TWIST BIOSCIENCE

Scaling up from Data Lakes to Data Oceans

STEFFEN HELLMOLD

JUNE 2022

DNA Data Storage
Market Opportunity

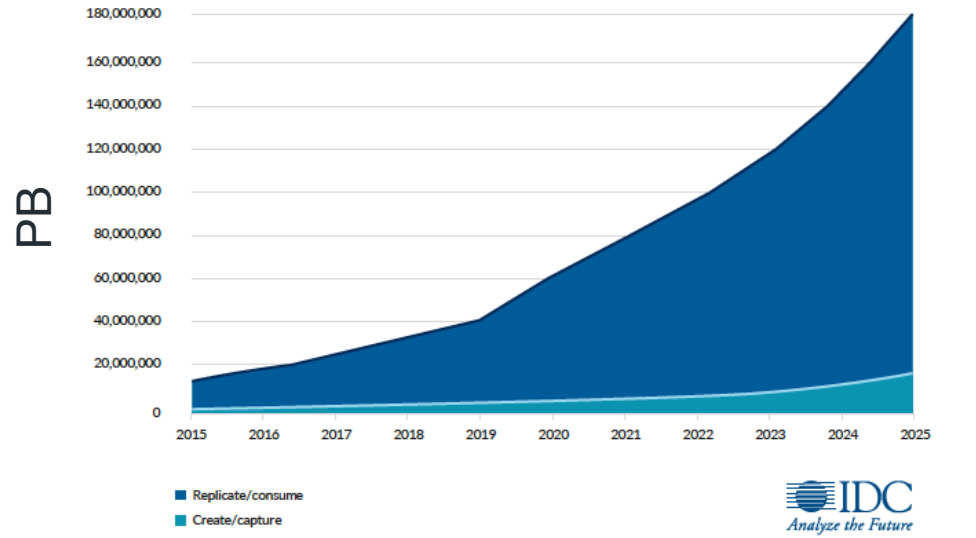
Data Creation & Storage Continues to Grow Exponentially

Data created, replicated and consumed:

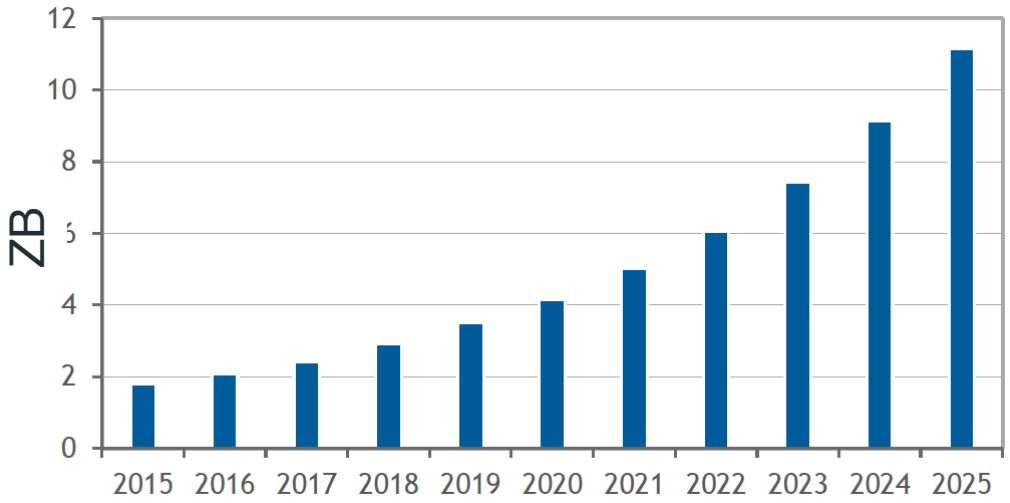
2021: 80 ZB → 2025: 180 ZB

Data storage worldwide capacity:

2021: 5 ZB → 2025: 11 ZB



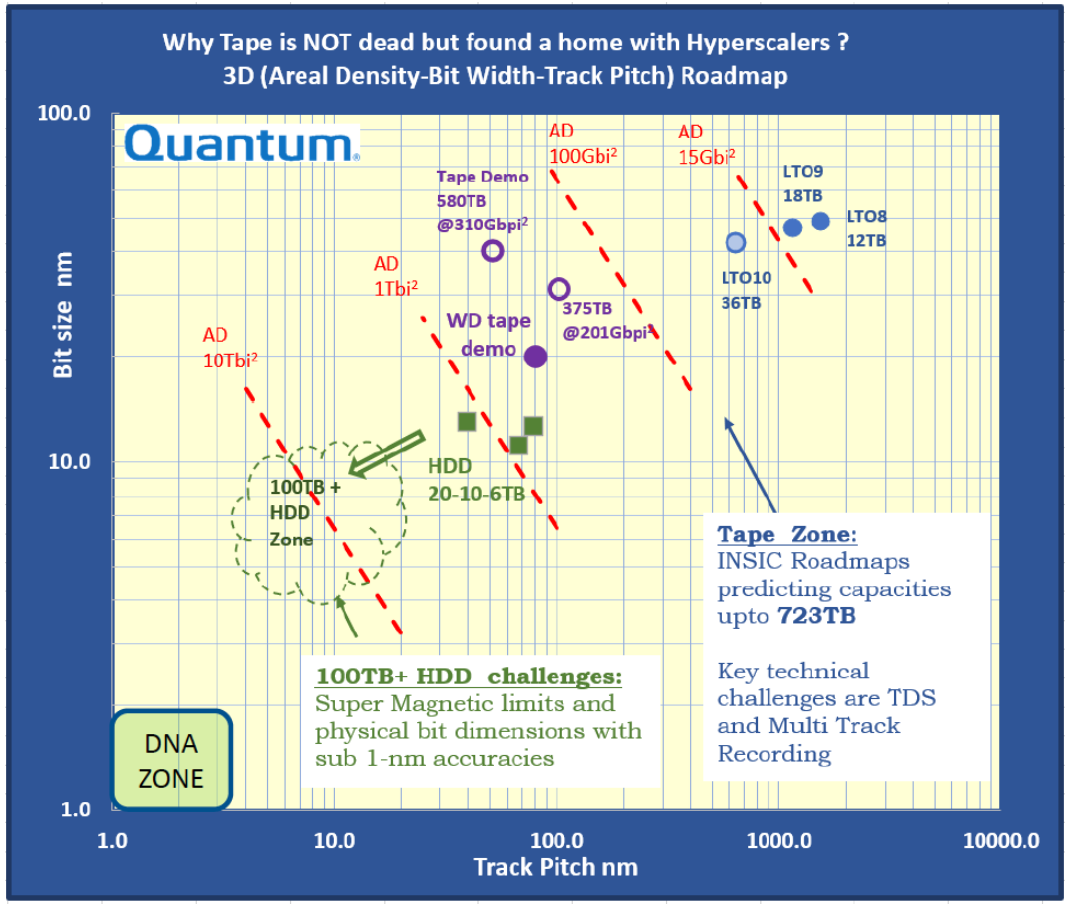
Source: IDC, Worldwide Global DataSphere Forecast, 2021–2025: The World Keeps Creating More Data – Now, What Do We Do with It All?, #US46410421



Source: IDC Global StorageSphere, 2021

Challenges of Current Data Storage Technologies, Scaling et al.

- Increasing physical scaling challenges
  - Magnetic storage scaling is slowing down
- Supply can't keep up with demand
  - ZB-scale supply gap in 2nd half of this decade (Gartner)
- Increasing demand for media diversity
  - Tape is the only true archive storage medium today
- Limited longevity of current data storage media
  - Require migration typically every 7 - 10 years
- Increasing sustainability considerations
  - Reducing resource utilization, energy & carbon footprint

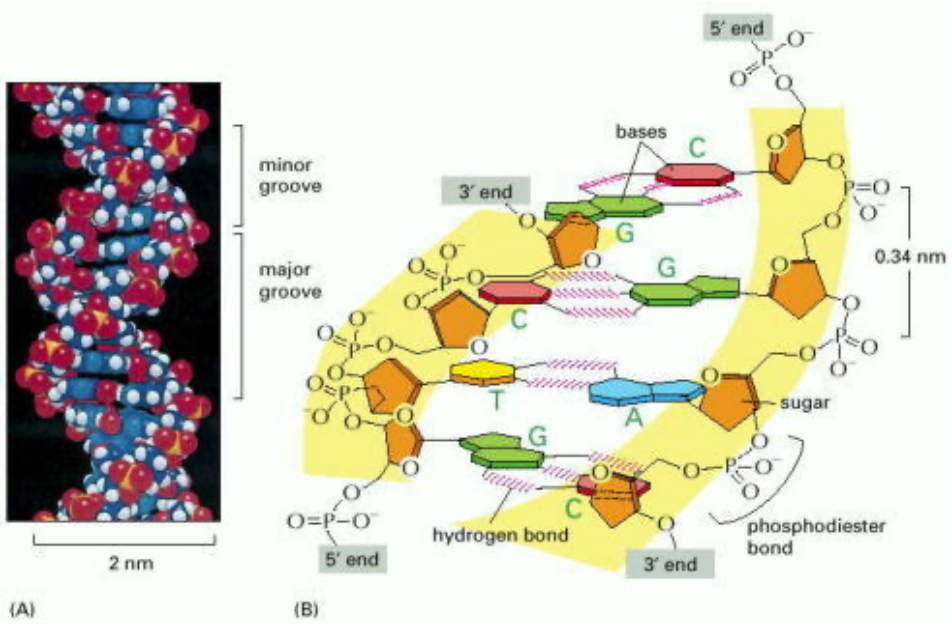


Source: https://www.snia.org/educational-library/lto-technology-and-two-dimensional-erasure-coded-long-term-archival-storage-rail

DNA enables high-density archive storage

DNA Data Storage – Designing Storage Using Nature's Playbook

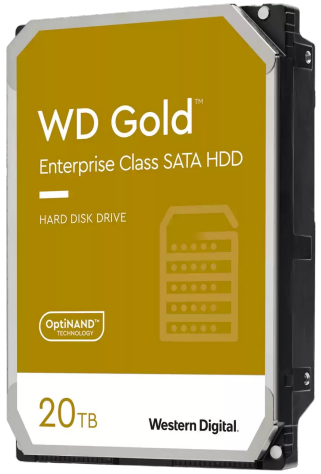
- The physics of DNA is well understood
- Synthesis & sequencing technologies exists
- DNA bases store bits: A, C, T, G → 00, 10, 01, 11
- Enabling century scale archive storage solutions
- Data is the Medium, *Software Defined Storage*
- Stable format, always able to read natural DNA
- Sustainable, lowest energy storage carbon footprint



Source: https://www.ncbi.nlm.nih.gov/books/NBK26821/

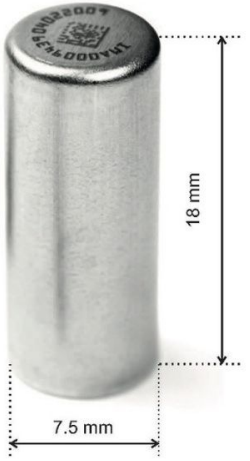
DNA Data Storage is delivering a unique value proposition, initially addressing deep archive to accessible archive use cases

Storage Capacity No Issue for DNA Data Storage



Capacity: 20 TB
By volume: 51.3 MB/mm^3
By weight: 29 GB/g

Capacity: 18 TB
By volume: 77.4 MB/mm^3
By weight: 90 GB/g



Capacity: 250 µl
By volume: ≈ **16.6 B/nm^3**
By weight: ≈ **450 EB/g**

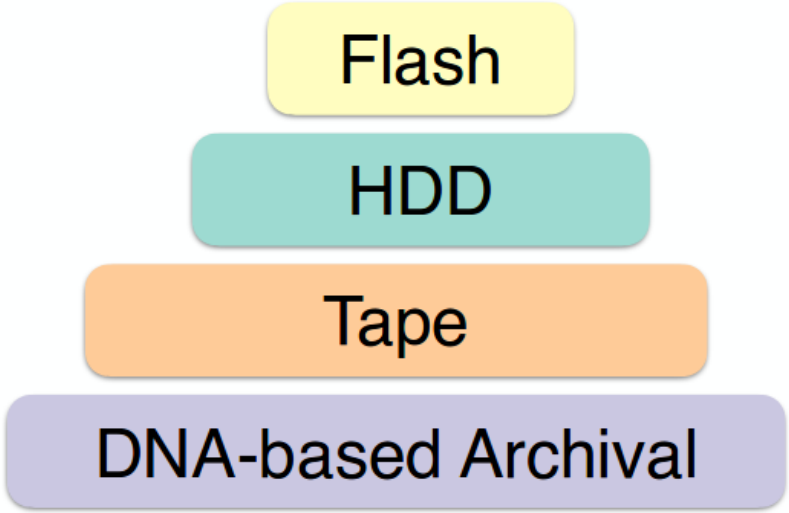
29,000x volumetric density
5,000,000x mass density
>10x migration longevity

The Decadal Plan for Semiconductor - Storage Grand Goal:

Discover storage technologies with **>100x storage density capability** and **new storage systems** that can leverage these new technologies

Source: https://www.src.org/about/decadal-plan/

DNA Data Storage Emerging as ‘Time Capsule’ Archive Storage



	Access Time	*Capacity*	*Durability*
Flash	*μs*-ms	TBs	~5 yrs
HDD	10s ms	100s TBs	~5 yrs
Tape	minutes	PBs	~10s yrs
DNA-based Archival	hours	ZBs	~100s yrs

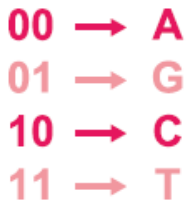
Source: https://pdfs.semanticscholar.org/7b06/ba3effa9fc7b2f194a355bcb69601ef1ea56.pdf

TODAY:
MB-class

SOON:
GB-class
Early Access Solutions

DNA Data Storage Technology

DNA Data Storage Workflow

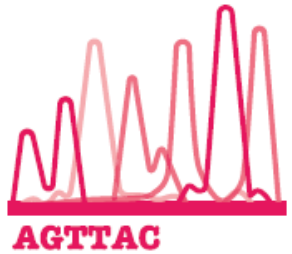


Coding

Synthesis

Storage

Retrieval



Sequencing

Decoding

Goal: Develop a Chip that Produces TB Scale Coded DNA

- DNA is synthesized on a chip
  - Use a 2D array of electrochemical reactors to synthesize strands of DNA
  - After synthesis, the DNA is washed into a tube, then amplified, purified, and packaged
- Chip capacity is limited by the array pitch and chip size
  - There is a scaling limit; each reactor needs to produce enough DNA to practically store
  - Given the scaling limit, 1 TB from a chip is the practical limit – otherwise the chip becomes too large
- Twist's chip capacity roadmap
  - 64 GB → 256 GB → 1 TB
  - *We are working on the 64 GB chip*
- Synthesis sustainability considerations
  - Enzymatic DNA Synthesis (EDS) ideal technology
  - Cost effective EDS technology is a key enabler

Packaging

Imagene's DNASHELL

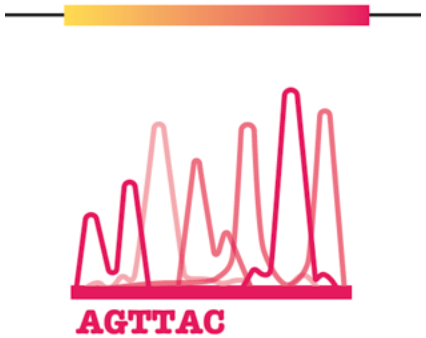
7.5mm x 18mm

- DNA degrades by oxidation
  - Hermetically packaging DNA leads to a long shelf life
  - The package can be checked periodically for leaks – no leaks, no degradation
- DNA is dense, but packaging needs to be practical
  - Industrial automation required for process steps
  - And tubes that can be laser welded shut
- Barcoded tubes can be packed in arrays
  - Arrays are configurable
  - Array sizes: 96 TB, 384 TB, or 1,536 TB per bio automation spec

96 DNASHELL Array

Sequencing

- Current sequencers focused on bio (genomics) applications
- Practical for up to GB-class DNA Data Storage
- Single run currently takes approximately 24 hours
- Overall sequencing cost depends on reading frequency
- As DNA sequencing cost is declining market will expand
- Multiple groups working on new sequencing technologies



Sequencing

DNA Data Storage Productization

DNA Data Storage Solutions Concepts

Vault

- Offline / Offsite data archiving solution
- Air gapped / Hacker safe
- Very low maintenance costs
- High density / Small footprint
- Immutable write once media
- Read with standard DNA sequencer
- Sustainable rugged solution
- Lowest long term TCO

Library

- Data Center ready solution
- Fully automated system with standard interface
- Integrates with existing storage applications
- Highest volumetric storage capacity
- Exceeds conventional data longevity capability
- Operated by IT team
- STaaS deployment
- Lower long term TCO

DNA Data Storage Solutions – Status

Vault

- Sampling today in MB scale
- GB scale pilots soon
- TB scale will follow
- Currently only available for select pilot customers
- Looking for innovative early-adopters, customers that will help shape the product

Library

- Requirements and design phase
- Estimated availability in several years
- Open for technology development collaboration
- Looking for innovative early-adopters, customers that will help shape the product

DNA Data Storage – Library: Concept System Requirements

- System outline:
  - Granular storage: capsules / tray
  - Data maps logically to physical location
  - Standard data center environmental conditions
  - System components field serviceable / replaceable
  - Maximizes DNA volumetric storage density
  - Maximizes write parallelism for throughput

Leveraging tape ecosystem key to achieving fastest TTM!

DNA Data Storage – Key Technology Enablers

Synthesis	TB scale TB per day Water-based
Storage/Retrieval	PB scale Automation Easy copy & store
Sequencing	TB scale TB per day Non-destructive
System	Data Center Ready Software integrated Object Storage APIs

DNA Data Storage
Customer Pilots

Twist DNA Data Storage Pilots

- Archiving example use cases:
  - Movie series, videos, images, performances, ancient & important documents and manuscripts
  - Artwork, NFT art, crypto currency, scripts, museum collection, national anthem
  - Human race and individual legacy preservation

NETFLIX

unicef

Montreux Jazz Festival

Microsoft

W

HARDDISKMUSEUM

NFSA
NATIONAL FILM AND SOUND ARCHIVE OF AUSTRALIA

LifeShip

LUARTS

BEYOND EARTH

BEELD & GELUID

SICPA

GENOMIKA

Yale University Library

TEMPLE UNIVERSITY
Beasley School of Law

DNA Data Storage Ecosystem

Building the DNA Data Storage Ecosystem

DNA Data Storage Alliance recently became a SNIA Technology Affiliate, with dedicated charter and P&P

History

- Formed in October 2020 by Illumina, Microsoft, Twist and Western Digital
- More than 50 member organizations across the entire eco system

Mission

- Create and promote an interoperable storage ecosystem based on DNA as a data storage medium

Scope

- Educate the DNA data storage market to create awareness and adoption
- Identify use cases in various markets/industries for the use of DNA data storage
- Develop an industry technology roadmap for DNA data storage
- Develop standards or specifications as needed by ecosystem

DNA DATA STORAGE ALLIANCE

https://dnastoragealliance.org/

DNA Data Storage
Conclusion

DNA Data Storage – What it is and What it is not

DNA is not…

- Storing all the world's data in a shoebox
- Coming to a DC nearby in the next 2 years
- A hot/warm storage medium
- Inexpensive to write (yet)

DNA is…

- A new, complementary cold layer in the storage pyramid
- An ideal medium for an offline copy and media diversity
- A medium lasting 100+ years in the right packaging
- Always readable, for as long as humanity reads DNA
- Eliminating migration; minimal maintenance, energy use
- Broadening the archive storage media choices available
- Offering the lowest long term TCO