SNIA DEVELOPER CONFERENCE

SDC 24

BY Developers FOR Developers

September 16-18, 2024
Santa Clara, CA

# DNA Data Storage

## An End-To-End System Concept

Olivier Lauvray, Biomemory

Shruti Sethi, Microsoft

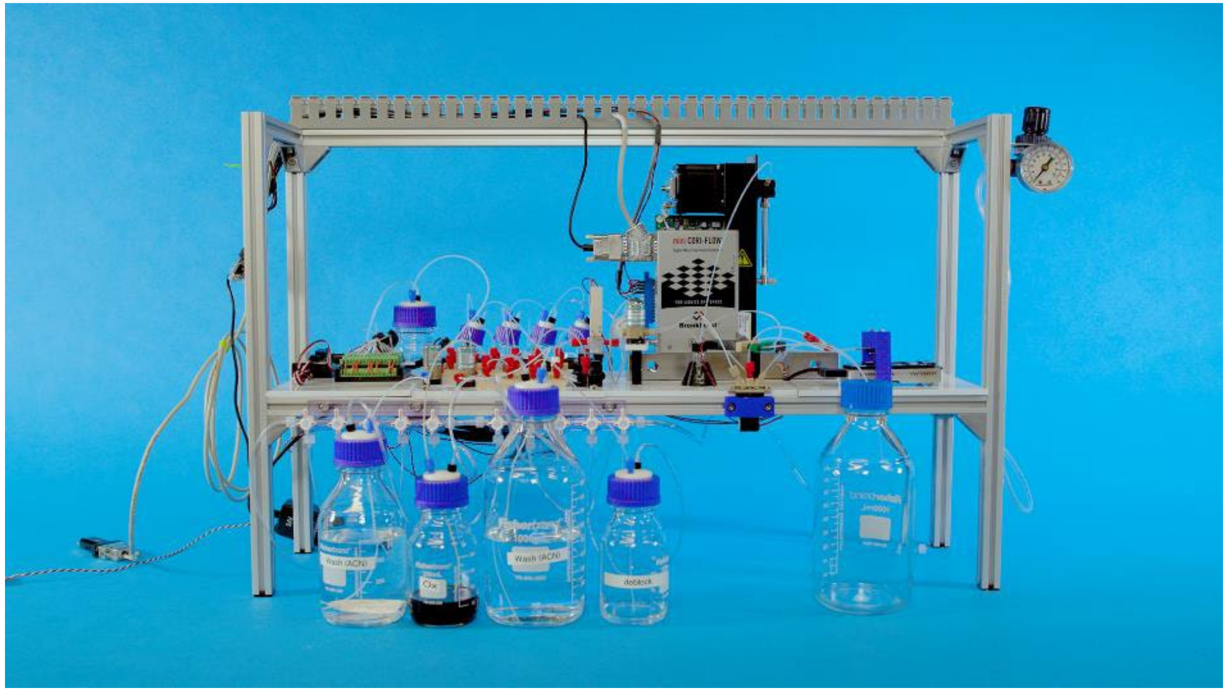
# Agenda

- From Lab to Datacenter Use Case
- Abstract Concept of a DNA Data Storage System
  - Defining an Interoperability Standard
  - Logical interfaces
  - System Operations
  - Type of DSS Racks
- System Implementation
  - Strategic Enablers
  - Integration in Datacenter
  - System Performance
  - Current Status

# DNA in the Datacenter

From here



To here

COST

SPEED

SCALABILITY

RELIABILITY

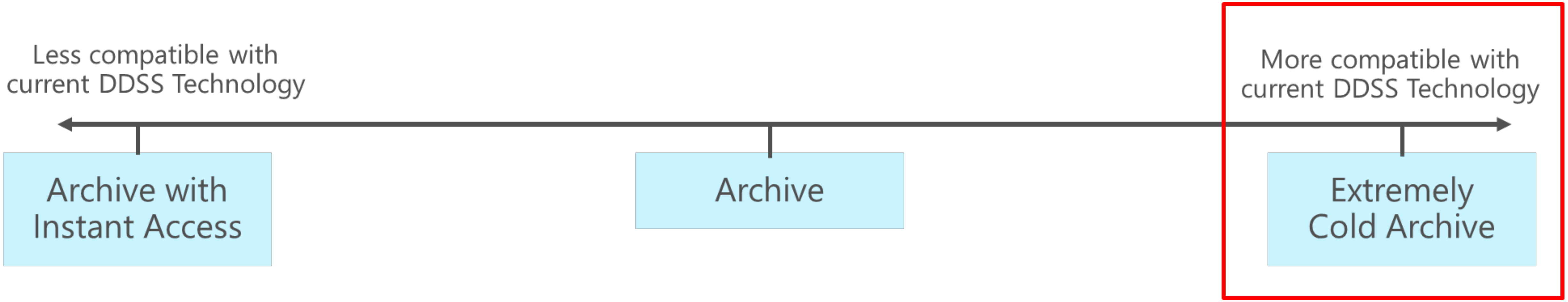
Interoperability



DNA DATA STORAGE ALLIANCE

SDC 24

# Datacenter Use Cases



DNA DATA STORAGE ALLIANCE

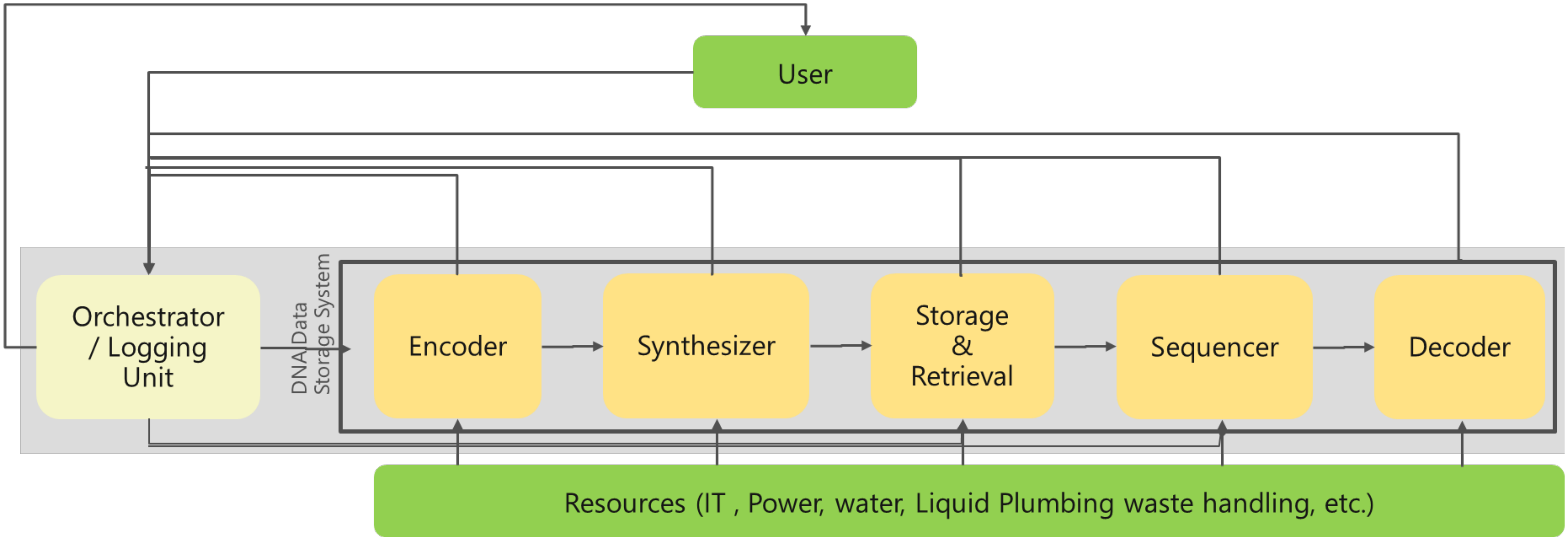
# Concept of a DNA Data Storage System

## Defining an Interoperability Standard

# End-To-End DNA Data Storage System (DDSS)

**DNA Data Storage System (DDSS)**

Sub-Blocks: Encoder, Synthesizer, Storage & Retrieval, Sequencer, Decoder

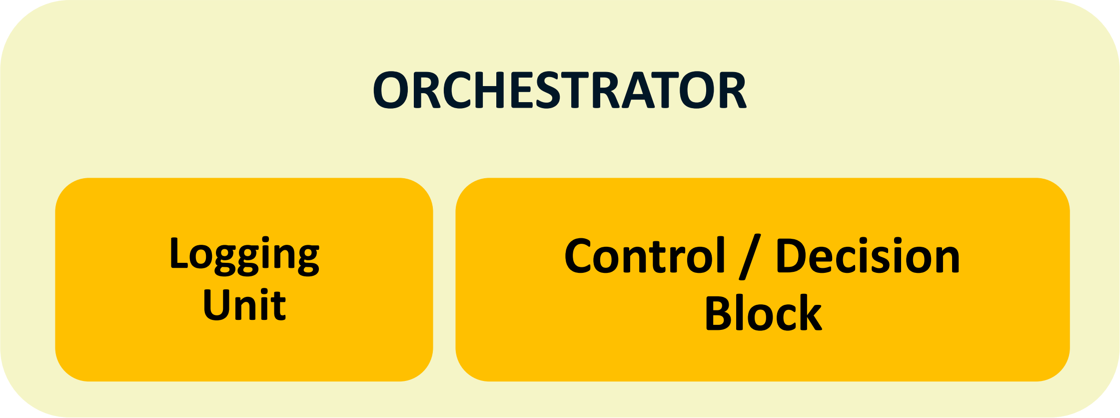


DNA DATA STORAGE ALLIANCE

SDC 24

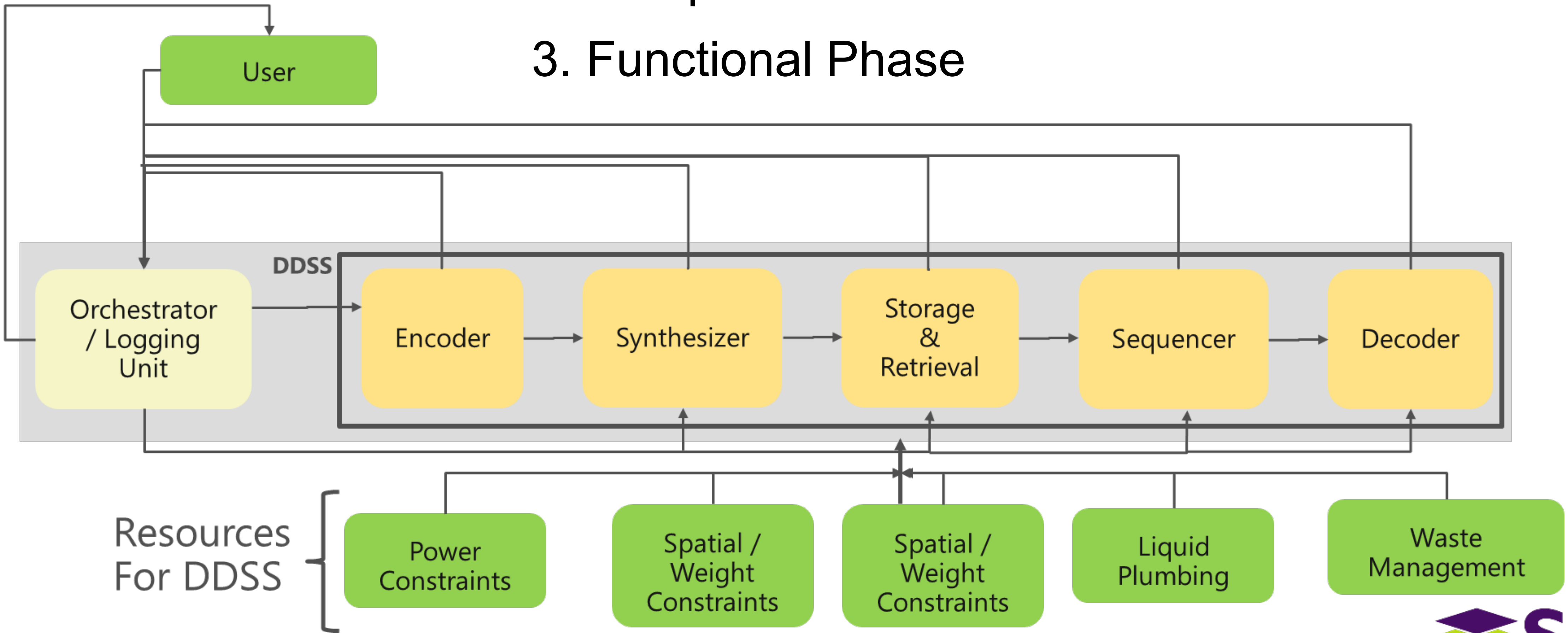
# End-To-End System : Orchestrator




- Main brain controlling the DDSS Setup & Functional Phases
- Channel through which all writes & reads are issued and read-data returned
- Tracks & logs the status and error feedback from each DDSS Sub-block. Logging used for decision making - Retry / Alarm decisions
- Monitors Accumulated Error at the System Level

# Defining an Interoperability Standard: **Phases of Operation**

1. System Assembly Phase
2. Setup / Initialization Phase
3. Functional Phase

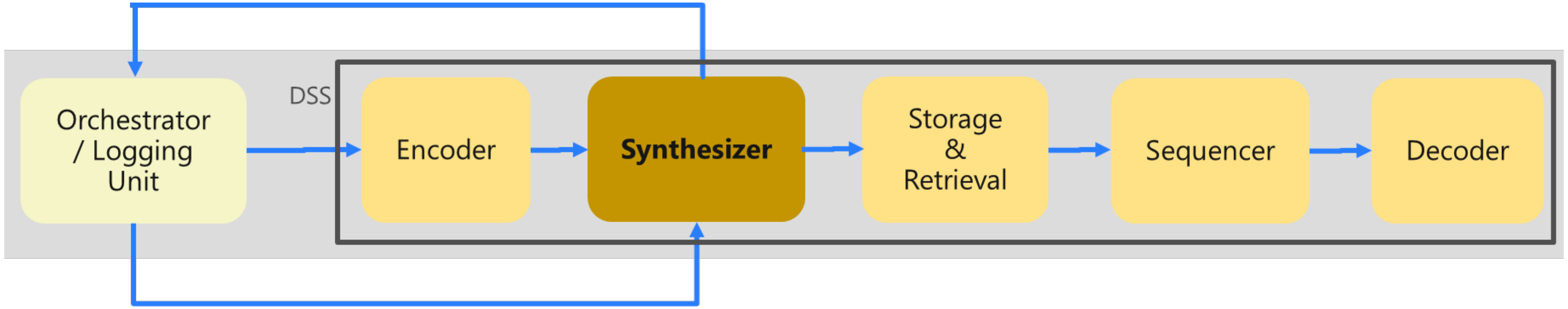


DNA DATA STORAGE ALLIANCE

SDC 24

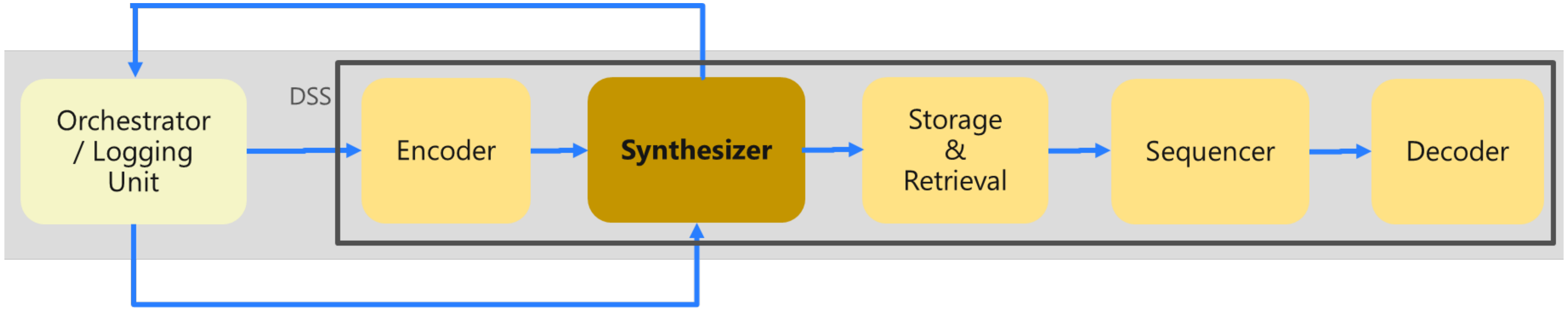
# Logical Interfaces - Functional Interface for “Synthesizer”




## Input Interface

| Data | Description | Source - Destination |
|---|---|---|
| Encoded Header | Name, Identification of customer, Metadata encoded into ATCG format as per CODEC (Primers included) | Encoder -> Synth |
| Encoded Addressing | Address encoded into ATCG format as per CODEC (Primers included) | Encoder → Synth |
| Encoded Customer Data | Customer Data encoded into ATCG (DNA) format as per CODEC (Primers included) | Encoder → Synth |
| Retry Signal | Orchestrator may issue a Retry signal in case of irreparable Error received from Synthesizer or Storage blocks | Orchestrator -> Synth |

# Logical Interfaces - Functional Interface for “Synthesizer”




## Output Interface

| Output | Description | Source - Destination |
|---|---|---|
| DNA Sequences/Strands | Sequences containing Address, Metadata and Customer Data | Synth -> Storage |
| Status metrics | Cycle Done, Key metrics | Synth -> Orchestrator |
| Flag: Est. Error Rate > Spec (optional) | Error flag that fires when the actual error rate during synthesis is higher than the Specification Error Rate of the synthesizer | Synth -> Orchestrator |
| Flag: Strand Length (optional) | Strand Length check using Bioanalyzer or Qubit post synthesis - for sample | Synth -> Orchestrator |
| Flag - Generic | General Error flag for issue in flow/control, reagent levels that may be detected via monitoring. | Synth -> Orchestrator |
| Quality Check (optional) | Quality Check can be executed using Sequencer, Bioanalyzer, or Qubit | Synth -> Orchestrator |
| Reagent Levels | Report Current Reagent Levels in terms of Number of Runs possible | Synth -> orchestrator |

# Defining an Interoperability Standard: **System Operations**

- **Write / Read**
- **Modify**
- **Erase / Delete**
  - Ability to delete part / all the previously written data. Implemented as:
    - ‘Erase’ at container granularity
    - ‘Erase’ partial data from a container
- **Self-Check (optional)**
  - During system IDLE, a known golden data file is written & read and checked for error rate.

Join us to work on the standards

@SNIA: www.snia.org/groups/snia-dna-technology-affiliate dnastoragealliance.org

@email: info@dnastoragealliance.org

# DNA Data Storage Racks

Defining an Interoperability Standard

# Type of DSS Racks

**Traditional Rack**

All 5 sub-components of the DDSS, along with the power, plumbing and management accessories needed, fit into a traditional Rack by specification

**‘N’ Racks act as Single Rack Unit**

The 5 sub-components of the DDSS may occupy N non-identical racks of standard power, weight and spatial specs.  This set of racks acts as a single unit and racks deployed only in multiples of N.

**Disaggregated Rack Unit**

The 5 sub-components of the DDSS occupy N non-identical racks. The system does not need to be deployed in multiples of N racks. A sub-portion of this Rack Unit can be scaled as required. Minimum N Racks footprint needed

# System Implementation

Example of the Biomemory DNA Data Storage appliance

# Pioneering DNA in Data Centers

**Rackable Enclosure Units**

**Consumables Cartridges**

**DNA Card Storage Containers**

**Storage Appliance**

✓ Store data in DNA for 150+ years @ room temp, with no energy consumption

✓ Suited for Data Centers

- Cost-effective
- Space-efficient
- Reliable
- Scalable
- Upgradable
- Resilient
- Interoperable
- Low-Power

***Disclaimer:***

*This presentation is for informational purpose only. It is meant to provide a preview of Biomemory's products, architecture, features and targeted performance, as **an example of implementation leveraging the recommendations from the DNA Data Storage Alliance.***

*It is provided as-is, without any expressed or implied warranty. The information in this presentation is not a commitment, promise or legal obligation of any kind from Biomemory.*

*Any feature, functionality or performance numbers are reflecting our best assessment at the date of the presentation. It may evolve in the course of our development and be modified, replaced or removed without notice in the final release of the products.*

# Strategic enablers

- **Low-cost** biosourced mass-production of non-hazardous biosafe DNA blocks used as consumables (low customer opex)

< 1000 times lower cost than life-science oligonucleotide synthesis technologies

- **Simple** assembly of a small number of DNA blocks, compatible with high-speed and massive parallelization.

Target > 1 billion reactions in parallel

- **Low error-rate** Write and Read processes, with efficient tolerance to errors (strong ECC)

90% error free process, with 100% error detection and recovery

- **Biotech adaptation to the system** (not the contrary)

- **Modular** software-driven system architecture aligned to Data Center requirements, with dedicated OAM channel

Compatible with existing Data Centers operations

- **Industrialization** and scalability, with focus on lowest risk, high-reliability and minimum custom hardware development

80% reuse and adaptation of industry-available components.

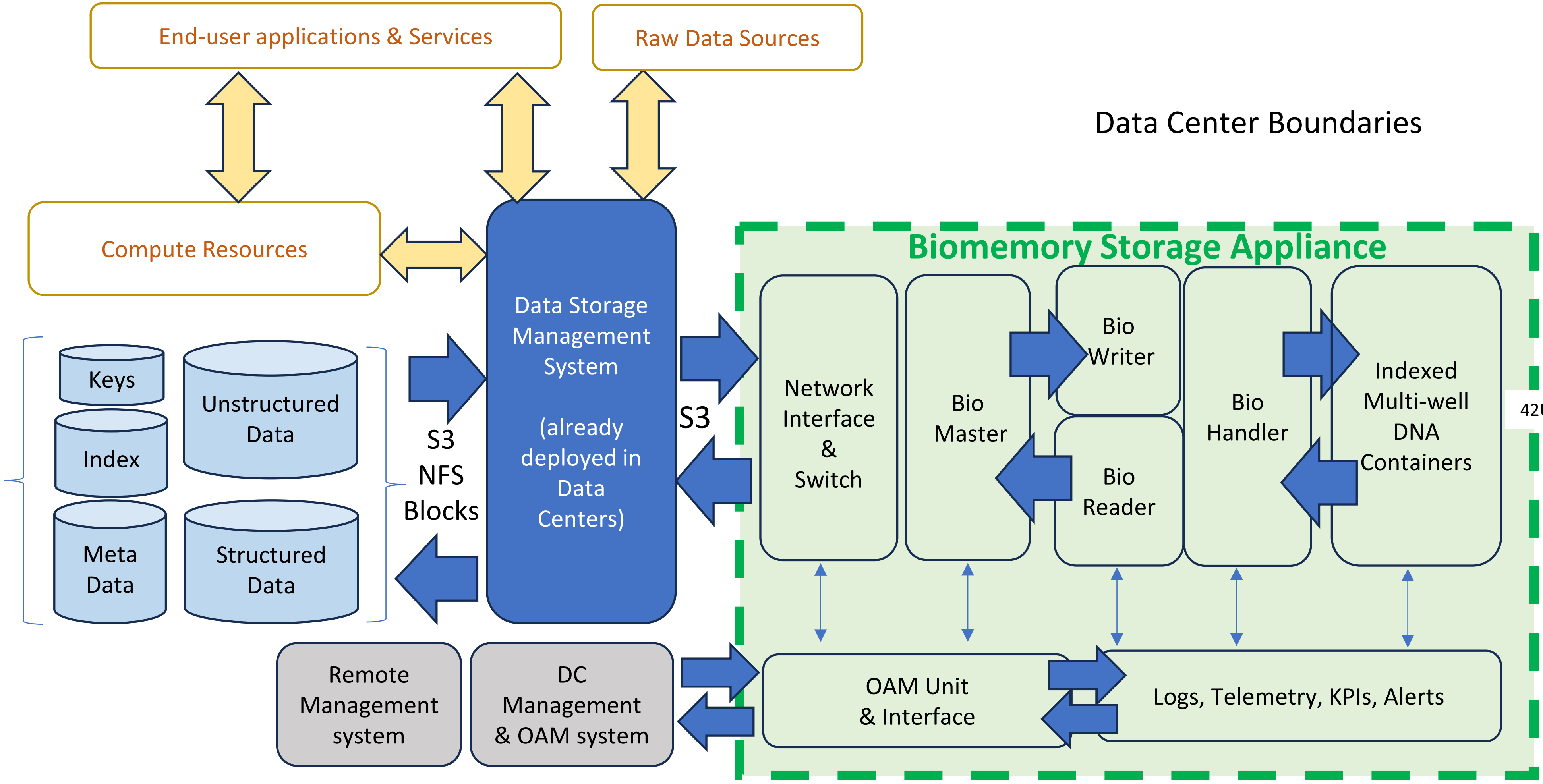
- **Aligned to SNIA DNA Data Storage Alliance** specifications and recommendations

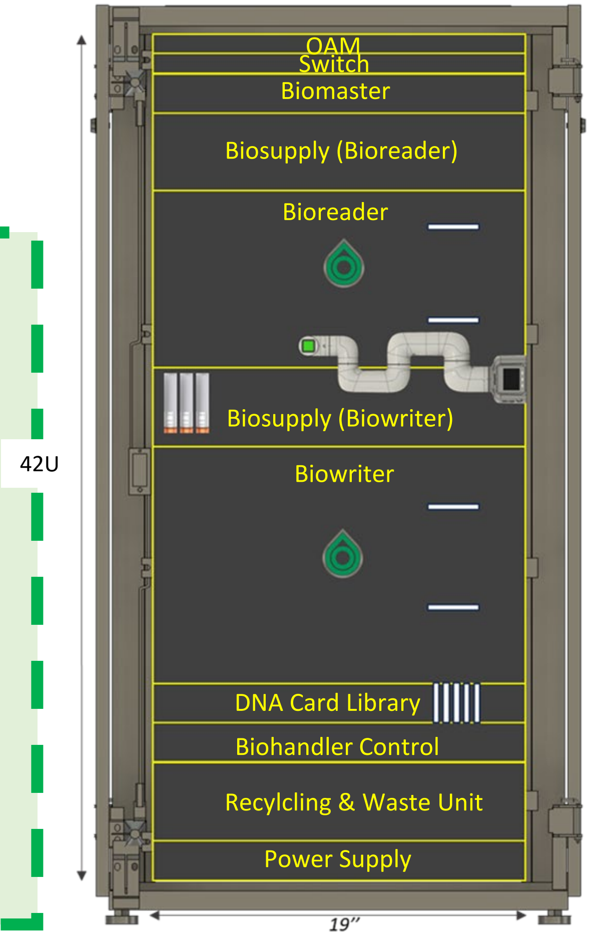
Interoperability, portability, reliability



DNA DATA STORAGE ALLIANCE

# Integration in Data Centers




 | 

DNA DATA STORAGE ALLIANCE

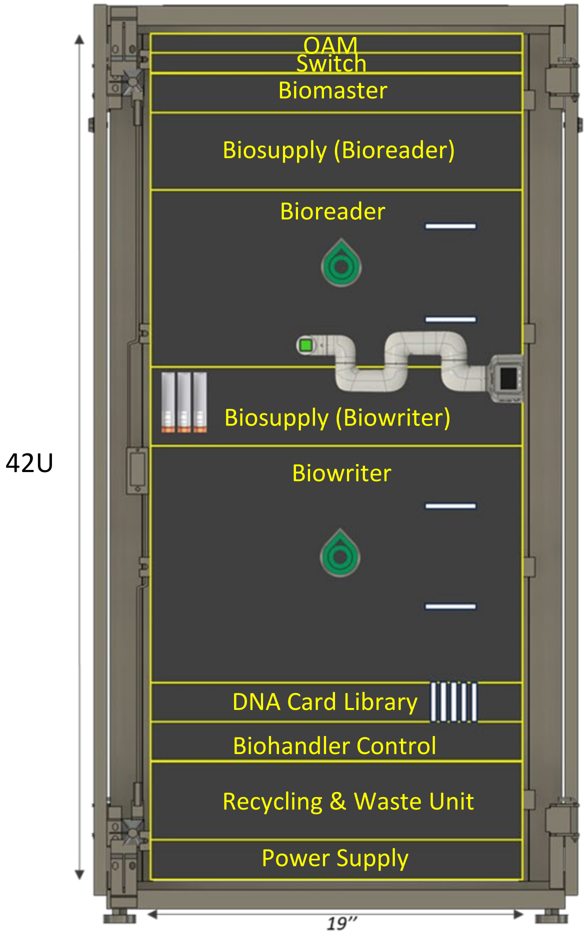
# System Implementation

Hardware & Fluidics

- 80+ % available or adapted from commercially-available components

Software / Firmware

- ~ 80 % to be developed or adapted




Biotech

- ~ 80 % available in Biomemory lab, ready to be industrialized and scaled up.
- Continuous roadmap to adapt and optimize the biotech components for Enterprise-grade Data Storage

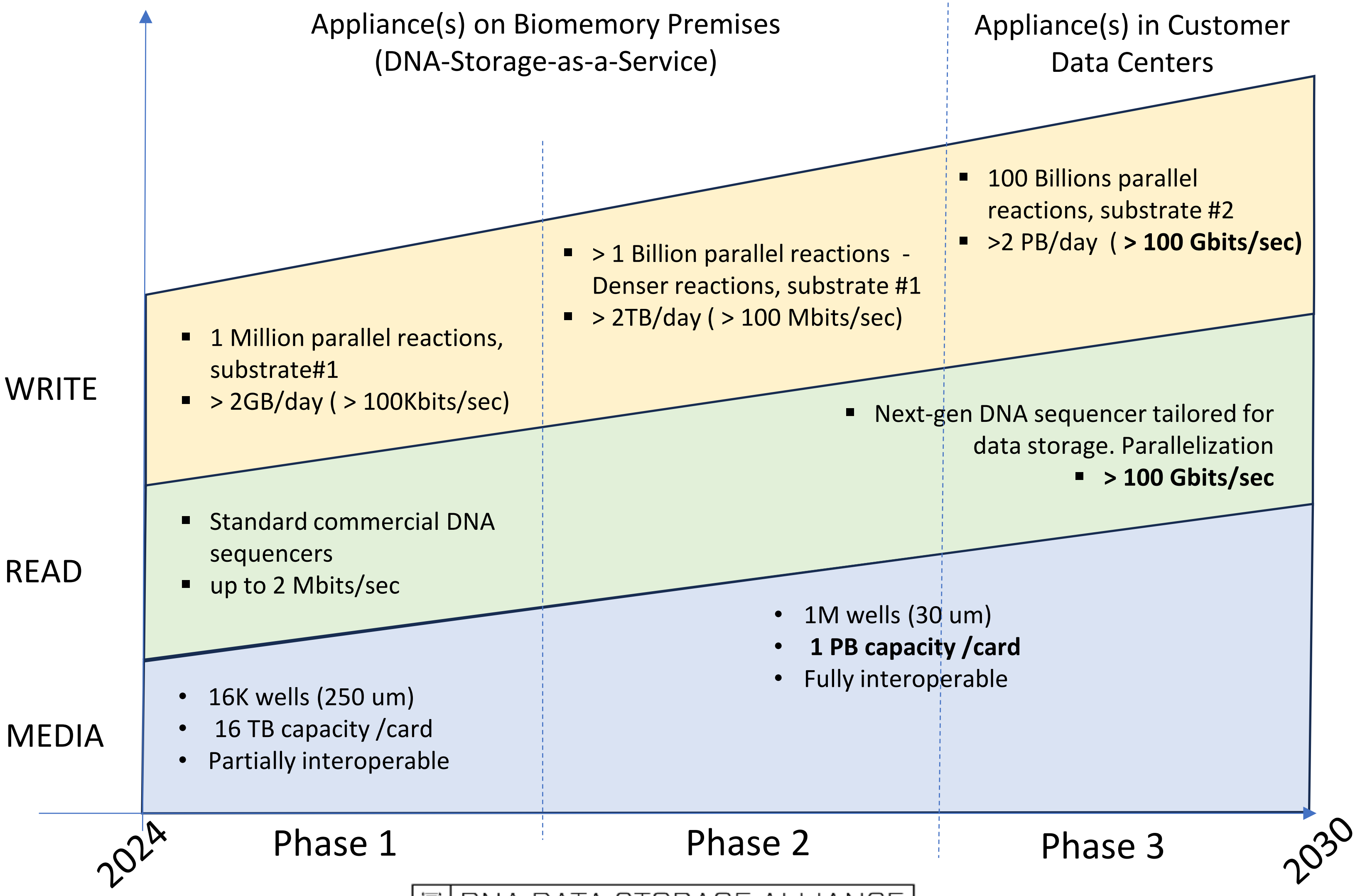
 

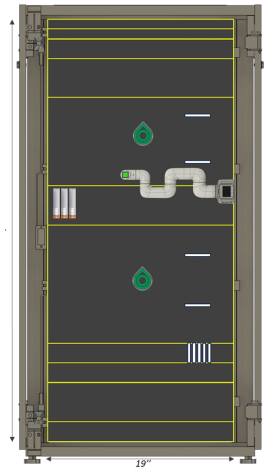
DNA DATA STORAGE ALLIANCE

SDC 24

# Clear path towards high-performance

Multi-rack system



Single-rack system

DNA DATA STORAGE ALLIANCE

# September 2024 status

**Miniaturization (Gen 1)**

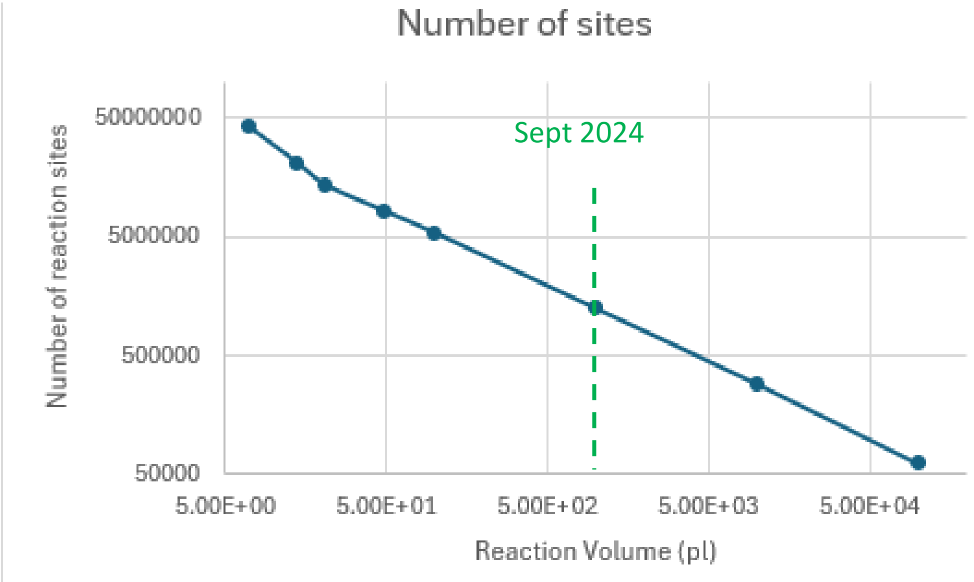
- DNA Assembly (Writing) with volume of 1nL.
- Validation = Writing / PCR / Reading cycle
- Enabling writing batches of ~1 Million sites (64 MB/batch)
- Moving towards 50 pL for ~10M sites/batch (640 MB/batch)

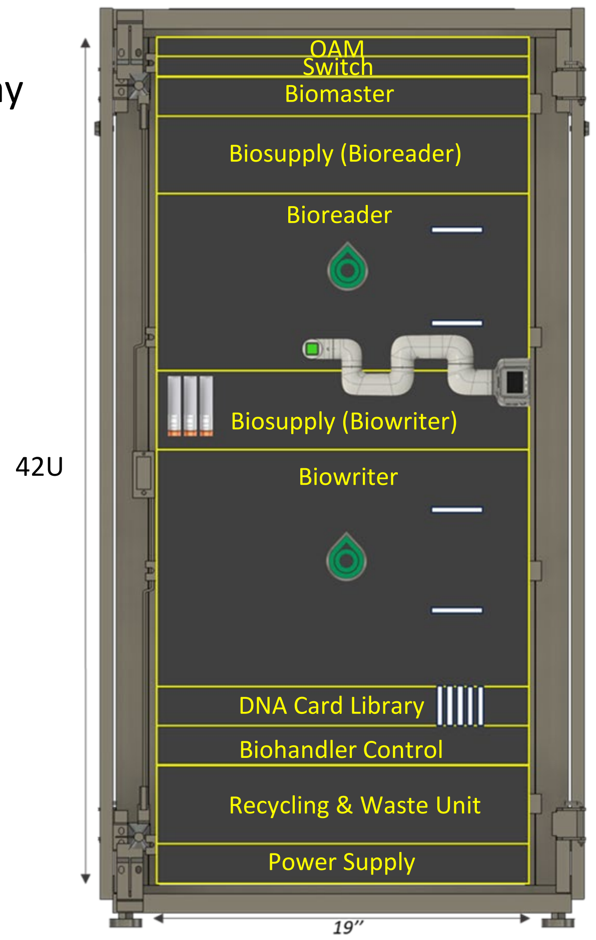
**Reaction Acceleration (Gen1)**

- Currently 1.3 day / batch
- Starting acceleration towards 3 batches/day (> 2GB/day by end of 2025)

**System Automation (Gen 1)**

- Biowriter system design completed
- Bioreader solutions under evaluation



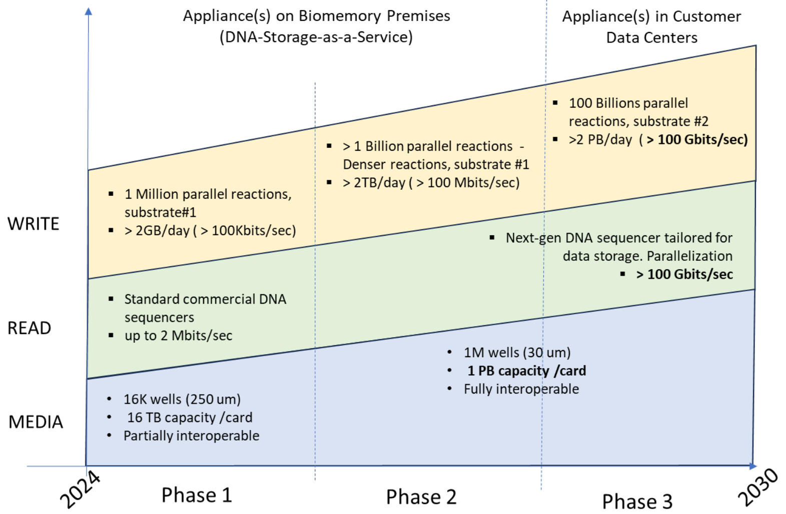


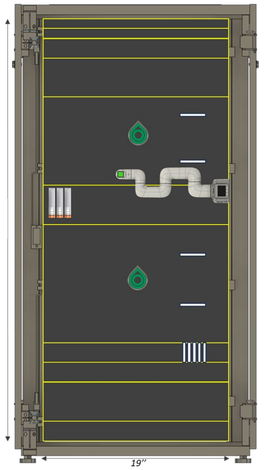

DNA DATA STORAGE ALLIANCE

# Going beyond cold storage

High-confidence to reach & exceed the 2030 industry **WRITE/STORE cost target of $1/TB**

With high-speed **READ exceeding 100 Gbit/sec**



2024 - 2025

2028 - 2030

**Frozen Reference**

Anti-counterfeiting, storage of boot code or keys, blockchain support

- NFT
- Military equipment
- Luxury goods
- Healthcare devices
- Pharma products
- Data in space
- …

**Extremely Cold storage**

From regulatory obligations to disaster recovery strategy

- Regulated activities
- Attorneys, notaries
- Administrative archives
- Critical software code
- Defense
- …

**Cold storage**

From conservatory to back-up, with fast access/recovery

- Libraries, museums,
- Government, agencies,
- Banking, insurance & finance
- Car or airplane OEMs,
- 3rd copy back-up
- …

**Warm storage**

Back-up, cloning and war chest raw data, pending AI readiness/analytics

- Cloud Storage Services
- Factory raw data for AI training
- Data Lakes for AI and Big Data
- IoT & rich media data
- Genomics data sets
- Logs & events
- …..

DNA DATA STORAGE ALLIANCE

# Thank you for your attention

Please take a moment to rate this session.

Your feedback is important to us.