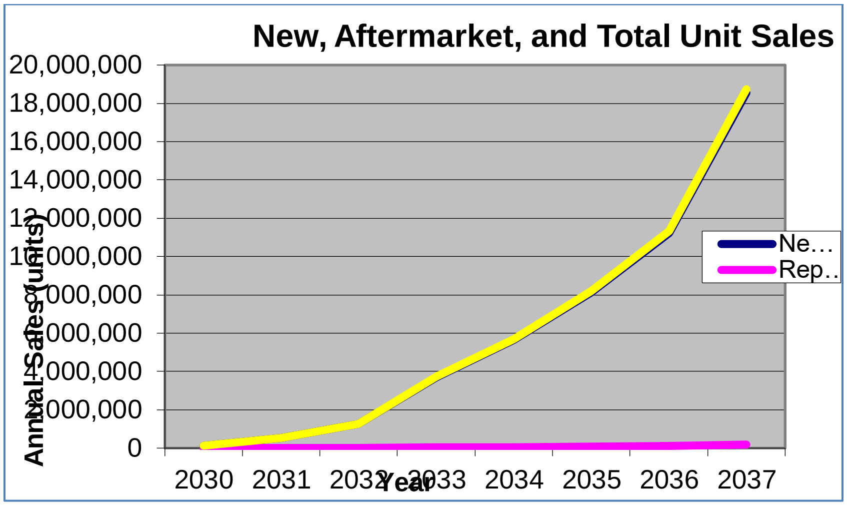


Figure 21. New, Aftermarket, and Total Unit Sales

11.4 Capital Requirements and Sources of Capital

The startup capital required will be approximately $130.000. This is the sum of the net discounted cash flow after we launched our product in 2032 for the first two years before the business starts making a profit. As our budget for R&D is $5 million, the ideal source of capital is internal funding from IBM, the host company. IBM has one of the largest R&D budgets in the technology sector. In 2023, IBM spent $6.78 billion on research and development (IBM Annual Report 2023), reflecting its commitment to pioneering advancements across a range of technologies, including artificial intelligence, cloud computing, quantum computing, and data storage solutions. IBM consistently ranks as a top company in patent filings, highlighting its focus on developing and owning new technologies. An additional advantage of internal funding is that obtaining the necessary capital from the host company is less costly than borrowing from external sources, as it doesn’t incur interest or require dividend payments. On the BCG Growth-Share Matrix, our technology falls within the “Question Mark” category, as the business unit has a small market share in a high growth market. Although competing products may require capital, DNA data storage offers a unique, long-term sustainable solution that could move to the **Star quadrant** as the technology advances and gains market traction. Given the capital-intensive nature of DNA data storage, we’ll pursue venture capital (VC) and strategic partnerships with firms specializing in biotech or data storage, such as Infinita Biotech and Illumina, who could potentially be equity partners. External sources, particularly VCs, are more likely to invest in a multi-round strategy, allowing us to secure additional funds as we achieve key milestones. This would allow the company to maintain a higher level of control over the product and the process. If investors provided capital in multiple rounds, the company would first request **$10 million** to validate enzyme synthesis, develop a prototype, and establish partnerships. In the next round, **$5 million** would fund R&D for scaling synthesis, optimizing protocols, and pilot projects. A final **$10 million** would support the full-scale product launch, customer acquisition, production expansion, and technical support scaling.

12 Project Valuation & Financing-Real Options Analysis

12.1 Real options projects

The path to full commercialization of a DNA data storage product begins with the R&D phase, where the primary focus is refining the technology to make it scalable, cost-effective, and user-friendly. Key milestones in this stage include demonstrating the feasibility of DNA as a reliable data storage medium, ensuring cost efficiency and energy savings compared to traditional storage methods, and developing a user interface that allows non-technical users to easily interact with the system. Additionally, a partnership with our prospective host company IBM will be vital for refining the product and securing intellectual property through patents. Once these objectives are met, the product will be ready for a testing phase.

The next phase involves testing and market validation, where the product will be introduced to a controlled market, such as a few firms in the BFSI sector. This stage allows for gathering valuable feedback to refine the technology and address unforeseen issues, while also securing further investment based on successful trials. Key milestones include establishing performance benchmarks for data retrieval, attracting early adopters, and generating publicity through marketing efforts. The final stage, full commercialization, focuses on scaling production, expanding the market reach, and capturing an initial 2% of market share. This phase will include partnerships with global firms, continuous innovation based on user feedback, and positioning DNA data storage as a mainstream solution for enterprises and data centers, ensuring its long-term success.

12.2Option value calculation

The value of the DNA data storage project was calculated using a decision tree analysis, which accounts for multiple stages of development and varying outcomes at each stage. This method allows for a more nuanced evaluation of the project's potential compared to traditional Discounted Cash Flow (DCF) analysis, as it considers best-, base-, and worst-case scenarios. In this analysis, each phase—R&D, testing and market validation, and full commercialization—has associated costs, revenues, and risks that influence the overall value of the initiative.

The R&D phase is projected to require an initial investment of $15 million over three years, with key milestones including demonstrating scalability, cost-efficiency, and developing a user-friendly interface. The assumption for the base case is that these milestones are met on time and within budget, enabling the project to move to the testing phase. In the testing and market validation phase, an additional $10 million investment is expected for trials and feedback collection from early adopters, primarily in the BFSI sector. If successful, this phase will position the product for full commercialization.

For the commercialization phase, an estimated $25 million is needed to scale production, establish partnerships with global firms, and capture 2% of the market share. The best-case scenario assumes rapid adoption, with market share growing to 5% within five years, driven by strong demand for green alternatives and increasing regulatory pressure on traditional storage technologies. In the worst-case scenario, the market share could remain low, and delays in product development or market acceptance could push the project’s NPV into negative territory.

The decision tree analysis integrates the flexibility to adapt to new information at each stage, allowing the project to pivot if needed. For example, if the R&D phase encounters unexpected delays or cost overruns, the project could be paused or abandoned, limiting potential losses. Conversely, success in earlier stages increases the likelihood of moving forward to the next phase, enhancing the project's overall value. This dynamic approach provides a more accurate and realistic projection of the project's potential outcomes, considering both the risks and opportunities inherent in the commercialization of DNA data storage.

12.3 Options space map

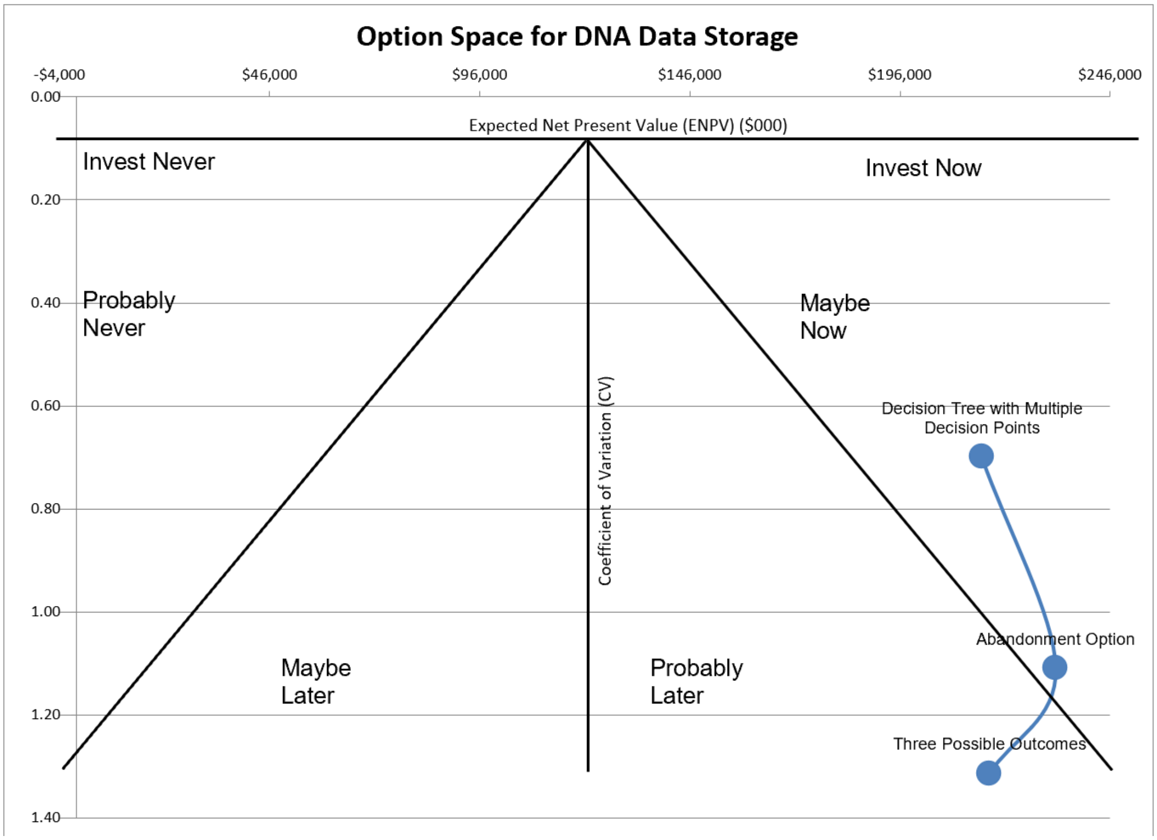


Figure 22. Option Space Diagram

In the R&D phase, the project focuses on proving the concept of DNA as a viable data storage medium, enhancing the technology for scalability, ensuring cost-effectiveness, and developing an accessible user interface. A partnership with IBM is crucial for refining the product and securing patents. The completion of these milestones sets the stage for the next phase.

The testing and market validation phase involves introducing the product to a controlled market segment, such as firms in the BFSI sector. This phase is critical for gathering feedback, which is used to refine the technology further and solve any unexpected problems. Achieving performance benchmarks, attracting early adopters, and creating market buzz through marketing efforts are the key milestones. Success in this phase is instrumental for securing additional investment.

The final phase, full commercialization, is where the product is scaled up for mass production and broader market introduction. The goal is to capture an initial 2% of the market share. This phase includes forming partnerships with global firms, innovating based on user feedback, and establishing the product as a mainstream solution for data storage. Achieving these milestones is expected to lead to the long-term success of the DNA data storage product.

Each phase is dependent on the successful completion of its milestones and is linked to the subsequent phase. The progression from one phase to the next is not linear but rather contingent upon meeting specific objectives that validate moving forward. This approach allows for flexibility and adaptability, ensuring that the project can respond to new information and changing market conditions.

Both the "decision tree with multiple decision points" and the "abandonment case" fall under the "Maybe Now" categories. This means that investing in these now could lead to a profit but waiting could potentially produce better returns. This decision model is the most likely to be profitable due to the multiple stages of investment which ensures that better outcomes are more likely. The "three possible choices" case falls into the "Probably later" category. This case might be profitable, but it has an uncertain future, so it makes the most sense to wait.

12.4Improvement over traditional DCF analysis

The decision tree analysis is a more accurate approach for evaluating the DNA data storage project than traditional DCF analysis because it incorporates multiple stages of investment and considers best, base, and worst-case scenarios. This flexibility allows for adjustments if milestones, such as cost-efficiency or market validation, are not achieved, something DCF analysis cannot accommodate.

In the R&D and testing phases, decision tree analysis accounts for technical and market feedback variations, enabling strategic pivots, while DCF assumes a single projected path. Additionally, the decision tree includes an abandonment option, critical for riskier technologies, ensuring a realistic valuation by considering potential failures or changes in market conditions.

Overall, the decision tree provides a dynamic and risk-aware framework, essential for an innovative project like DNA data storage, where success depends on sequential milestones and adaptability.

13 References

"2024 Poll: Americans' Views on Climate Change and Policy in 12 Charts." *EPIC*, June 2024, https://epic.uchicago.edu/insights/2024-poll-americans-views-on-climate-change-and-policy-in-12-charts/.

"20TB Ultrastar DC HC560." *Western Digital*, www.westerndigital.com/products/internal-drives/data-center-drives/ultrastar-dc-hc560-hdd?sku=0F38785. Accessed 12 Nov. 2024.

Jan 18, 2023. "Infinita Biotech Projects 30% Growth in Enzyme Market - Times of India." *The Times of India*, TOI, timesofindia.indiatimes.com/business/india-business/infinita-biotech-projects-30-growth-in-enzyme-market/articleshow/97082532.cms. Accessed 4 Nov. 2024.

"AI Is Poised to Drive 160% Increase in Data Center Power Demand." *Goldman Sachs*, 14 May 2024, www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand.

Alex Hoose et Al. "DNA Synthesis Technologies to Close the Gene Writing Gap." *Nature Reviews Chemistry*, Jan. 2023, https://www.nature.com/articles/s41570-022-00456-9.

Andrea Doricchi. "Emerging Approaches to DNA Data Storage: Challenges and Prospects." *ACS Publications*, Oct. 2022, https://pubs.acs.org/doi/10.1021/acsnano.2c06748.

Athow, Desire. "DNA Storage Finally Reaches Mainstream (Well, Sort of) — but It Will Cost You a Whopping $1,000 per KB (Yes, Kilobyte)." *Techradar Pro*, 4 Dec. 2023, https://www.techradar.com/pro/dna-storage-finally-reaches-mainstream-well-sort-of-but-it-will-cost-you-a-whopping-dollar1000-per-kb-yes-kilobyte.

"Annual Reports." *Equinix, Inc.*, investor.equinix.com/sec-filings/annual-reports?form_type_custom=https%3A%2F%2Finvestor.equinix.com%2Fsec-filings%2Fannual-reports&year=2001##document-478-0001012870-01-001327-2. Accessed 12 Nov. 2024.

Avoiding Data Storage Crisis. https://ciosea.economictimes.indiatimes.com/news/next-gen-technologies/how-can-we-avoid-a-data-storage-crisis/93516724.

Bellingham, Rob. "The Growing Energy Demands of Data Storage." *Dataversity*, May 2022, https://www.dataversity.net/the-growing-energy-demands-of-data-storage/.

"BFSI Security Market Report by Type (Physical Security, Cyber Security), Physical Security Offering (System, Service), Information Security Offering (Solution, Service), Enterprise Size (Large Enterprises, Small and Medium Enterprises), End User (Banks, Insurance Companies, and Others), and Region 2024-2032." *Imarc*, https://www.imarcgroup.com/bfsi-security-market.

"Catalog Secures $35 Million in Series B Funding, DNA-Based Computation Expected next Year." *Catalog*, catalogdna.com/blog-posts/catalog-secures-35-million-series-b-funding. Accessed 12 Nov. 2024.

Carlson, Rob. "The Quest for a DNA Data Drive." *IEEE Spectrum*, Feb. 2024, https://spectrum.ieee.org/dna-data-storage.

Cell Press: Trends in Cell Biology, www.cell.com/trends/cell-biology/home. Accessed 13 Nov. 2024.

Church, George, and Don Ingber. "DNA Data Storage." *Wyss Institute,* https://wyss.harvard.edu/technology/dna-data-storage/.

"Could All Your Digital Photos Be Stored as DNA?" MIT News | Massachusetts Institute of Technology, 10 June 2021, news.mit.edu/2021/dna-data-storage-0610.

"Data Storage Market Size, Share & Industry Analysis, By Storage Medium (Direct Attached Storage, Network Attached Storage, Software Defined Storage and

Hyper Converged Storage, Cloud Storage, and Storage Area Network), By Industry (BFSI, IT and Telecommunication, Governments and Public Sector, Manufacturing, Healthcare and Life Sciences, Retail and Consumer Goods, Media and Entertainment, and Others), By Enterprise Type (SoHo, Mid-Size, Large Enterprises), and Regional Forecast 2024-2032 Source: Https://Www.Fortunebusinessinsights.Com/Data-Storage-Market-102991." *Fortune Business Insights*, Sept. 2024, https://www.fortunebusinessinsights.com/data-storage-market-102991.

"DNA Data Storage Market Size - By Technology (Sequence Based DNA Data Storage, Structure Based DNA Data Storage), End-Use Industry (Biotechnology and Healthcare, Banking & Finance, Government and Defense, Media and Entertainment), Type & Forecast, 2024 - 2032." *Global Market Insights*, Jan. 2024, https://www.gminsights.com/industry-analysis/dna-data-storage-market.

Dutta, Tapapriya. "How Can We Avoid a Data Storage Crisis?" *CIO.Com*, https://ciosea.economictimes.indiatimes.com/news/next-gen-technologies/how-can-we-avoid-a-data-storage-crisis/93516724.

Education, IBM Cloud. "Hard disk drive vs Solid state drive." IBM, 2 Feb. 2022, www.ibm.com/think/topics/hard-disk-drive-vs-solid-state-drive. Accessed 23 Aug. 2024.

Education, IBM Cloud. "What Is Operational Excellence?" IBM, 23 Aug. 2024, www.ibm.com/think/topics/operational-excellence.

Eojin Yoo et Al. "Mini Review: Enzyme-Based DNA Synthesis and Selective Retrieval for Data Storage." *Science Direct*, Apr. 2021, https://www.sciencedirect.com/science/article/pii/S2001037021001690.

Foote, Keith D. "A Brief History of Data Storage." *DATAVERSITY*, 4 Jan. 2023, www.dataversity.net/brief-history-data-storage/. Accessed 23 Sept. 2024.

Funk. "What Americans Think about COVID-19 Vaccines." *Pew Research Center*, May 2023, https://spectrum.ieee.org/dna-data-storage.

Gervasio. "How Close Are We to Storing Data in DNA?" *Trends in Biotechnology*, https://www.dataversity.net/the-growing-energy-demands-of-data-storage/.

Hawkins, Kerry, and Michael Restivo. "Data Centers: Expensive to Build, but Worth Every Penny." *JLL*, https://www.us.jll.com/en/views/data-centers-expensive-to-build-but-worth-every-penny#:~:text=Core%20and%20Shell,of%20%24200%20per%20square%20foot.

"IBM Cloud Market Share." 6sense, https://6sense.com/tech/web-and-application-servers/ibm-cloud-market-share. Accessed 30 Sept. 2024.

IBM. "Explore the 2023 IBM Annual Report Today." *2023 IBM Annual Report*, 12 Mar. 2024, www.ibm.com/annualreport/.

IBM. 2023 IBM Annual Report. IBM, https://www.ibm.com/annualreport/. Accessed 30 Sept. 2024.

Ionkov, Latchesar, and Bradley Settlemyer. "DNA: The Ultimate Data-Storage Solution." *SCIAM*, May 2021, https://www.scientificamerican.com/article/dna-the-ultimate-data-storage-solution/#:~:text=By%20comparison%2C%20DNA%20storage%20is,environs%

20of%20a%20data%20center.

Karishma Matange et Al. “DNA Stability: A Central Design Consideration for DNA Data Storage Systems.” *Nature Communications*, Mar. 2021, https://www.nature.com/articles/s41467-021-21587-5.

Ng, Cheuk Chi A et al. “Data storage using peptide sequences.” *Nature communications* vol. 12,1 4242. 13 Jul. 2021, doi:10.1038/s41467-021-24496-9

“North America Data Center Trends H2 2023.” *CBRE*, Mar. 2024, https://www.cbre.com/insights/reports/north-america-data-center-trends-h2-2023.

McDowell, Steve. “IBM Realigns Its Storage Business to Match Data-Driven Enterprise Needs.” *Forbes*, Forbes Magazine, 11 Sept. 2023, www.forbes.com/sites/stevemcdowell/2023/03/07/ibm-realigns-its-storage-business-to-match-data-driven-enterprise-needs/.

Philippidis, Alex. “Illumina, Microsoft, Twist Lead New DNA Data Storage Alliance.” *GEN*, GEN - Genetic Engineering and Biotechnology News, 2 June 2023, www.genengnews.com/news/illumina-microsoft-twist-lead-new-dna-data-storage-alliance/.

Pradeep Vancheeswaran. “Data Modernization in BFSI: Overcoming Legacy Challenges.” *Experion Technologies*, https://experionglobal.com/data-modernization/.

“PRESERVING OUR DIGITAL LEGACY: AN INTRODUCTION TO DNA DATA STORAGE.” *DNA Data Storage Alliance*, June 2021, https://dnastoragealliance.org/dev/wp-content/uploads/2021/06/DNA-Data-Storage-Alliance-An-Introduction-to-DNA-Data-Storage.pdf.

Rawat, Sachin. *Is DNA Data Storage Ready for Data Centers*. Nov. 2023, https://www.synbiobeta.com/read/is-dna-data-storage-ready-for-data-centers.

"Public Cloud in BFSI Market Size & Share, by Type (Software as a Service (SAAS), Platform as a Service (Paas), Infrastructure as a Service (Iaas)); Component; Enterprise Size; End Use - Global Supply & Demand Analysis, Growth Forecasts, Statistics Report 2025-2037." *Research Nester*, www.researchnester.com/reports/public-cloud-in-bfsi-market/6482. Accessed 12 Nov. 2024.

Rigby, Darrell K., and Jeff Sutherland. "How IBM, Intuit, and Rich Products Became More Customer-Centric." Harvard Business Review, 24 June 2015, https://hbr.org/2015/06/how-ibm-intuit-and-rich-products-became-more-customer-centric. Accessed 30 Sept. 2024.

Saadia Arshad et Al. "Analysis of Security and Privacy Challenges for DNA-Genomics Applications and Databases." *Science Direct*, May 2021, https://www.sciencedirect.com/science/article/pii/S1532046421001441.

Satia, Sanyam. "Data Lake at Robinhood." *Robinhood Newsroom*, 29 July 2022, newsroom.aboutrobinhood.com/data-lake-at-robinhood/#:~:text=We%20use%20PostgreSQL%20and%20AWS,of%20each%20of%20these%20systems. Accessed 26 Oct. 2024.

Sec.Gov, www.sec.gov/Archives/edgar/data/51143/000155837023002376/ibm-20221231x10k.htm#Item1Business_962012. Accessed 30 Sept. 2024.

Shevlin, Ron. "JPMorgan Chase To Spend $12 Billion On Technology...And Why Other Banks Can't Keep Up." *Forbes*, Apr. 2022, https://www.forbes.com/sites/ronshevlin/2022/01/18/jpmorgan-chase-to-spend-

12-billion-on-technologyand-why-other-banks-cant-keep-up/.

Taylor, Petroc. "Forecast Number of Mobile Users Worldwide from 2020 to 2025." *Statista*, Nov. 2023, https://www.statista.com/statistics/218984/number-of-global-mobile-users-since-2010/#:~:text=In%202021%2C%20the%20number%20of,projected%20to%20reach%207.49%20billion.

Tozzi, Christopher. "Why Communities Are Protesting Data Centers – And How the Industry Can Respond." *Data Center Knowledge*, June 2024, https://www.datacenterknowledge.com/data-center-construction/why-communities-are-protesting-data-centers-and-how-the-industry-can-respond.

"Why Hdds Dominate Hyperscale Cloud Architecture: Seagate US." *Seagate.Com*, www.seagate.com/blog/why-hdds-dominate-hyperscale-cloud-architecture/. Accessed 23 Sept. 2024.

"IBM: Reasonable Price, But You Need To Manage Your Risk." *Seeking Alpha*, Sept. 2024, https://seekingalpha.com/article/4724297-ibm-stock-reasonable-price-but-need-to-manage-risk.

"IBM Releases Second-Quarter Results." *IBM Newsroom*, July 2024, https://newsroom.ibm.com/2024-07-24-IBM-RELEASES-SECOND-QUARTER-RESULTS?sf195888222=1#:~:text=IBM%20(NYSE:%20IBM)%20today%20announced%20second-quarter%202024%20earnings%20results.%20%22We.

"IBM Stock Overview." *Simply Wall St*, Sept. 2024, https://simplywall.st/stocks/us/software/nyse-ibm/international-business-machines.

"International Business Machines Corporation (IBM) Is Attracting Investor Attention:

Here Is What You Should Know." *Zacks Equity Research*, Sept. 2024, https://www.zacks.com/stock/news/2343215/international-business-machines-corporation-ibm-is-attracting-investor-attention-here-is-what-you-should-know#:~:text=The%20Zacks%20Computer%20-%20Integrated%20Systems%20industry,%20to%20which%20IBM.

Bernadette Tansey. "DNA Edges Forward As Data Storage Option." *Semiconductor Engineering*, Sept. 2022, https://semiengineering.com/dna-edges-forward-as-data-storage-option/#:~:text=DNA%20data%20storage%20technology%20builds%20on%20the%20advances%20of%20those.

Matt Shipman. "For First Time, DNA Tech Offers Both Data Storage and Computing Functions." *NC State*, Aug. 2024, https://news.ncsu.edu/2024/08/functional-dna-computing/#:~:text=Researchers%20from%20North%20Carolina%20State%20University%20and%20Johns%20Hopkins%20University.

Michael Wu. "5 Ways Data Is Transforming the Financial Industry." *Fast Company*, Nov. 2023, https://www.fastcompany.com/90985394/5-ways-data-is-transforming-the-financial-industry.

Muhammad Hassan Raza, et al. "An Outlook on the Current Challenges and Opportunities in DNA Data Storage." *Science Direct*, vol. 66, Apr. 2024, https://www.sciencedirect.com/science/article/abs/pii/S0734975023000629.

North Carolina State University. "The Next Frontier: DNA Emerges as a Powerhouse for Data Storage and Computing." *SciTechDaily*, Aug. 2024, https://scitechdaily.com/the-next-frontier-dna-emerges-as-a-powerhouse-for-data-storage-and-computing/#:~:text=A%20new%20technology%20using%20DNA%20for%20data%20storage%20and%20computing.

Piper, Kelsey. "The next Deadly Pathogen Could Come from a Rogue Scientist. Here's How We Can Prevent That." *Vox*, 11 Feb. 2020, www.vox.com/future-perfect/2020/2/11/21076585/dna-synthesis-assembly-viruses-biosecurity. Accessed 22 Oct. 2024.

14 Appendices

14.1 Contact Log

Name	Company/Institution	Status
Nathaniel Chism	CATALOG	Interview
Karin Strauss	Microsoft	Response
Emily Leproust	Twist Bioscience	No Response
Albert Keung	NC State	Response
Paul Hook	Johns Hopkins	Response
George Church	Harvard	Interview
Winston Timp	John Hopkins	No Response
Orlin Velev	NC State	No Response

Figure 23. Contact Log

14.2 Interview Summaries

Interview 1:

Question: What are some of the primary obstacles you're facing in scaling the production of synthetic DNA for large-scale data storage?

Nathaniel Chism: The main challenge lies in the high complexity and cost of producing synthetic DNA at scale. While the technology is promising, the methods we use now aren't fast or cost-effective enough for large-scale production. Enzymatic synthesis shows potential as an alternative, but it's not yet efficient or automated enough to support high-volume manufacturing. We need to increase speed, reduce costs, and ensure low error rates to make this commercially viable.

Interviewer: Are there any breakthroughs in enzymatic synthesis that you're particularly excited about?

Nathaniel Chism: Absolutely. Enzymatic synthesis has the potential to greatly speed up DNA production while lowering costs. We're particularly excited about advancements that are making this process more automated and increasing the throughput of enzymes. If we can speed up synthesis without sacrificing accuracy, it will open the door for large-scale DNA data storage. But we still need a lot of progress before it becomes a viable alternative to chemical methods.

Interviewer: From a business standpoint, how do you see DNA data storage being applied in the banking industry, and what are the main challenges to its adoption?

Nathaniel Chism: The banking sector has very specific needs when it comes to data storage—primarily security, reliability, and scalability. DNA storage has great potential in terms of data density and durability, but it faces several challenges. For example, data retrieval speed is currently a limiting factor. Banks need fast access to their data, and with current sequencing technology, that's not feasible yet. There's also the issue of integrating DNA storage with existing financial systems, as well as meeting strict regulatory requirements. We're working closely with financial institutions to address these concerns and build a business model that supports DNA storage within the banking industry.

Interview 2:

Interviewer: What do you think is the biggest hurdle to making DNA data storage commercially viable?

George Church: The major challenge is reading the information stored in DNA and converting it back into digital data. Sequencing technology is improving, but it's still too slow and expensive for practical use at scale. Additionally, reducing the errors in sequencing is crucial for ensuring the data is reconstructed accurately. Until we solve these issues, DNA data storage will remain in the experimental phase.

Interviewer: In terms of integration, what challenges do you foresee when merging DNA data storage with current digital storage systems, especially in sectors like banking?

George Church: Compatibility with existing digital storage systems is one of the key challenges. DNA data storage requires a different approach, and there's no universal method for converting digital data into DNA sequences and back yet. For industries like banking, which rely on secure and standardized systems, it's critical that DNA storage can integrate seamlessly while maintaining data security and integrity. Developing standardized protocols to bridge this gap will be essential.

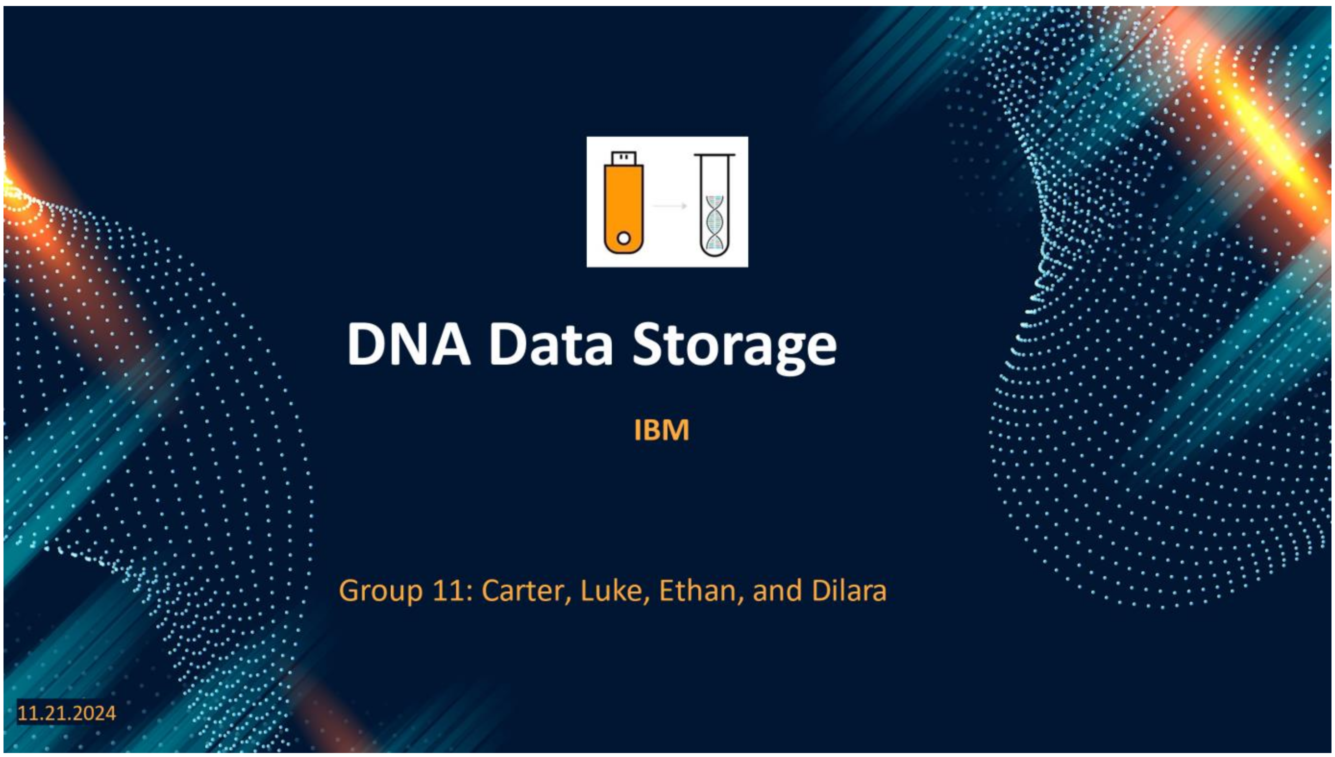
Interviewer: Looking ahead, how do you envision the future of DNA data storage? What's needed for it to become a mainstream solution?

George Church: The next few years will be critical for DNA data storage to transition from a research concept to a commercially viable solution. It's not just about the science; it's also about creating the infrastructure and standards for widespread adoption. We're in the early stages, but with ongoing innovation, DNA storage will likely become a viable option for niche applications, and eventually, mainstream use.

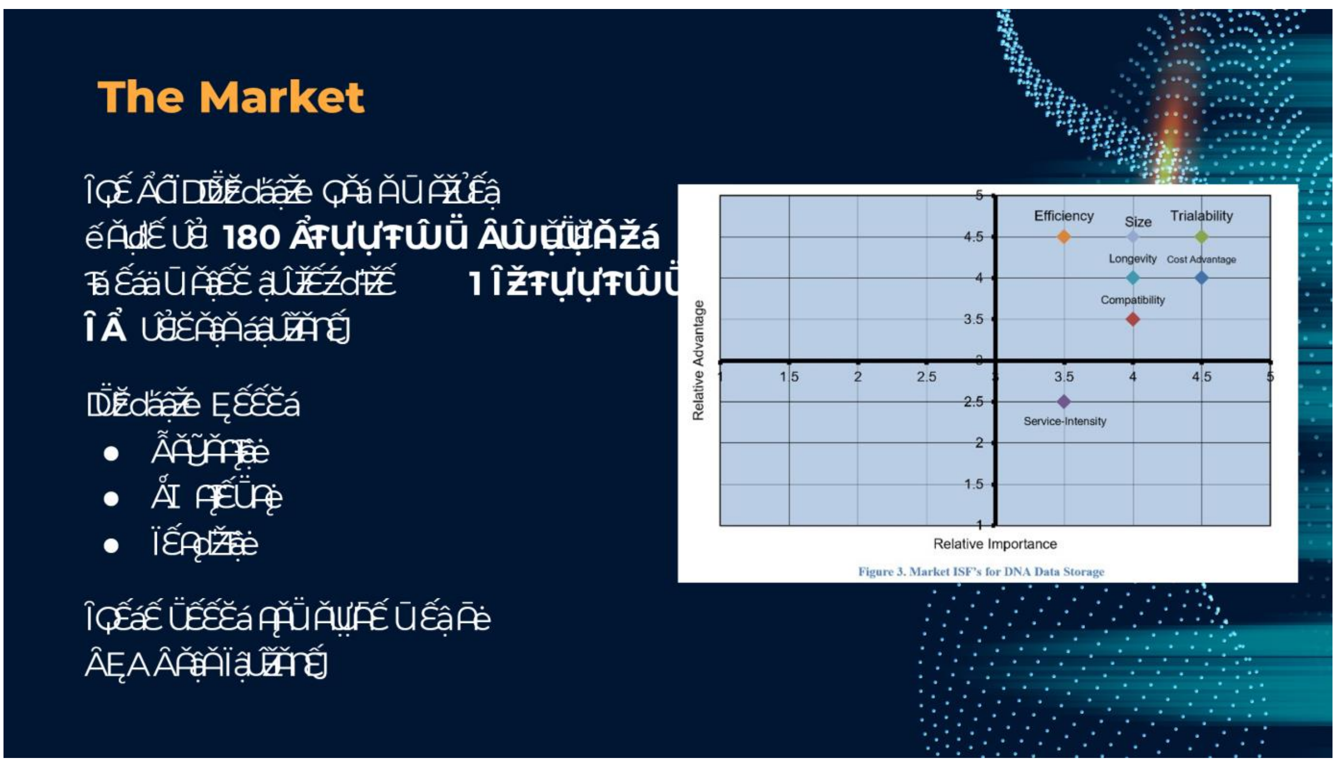
14.3 Other Appendices

Pro Forma Financial Statement Access

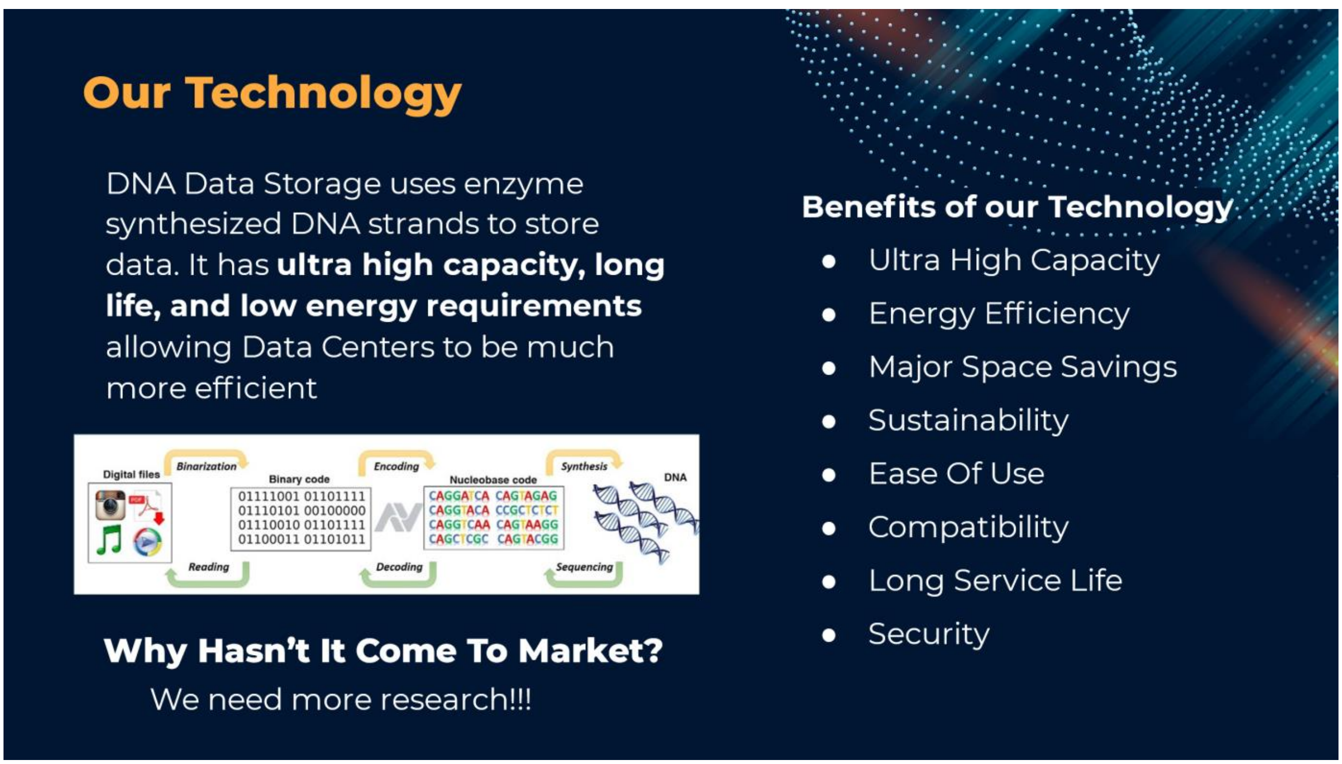
14.4Final Presentation Handout



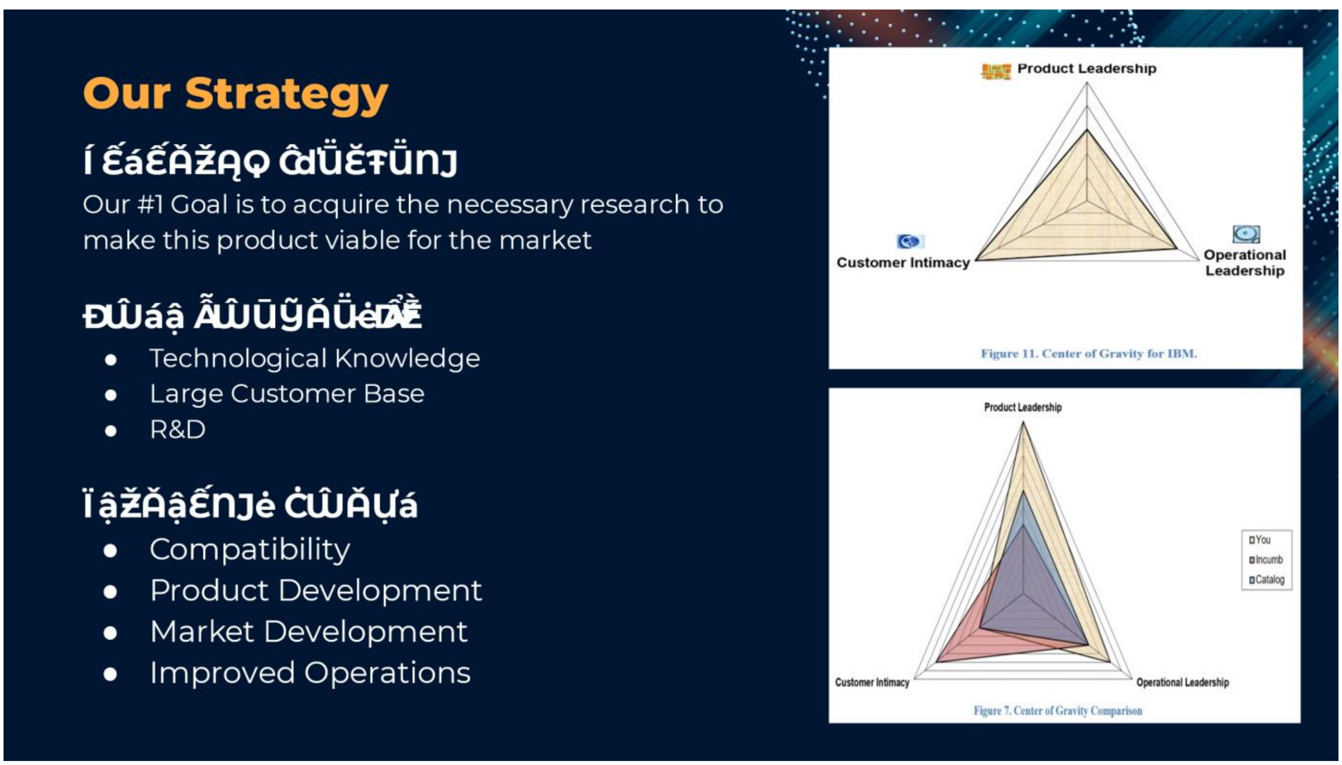

The Problem
Growing Demand
Demand for data storage is growing too fast for current storage methods to handle
Sustainability
Current data centers take up too much space and use far too much power at ~50 Megawatts of power per center
DNA Data Storage uses enzyme based DNA synthesis to store data with extreme efficiency

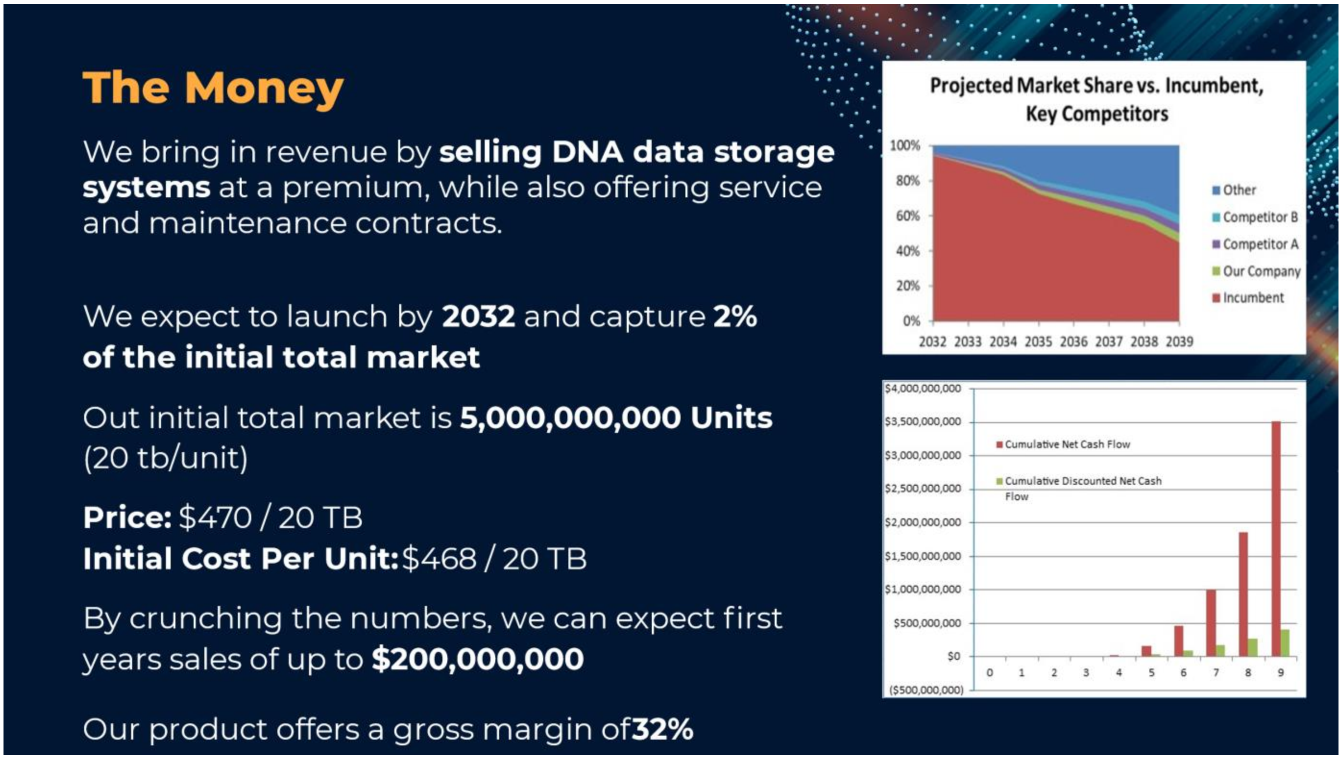

The Market
180
Relative Advantage
Relative Importance
Efficiency
Size
Trialability
Longevity
Cost Advantage
Compatibility
Service-Intensity
Figure 3. Market ISF's for DNA Data Storage


Historical Solutions
Current Data Centers use thousands of traditional storage devices including Hard Disk Drives and Solid State Drives that have lower capacity, shorter life, and higher energy requirements
Hard disk
Flash memory
Bacterial DNA
Read-write speed (μs per bit) ~3,000–5,000 ~100 <100
Data retention (years) >10 >10 >100
Power usage (watts per gigabyte) ~0.04 ~0.01–0.04 <10^-10
Data density (bits per cm³) ~10^13 ~10^16 ~10^19
Our Competitors
Traditional Storage
AWS
Microsoft
IBM
Other Technologies
Quantum
Diamond
5D Optical
Holographic


Our Technology
DNA Data Storage uses enzyme synthesized DNA strands to store data. It has ultra high capacity, long life, and low energy requirements allowing Data Centers to be much more efficient
Digital files
Binarization
Binary code
Encoding
Nucleobase code
Synthesis
DNA
Reading
Decoding
Sequencing
Why Hasn't It Come To Market?
We need more research!!!
Benefits of our Technology
Ultra High Capacity
Energy Efficiency
Major Space Savings
Sustainability
Ease Of Use
Compatibility
Long Service Life
Security


What we offer
We aim to offer Premium Data Storage Solutions via our DNA Data Storage Systems to members of the BFSI Industry
Our Target Market
Our product is perfect for the BFSI Industry meeting and even exceeding the markets Data Storage needs
Value
With a projected 27% increase in data storage needs, the value of a premium technology like this is extremely high
4
1
3
2
Relative Advantage
Relative Importance
Figure 12. ISF Matrix for DNA Data Storage Portraying the Strengths, Weaknesses, Opportunities, and Threats.


Our Strategy
Our #1 Goal is to acquire the necessary research to make this product viable for the market
Technological Knowledge
Large Customer Base
R&D
Compatibility
Product Development
Market Development
Improved Operations
Product Leadership
Customer Intimacy
Operational Leadership
Figure 11. Center of Gravity for IBM.
Product Leadership
You
Incumb
Catalog
Customer Intimacy
Operational Leadership
Figure 7. Center of Gravity Comparison


The Money
We bring in revenue by **selling DNA data storage systems** at a premium, while also offering service and maintenance contracts.
We expect to launch by **2032** and capture **2% of the initial total market**
Out initial total market is **5,000,000,000 Units** (20 tb/unit)
Price: $470 / 20 TB
Initial Cost Per Unit: $468 / 20 TB
By crunching the numbers, we can expect first years sales of up to **$200,000,000**
Our product offers a gross margin of **32%**
Projected Market Share vs. Incumbent, Key Competitors
100%
80%
60%
40%
20%
0%
2032 2033 2034 2035 2036 2037 2038 2039
Other
Competitor B
Competitor A
Our Company
Incumbent
$4,000,000,000
$3,500,000,000
$3,000,000,000
$2,500,000,000
$2,000,000,000
$1,500,000,000
$1,000,000,000
$500,000,000
$0
($500,000,000)
Cumulative Net Cash Flow
Cumulative Discounted Net Cash Flow
0 1 2 3 4 5 6 7 8 9


Resources Required
Round 1:
for just **5 Million Dollars**, we can validate enzyme synthesis, develop a prototype, and establish partnerships
Round 2:
With **just 4 Million Dollars**, we can fund R&D for scaling synthesis, optimizing protocols, and pilot projects
Round 3:
With **3 Million Dollars**, we can support the full-scale product launch, customer acquisition, production expansion, and technical support scaling.