DNA Data Storage System

An End-to-End System Concept

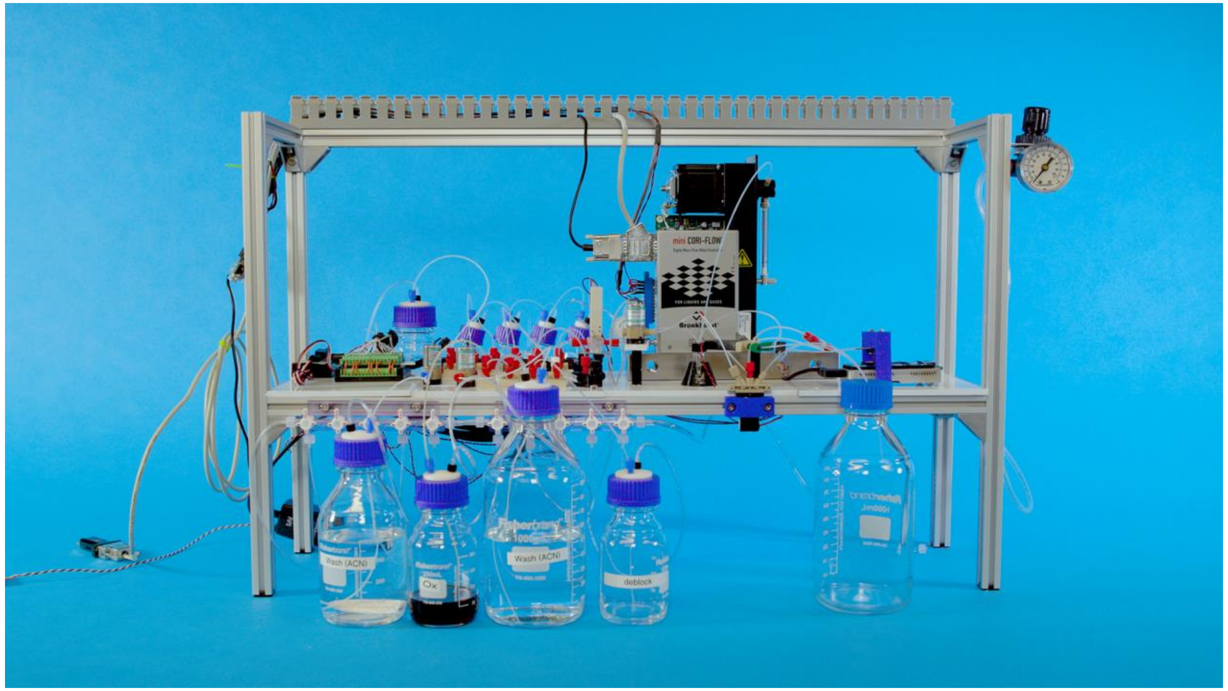
DNA DATA STORAGE ALLIANCE

Presenters:

- Shruti Sethi, Microsoft
- Olivier Lauvray, Biomemory
- David Landsman, Western Digital

DNA in the Datacenter

From here



To here

Key Influencers

COST

SPEED

SCALABILITY

RELIABILITY

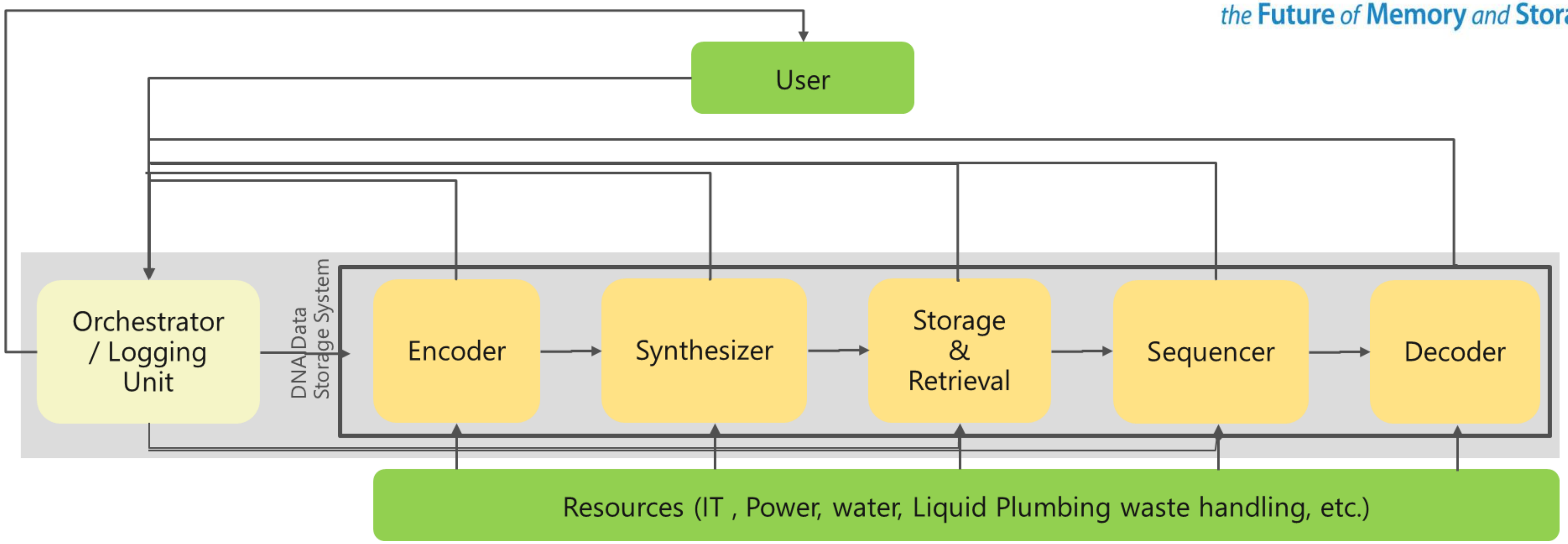
Interoperability

Defining an interoperability standard

the Future of Memory and Storage

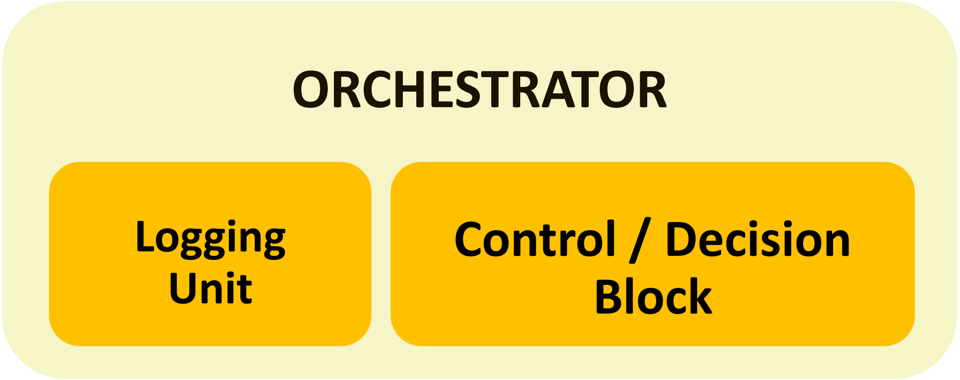
DNA Data Storage System (DDSS):

- Inputs/outputs for each phase



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

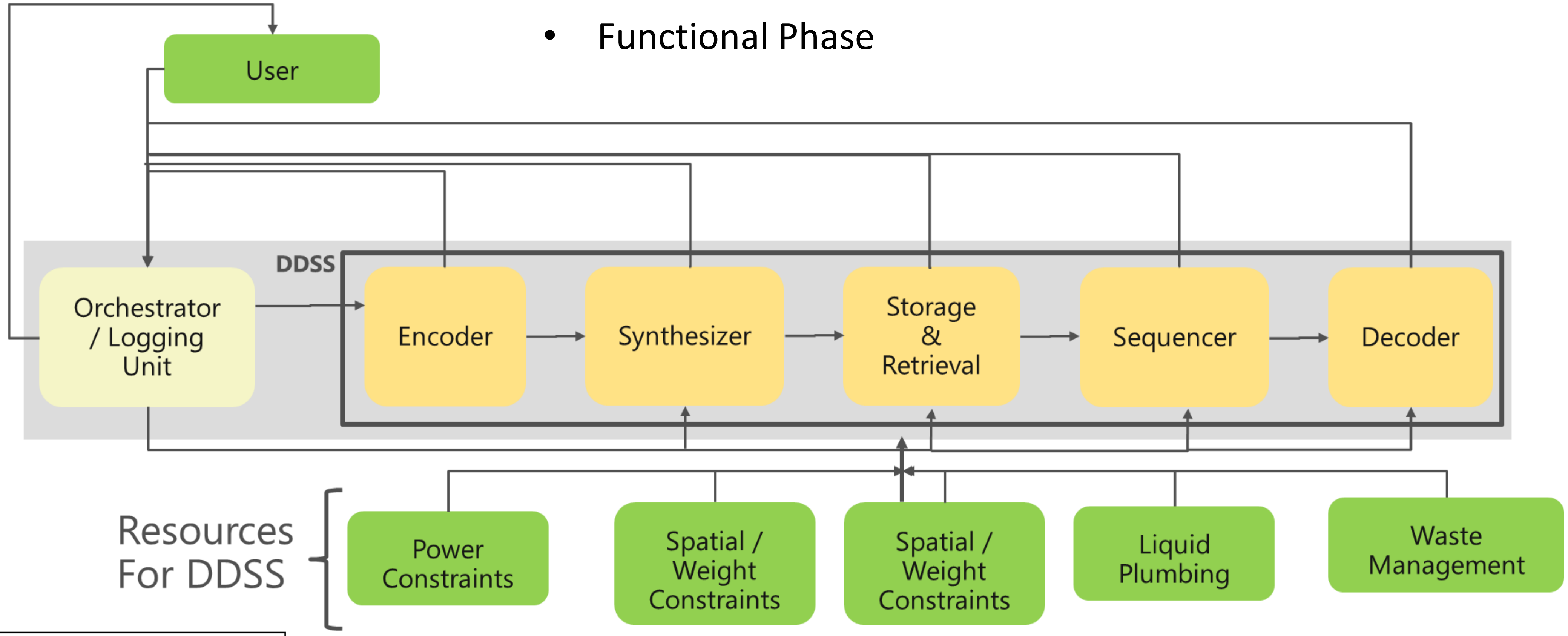


Defining an interoperability standard

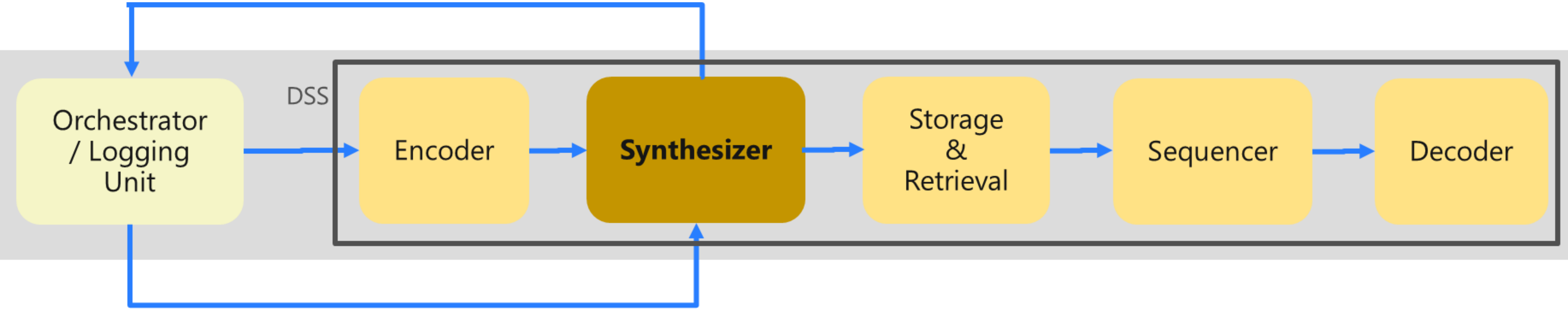
the Future of Memory and Storage

Phases of Operation:

- System Assembly Phase
- Setup / Initialization Phase
- Functional Phase



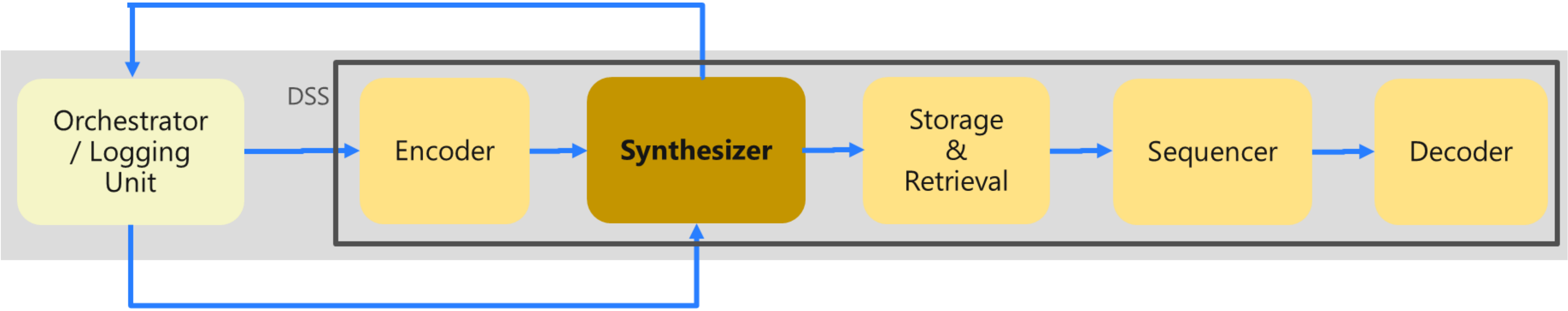
Example:
Functional Interface for “Synthesizer” - Inputs



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

Example:
Functional Interface for “Synthesizer” - Outputs

the Future of Memory and Storage



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

DDS Rack Types

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 Operations

- **Write, Read, Modify**
- **Erase / Delete**

  Ability to delete part / all the previously written data:
  - ‘Erase’ at container granularity
  - ‘Erase’ partial data from a contain

- **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

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

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

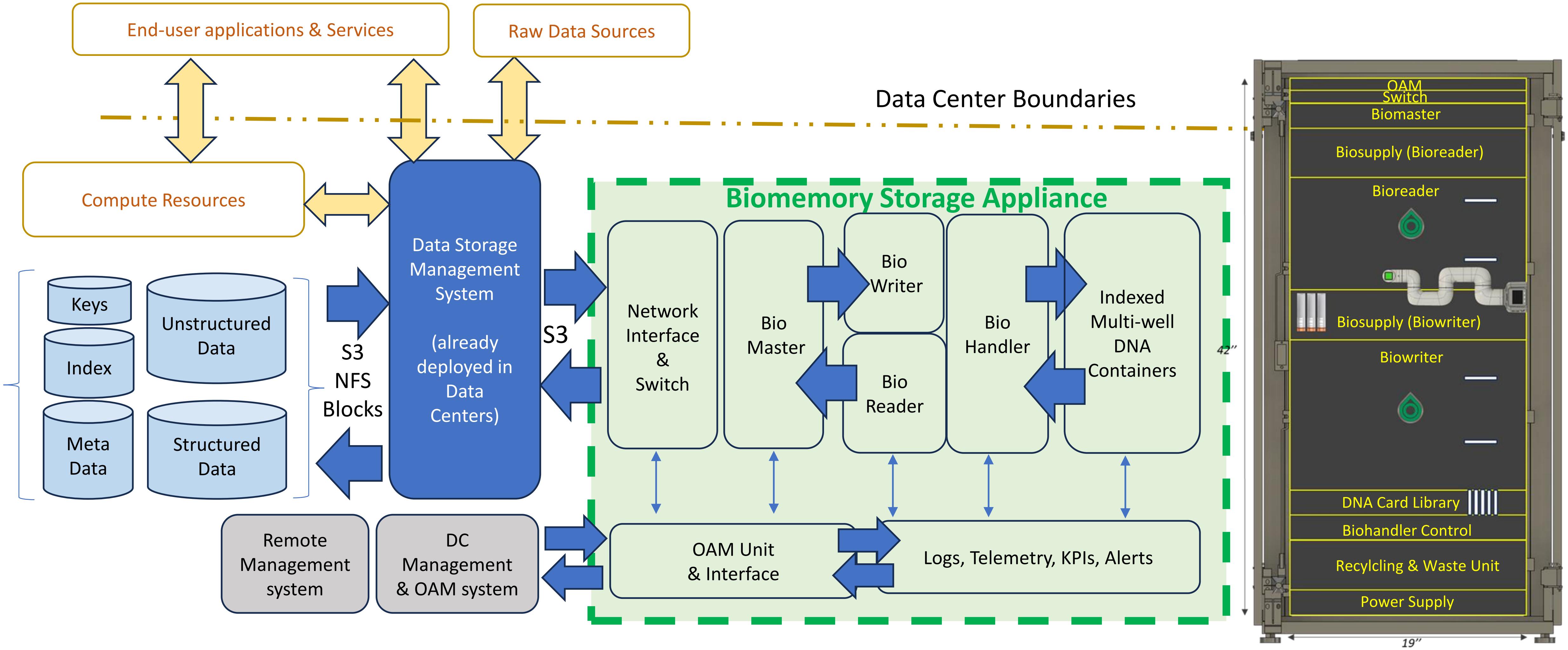
❖ **Industrialization** and scalability, with focus on lowest risk, high-reliability and minimum custom hardware development

80% reuse and adaptation of industry-available components.

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

Interoperability, portability, reliability

Integration in Data Centers



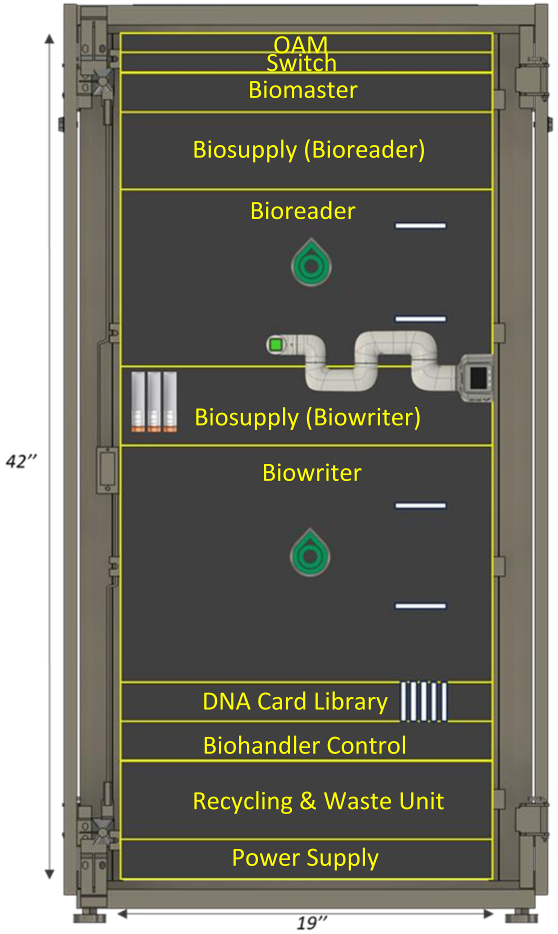
System Implementation

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

Clear path towards high-performance

Multi-rack system

Appliance(s) on Biomemory Premises (DNA-Storage-as-a-Service)

Appliance(s) in Customer Data Centers

WRITE

- Phase 1:
  - 1 Million parallel reactions, substrate#1
  - > 2GB/day (> 100Kbits/sec)
- Phase 2:
  - > 1 Billion parallel reactions - Denser reactions, substrate #1
  - > 2TB/day (> 100 Mbits/sec)
- Phase 3:
  - 100 Billions parallel reactions, substrate #2
  - >2 PB/day (**> 100 Gbits/sec)**

READ

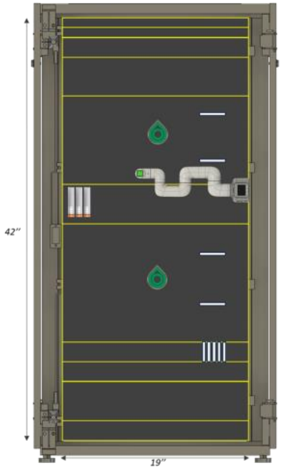
- Standard commercial DNA sequencers
- up to 2 Mbits/sec
- Next-gen DNA sequencer tailored for data storage. Parallelization
  - **> 100 Gbits/sec**

MEDIA

- 16K wells (250 um)
- 16 TB capacity /card
- Partially interoperable

- 1M wells (30 um)
- **1 PB capacity /card**
- Fully interoperable

2024 — Phase 1 — Phase 2 — Phase 3 — 2030

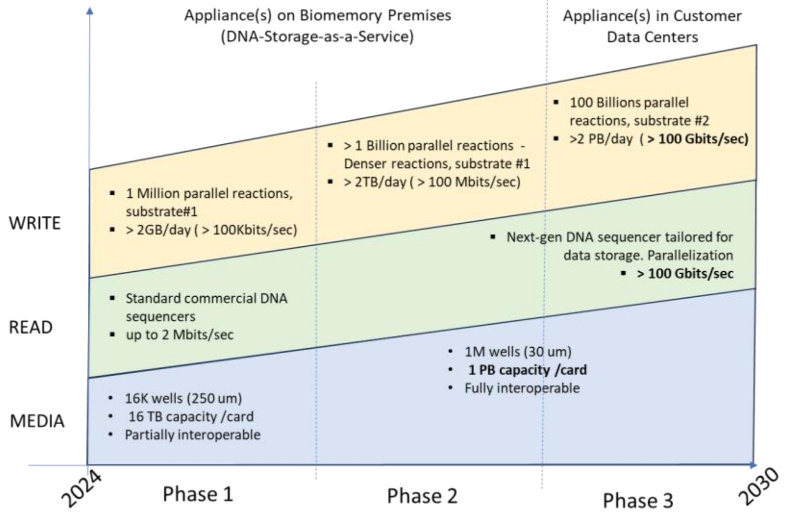


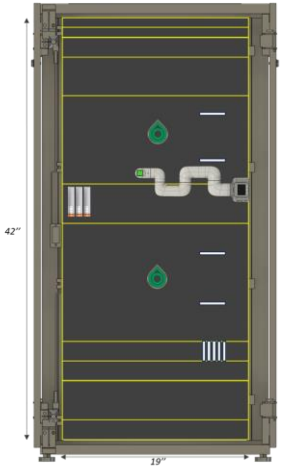
Single-rack system

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
- …..

Thank you for your attention

DNA Data Storage System

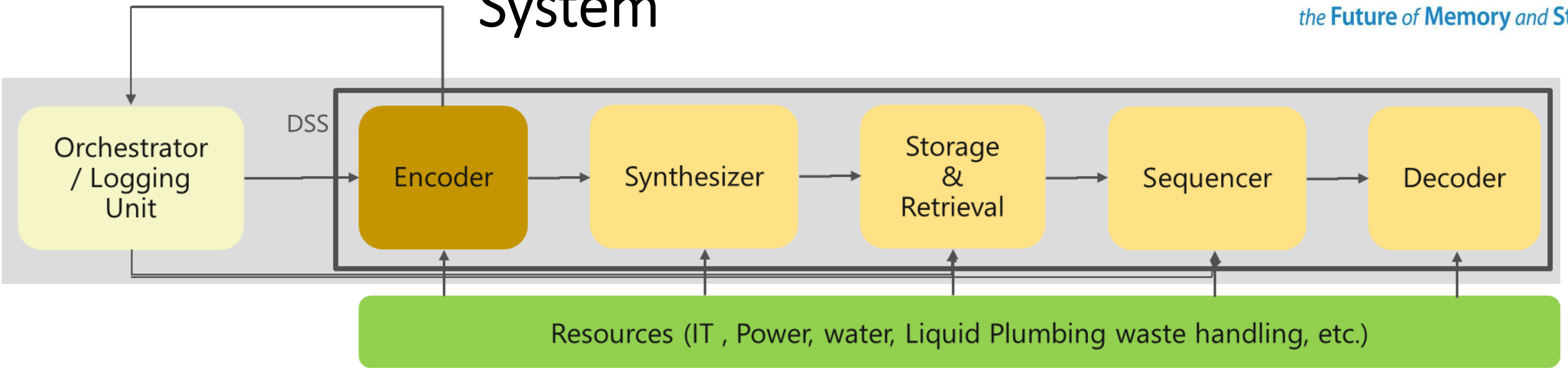
An End-to-End System Concept

Presenters:

- Shruti Sethi, Microsoft
- Olivier Lauvray, Biomemory
- David Landsman, Western Digital

BACKUP

Setup/Initialization for "Encoder" in an End-to-End System



SETUP

Data	Description	Source - Destination
CODEC & PARAMETERS	Refer to Sector 0 Spec for full Definition	Logging -> Encoder
Configuration	Encoder Config enabled - as per seen Synth Error Rate	
Granularity of Writes & Reads	Eg: 64K pages	Logging -> Encoder

Type of Operations Handled

- **Write**
- **Read**
- **Modify**
  - Implemented as ‘Read’ + ‘Re-write with modification’ + ‘Erase old’
- **Erase / Delete**
  - Ability to delete part / all the previously written data
  - This could be implemented as :
    - **‘Erase’ at container granularity** -> the entire container can be marked invalid and the DNA content flushed (Garbage Collection).
    - **‘Erase’ partial data from a container** -> it may require reading and re-synthesizing the data that has to be retained from the container. The initial container is then marked invalid and its entire DNA content flushed (Garbage Collection).
- **Self-Check (optional)**
  - During system IDLE, a known golden data file is written & read and checked for error rate.
  - Can be used by the Orchestrator for re-calibrating the Accumulated System Error, Sequencer Coverage, Encoder/Decoder parameters

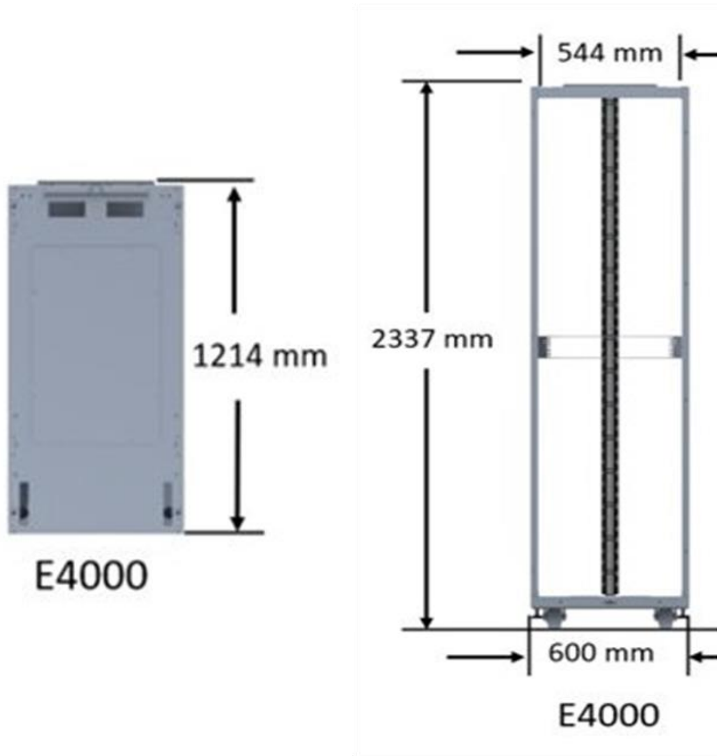
DDS Rack Types

(Power, Weight & Spatial (42OU) constraints exist per rack)

the Future of Memory and Storage

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 may be deployed only in multiples of N.

Disaggregated Rack Unit

The 5 sub-components of the DDSS may occupy N non-identical racks of standard power, weight and spatial specs. 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. However, the Rack Unit needs N minimum racks.