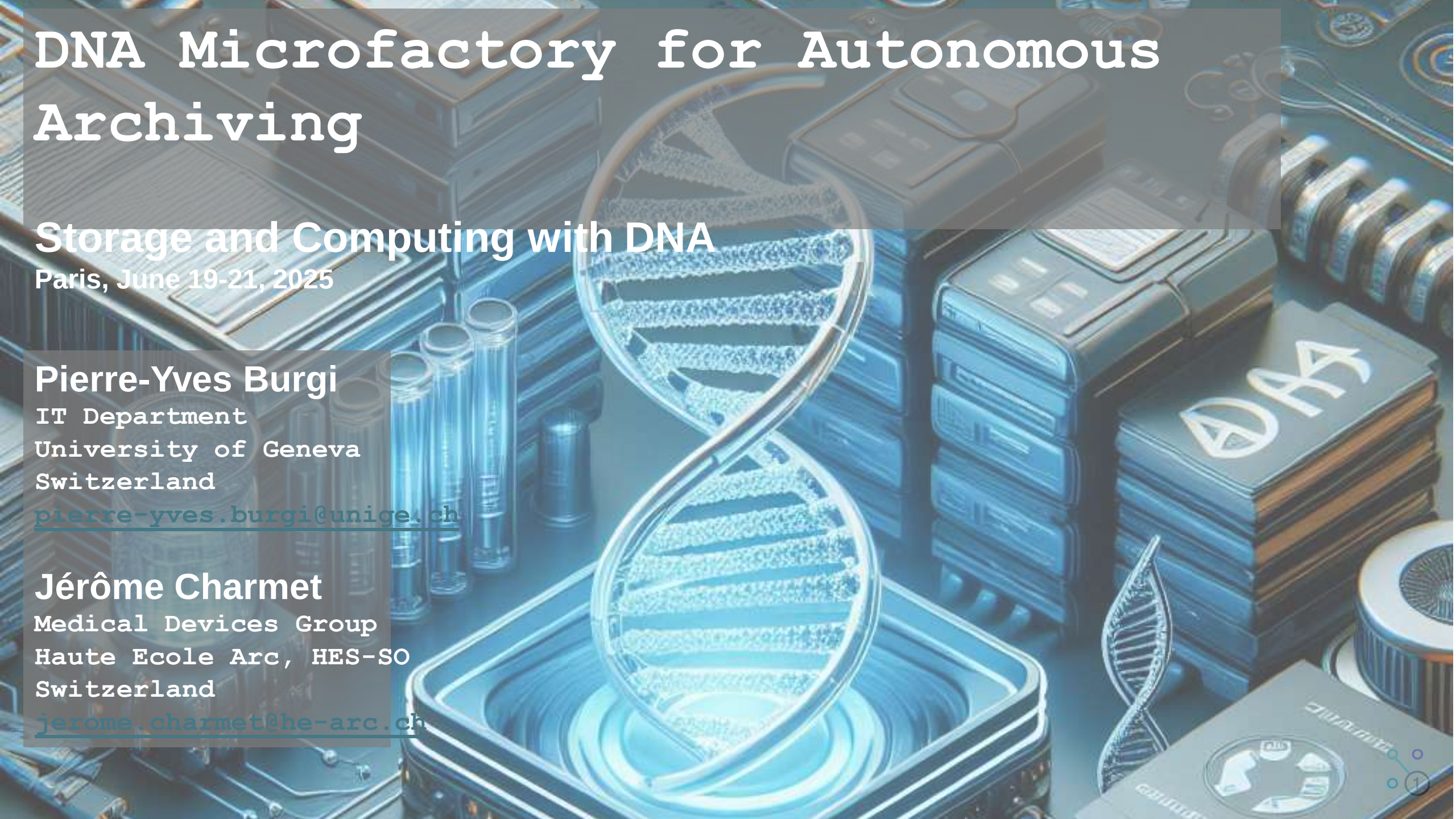
# DNA Microfactory for Autonomous Archiving



## Storage and Computing with DNA

Paris, June 19-21, 2025


**Pierre-Yves Burgi**
IT Department
University of Geneva
Switzerland
pierre-yves.burgi@unige.ch

**Jérôme Charmet**
Medical Devices Group
Haute Ecole Arc, HES-SO
Switzerland
jerome.charmet@he-arc.ch

# DNAMIC.org


GA no. 101115389
European Path Finder challenge for DNA-based digital data storage

Genomika

kaunas university of technology
ktu
1922

TECHNISCHE UNIVERSITÄT MÜNCHEN

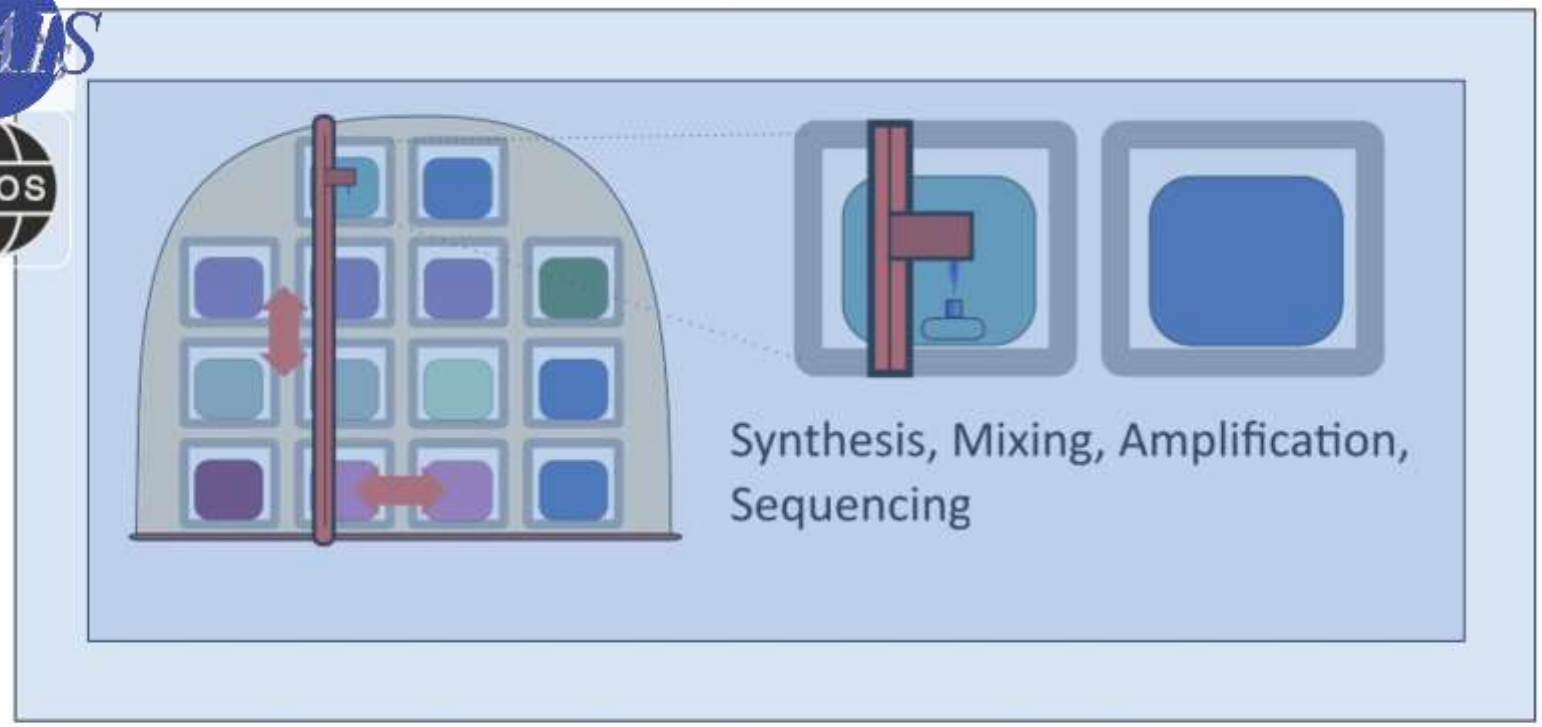
kilobaser

UNIVERSITÉ DE GENÈVE

Imperial College London

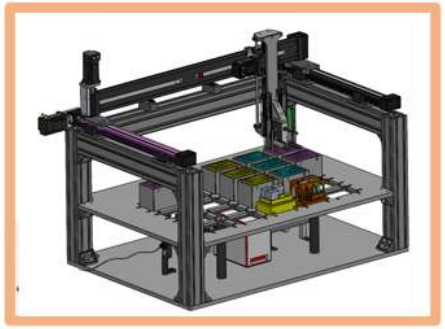
Hes·so
haute école
neuchâtel berne jura
arc

DNAMIC




OIOS

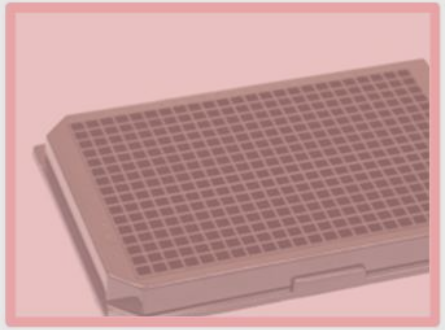
Wet Bench 1



Wet Bench 2

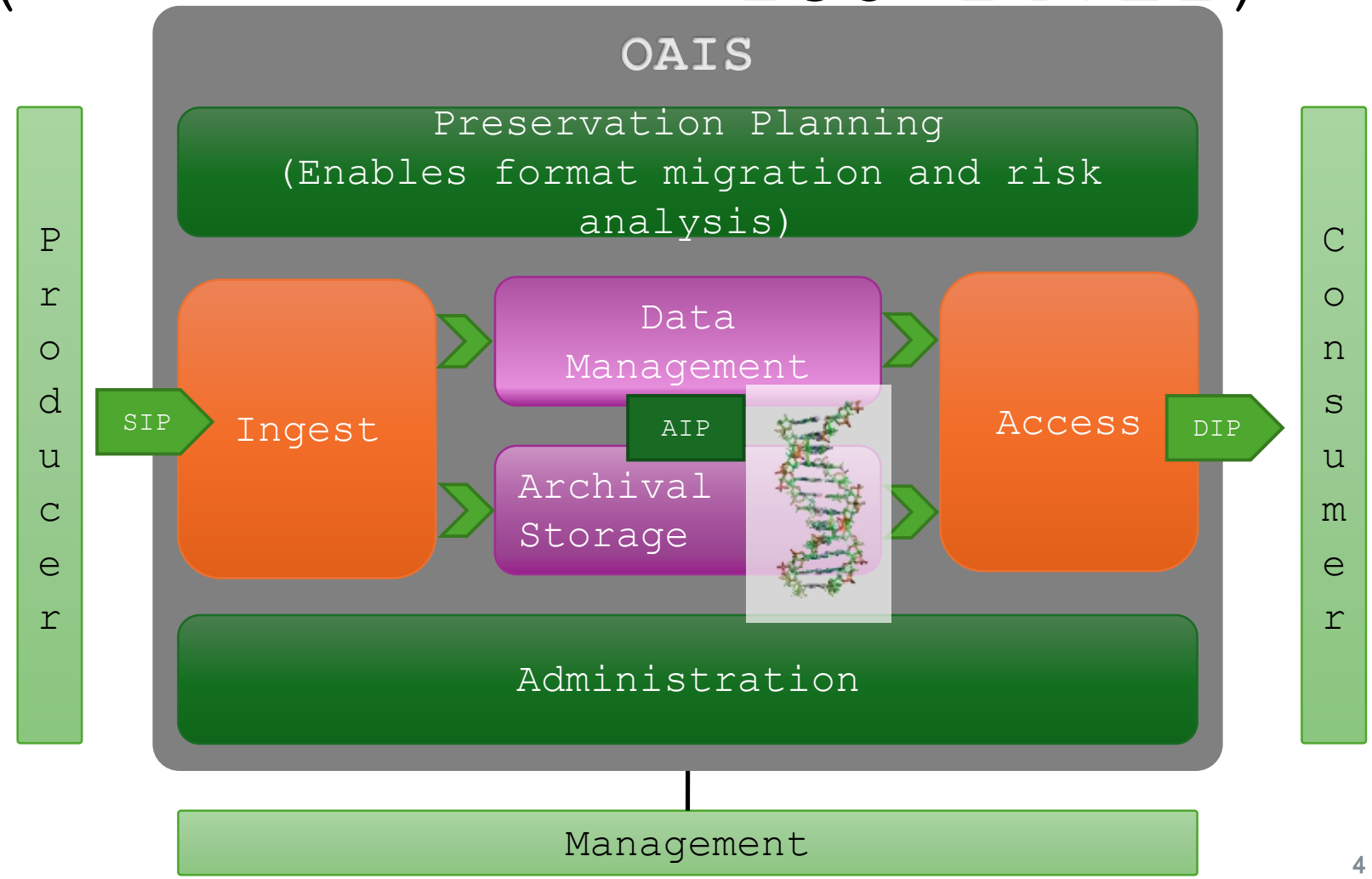
Storage

Physical archive



Not yet implemented

# The Open Archival Information System standard (OAIS – ISO 14721)



# Key features of the OAIS Standard

The Archival Information Package (AIP) includes:

- **Content Information** (the actual data)
- **Preservation Description Information** :
  - Provenance
  - Fixity
  - Context
- **Representation Information :**
  - Syntax (e.g., file format: TIFF, PDF/A, etc.)
  - Semantics (e.g., what a “record” or “field” means)
  - Decoding parameters (e.g., for physical media: bit density, signal modulation, byte alignment) --> CODEC parameters
  - Software tools, format registries, or emulators

# Self-Description Principle

- In the context of OAIS, **self-description** refers to the archive's ability to **describe its own contents, structure, and services** in a way that is understandable and usable by both:
  - **A Designated Community**
  - **Future systems or custodians**, even after decades of technological change
- This enables disaster recovery based on **Representation Information**

# Self-description is dependent on:

- The **complexity of describing**: it it measured by the difficulty associated with extracting the description of a system
- The **complexity of interpreting**: it is measured by the difficulty associated with extracting the interpretation (meaning) of a description

A DNA CODEC should be conceived so as **to minimize both complexities :**

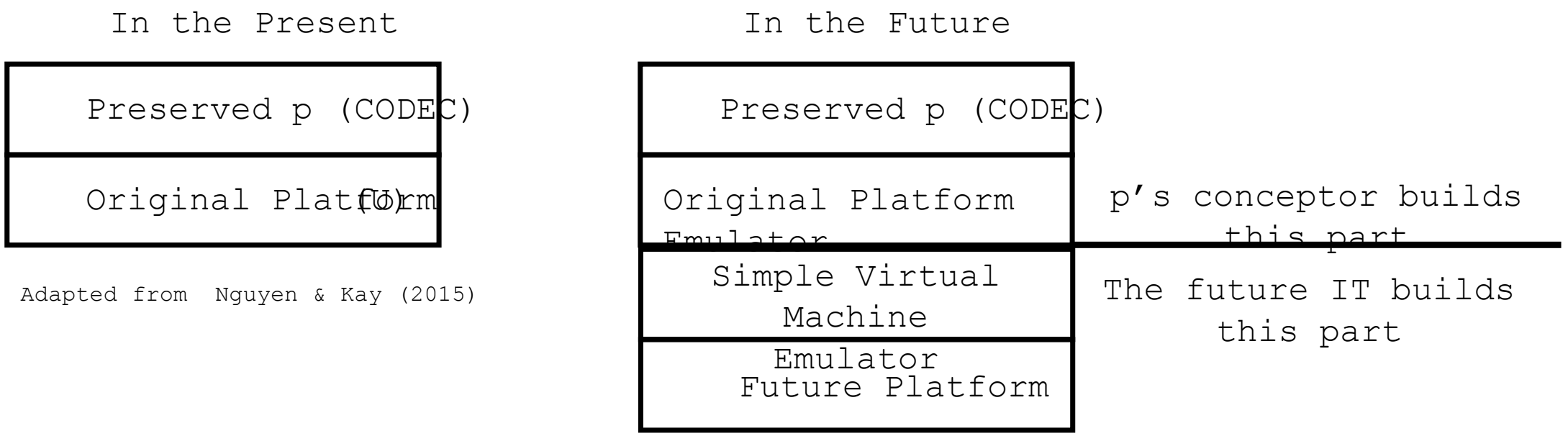
- To ensure a compact description into DNA (or other sustainable material, e.g. micro-film)
- To ensure our future generations will be able to decode the information

The algorithmic complexity can be measure based on **Kolmogorov complexity** K(x):

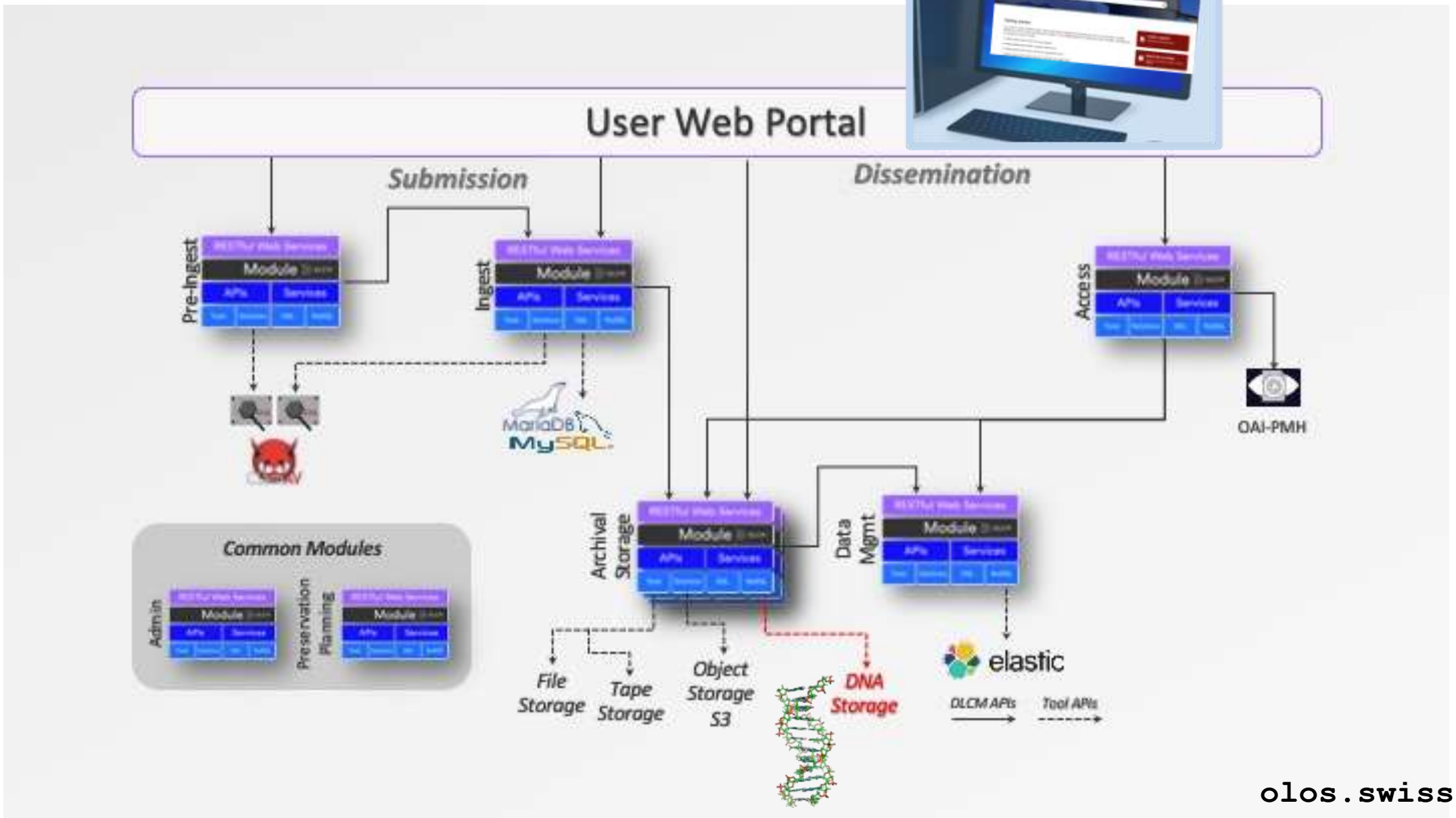
$$K(x)=\min\{len(p):U(p)=x\}$$

where

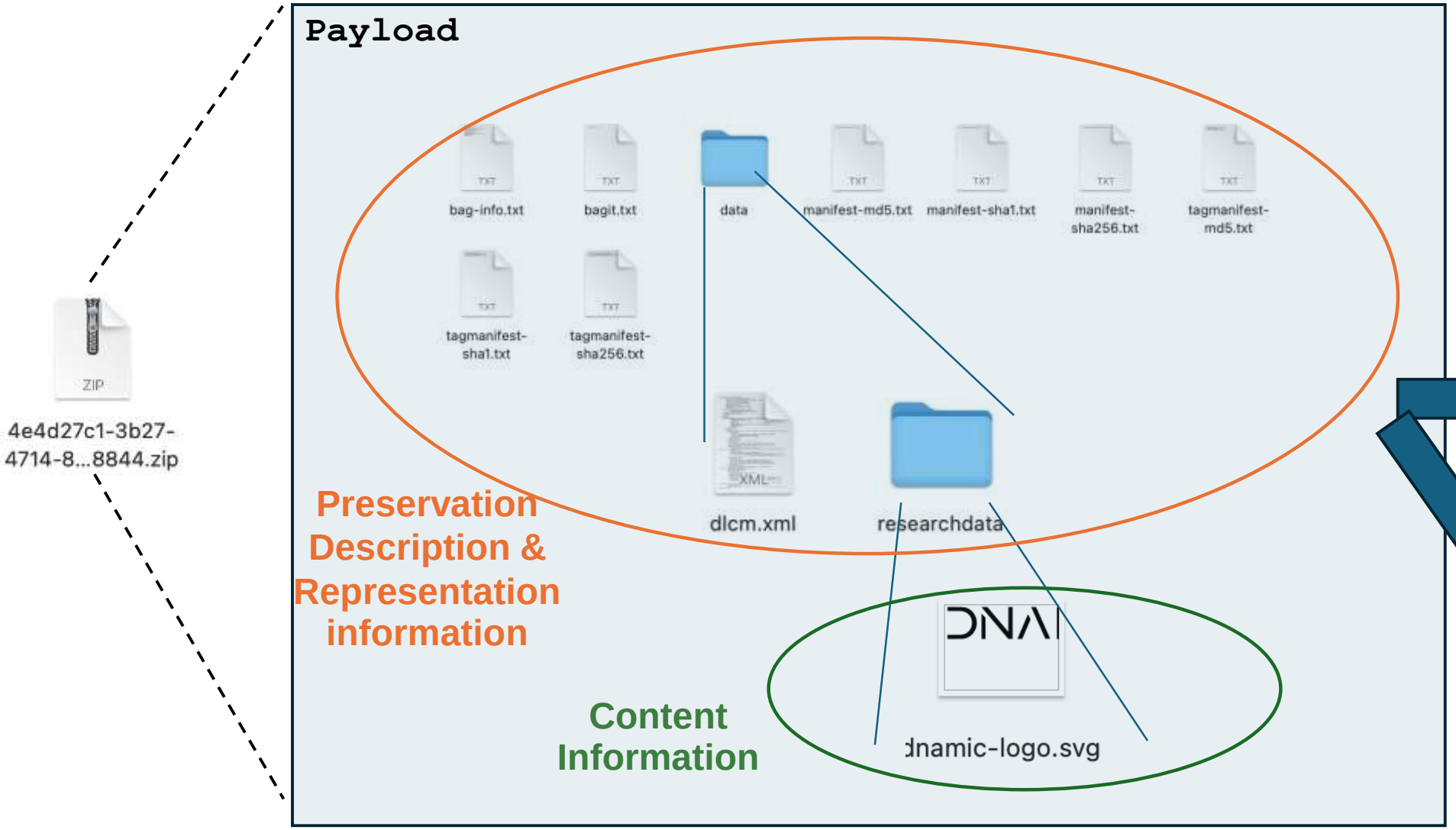
- U is a **universal Turing machine**
- p is a **program** (a finite binary string) that, when run on U, produces x

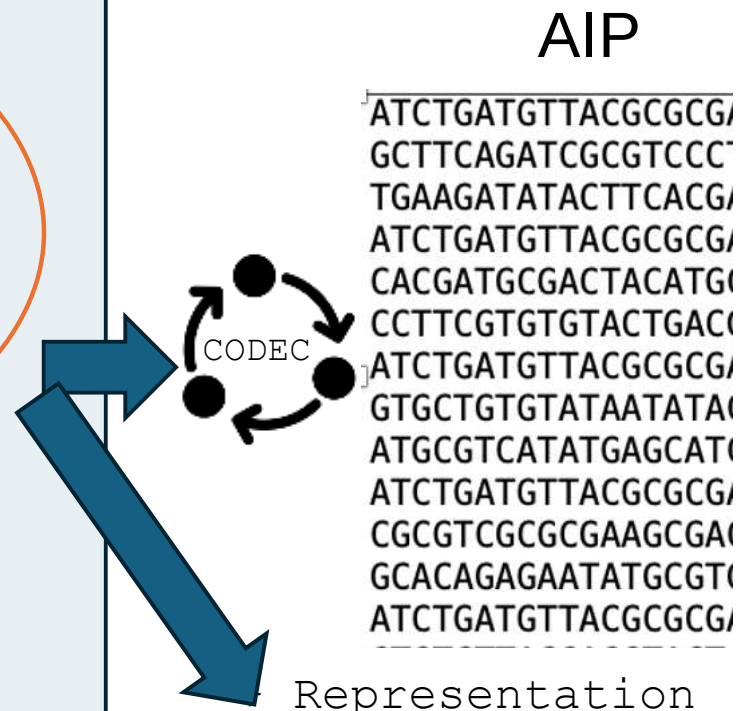


# OAIS Architecture DLC



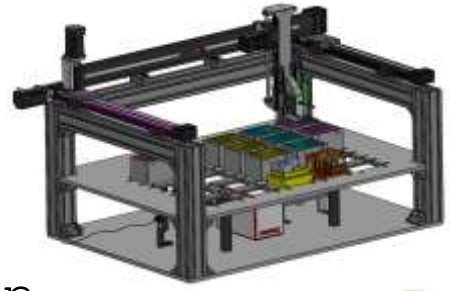
# AIP file structure (zip files)



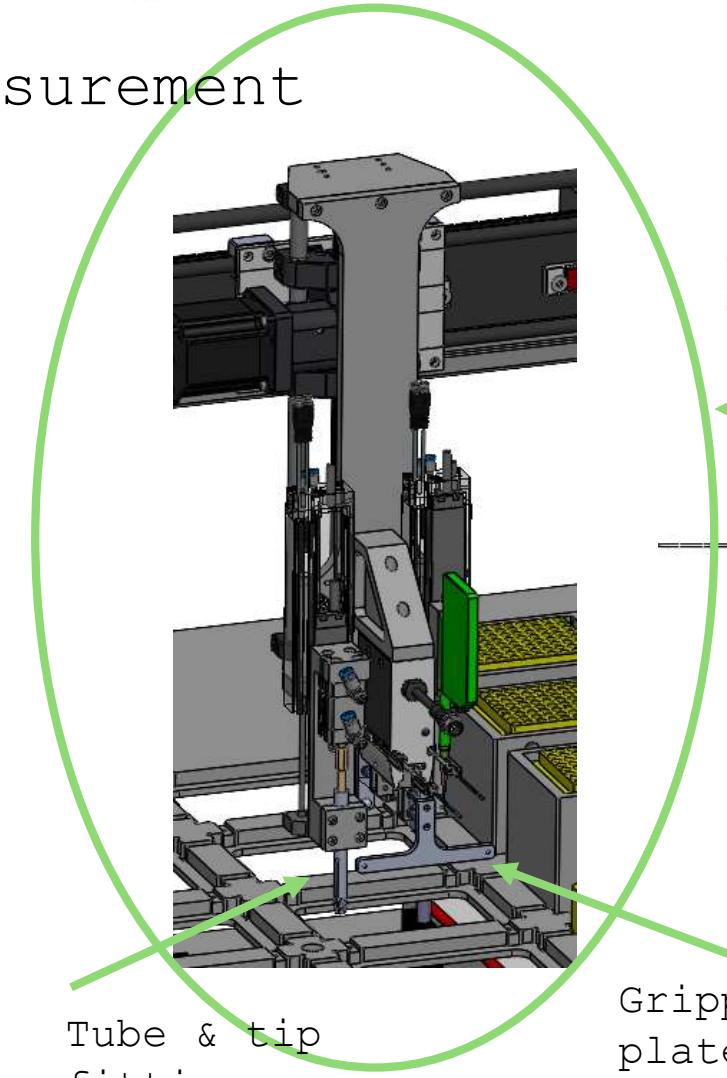



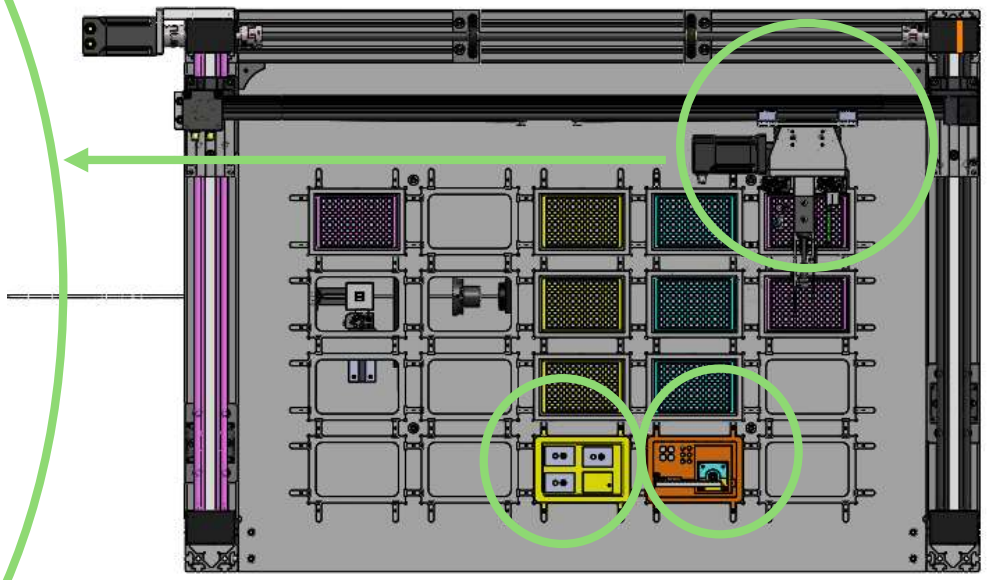

Representation Information needed to decode the AIP (includes the

# Wet Bench



- DNA amplification
- Concentration measurement





Tube & tip fitting picking head

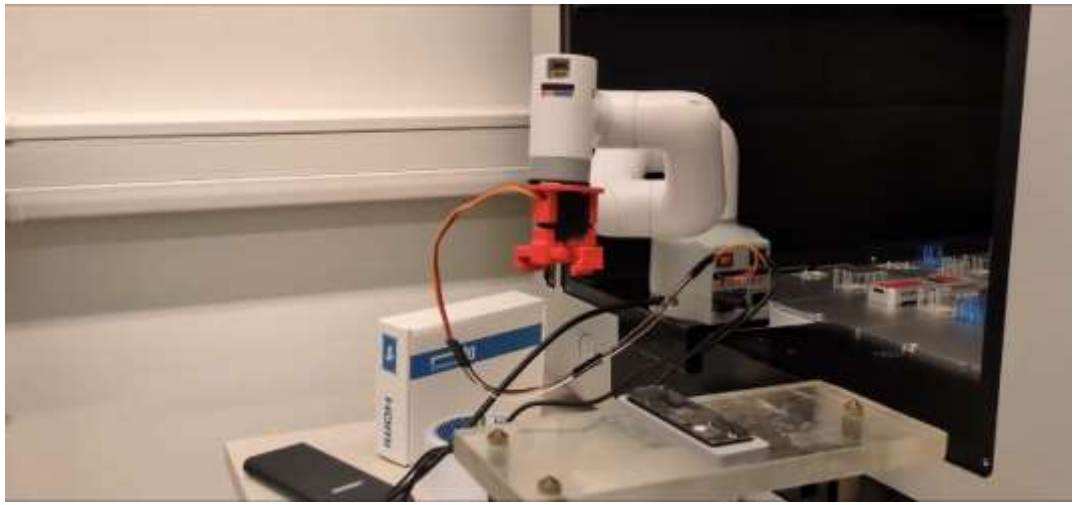
Gripper, well plate footprint compatible

Thermal bath,

Filtration unit, HES-SO

# Wet Bench 2 - Opentrons Based

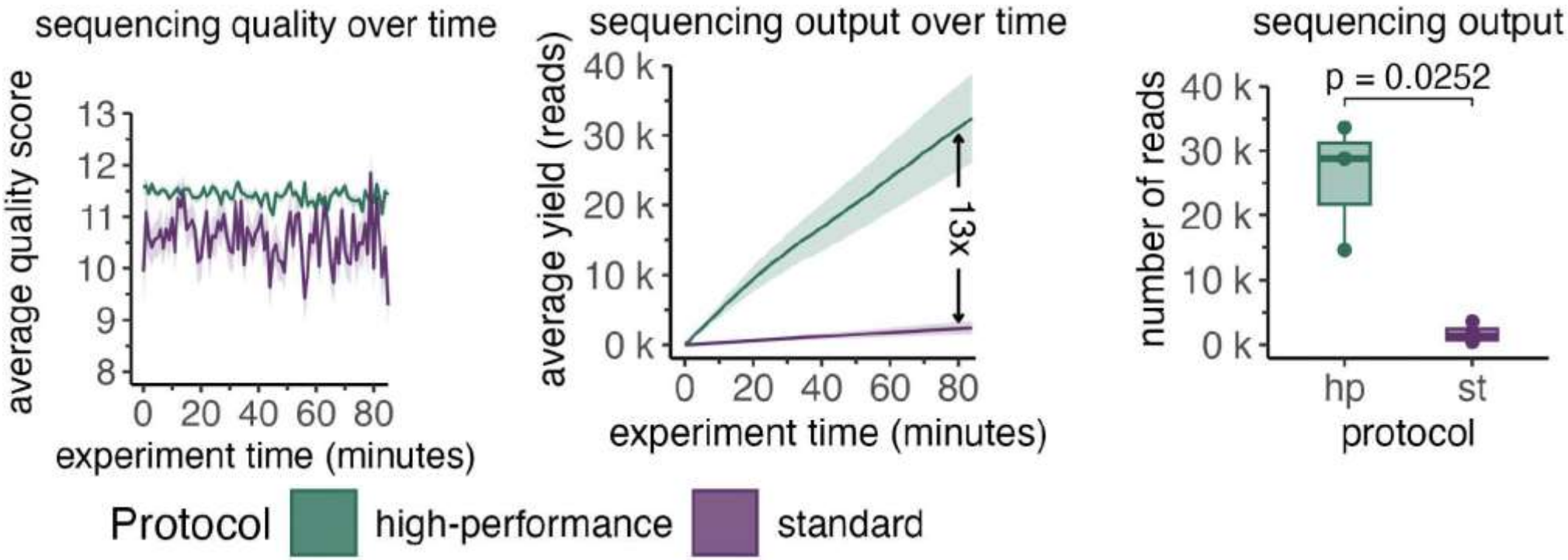
- Primer synthesis
- DNA Library preparation and sequencing implementation according to the Genomica's wet lab protocol
- Concentration measurement



Proprietary technology :

- Thermal unit and filtration unit
- Universal arm for ONP MinION cartridge

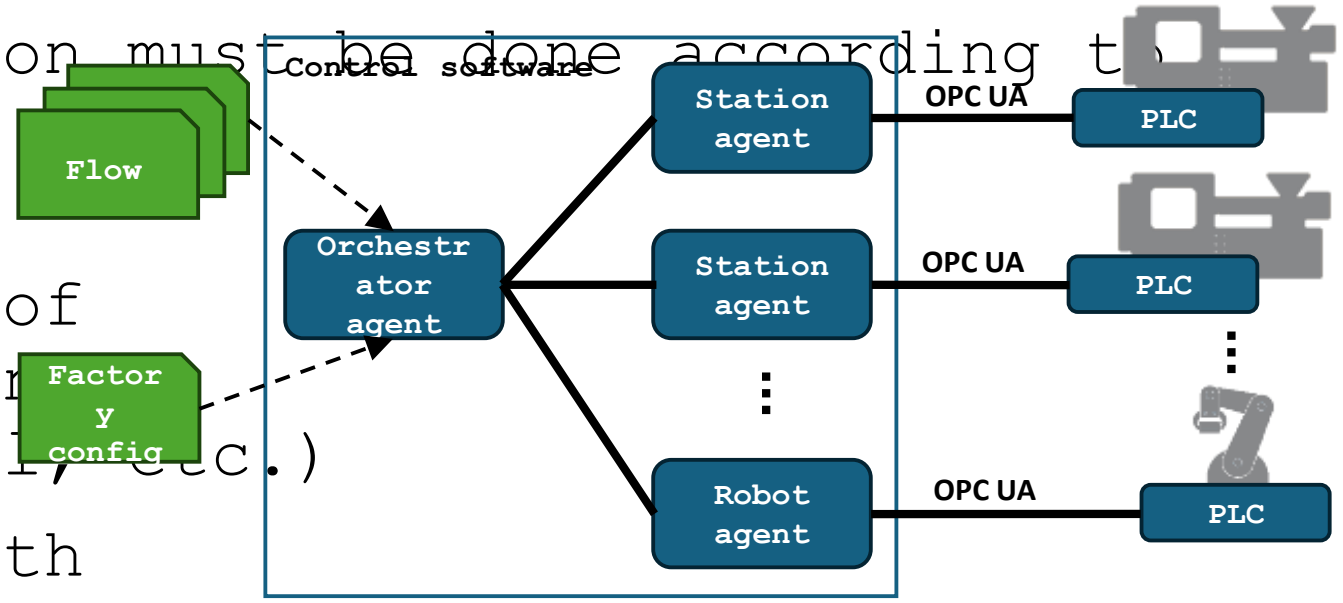
# Genomica's wet lab protocol[1] for nanopore sequencing



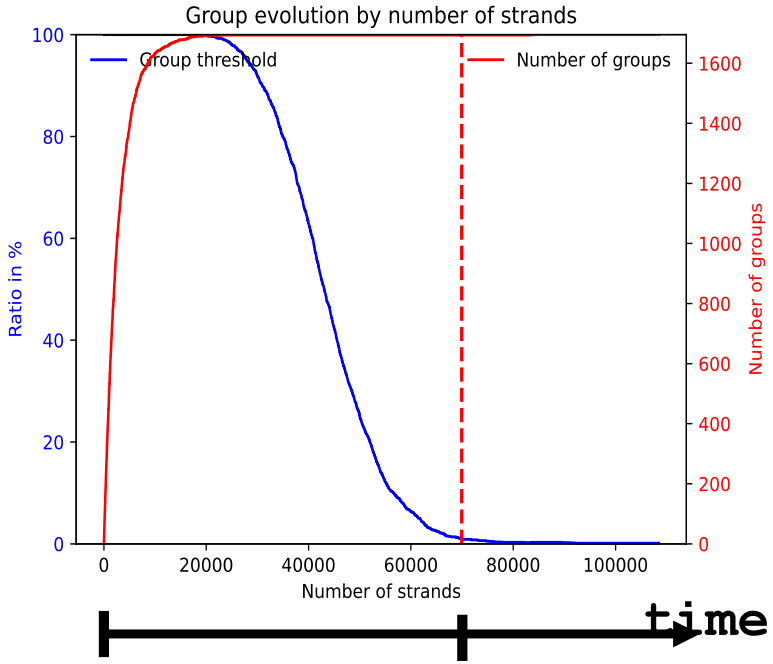

(1) Žemaitis L et al. (2025) High-performance protocol for ultra-short DNA sequencing using Oxford Nanopore Technology (ONT). PLoS One 20(4): e0318040. https://doi.org/10.1371/journal.pone.0318040

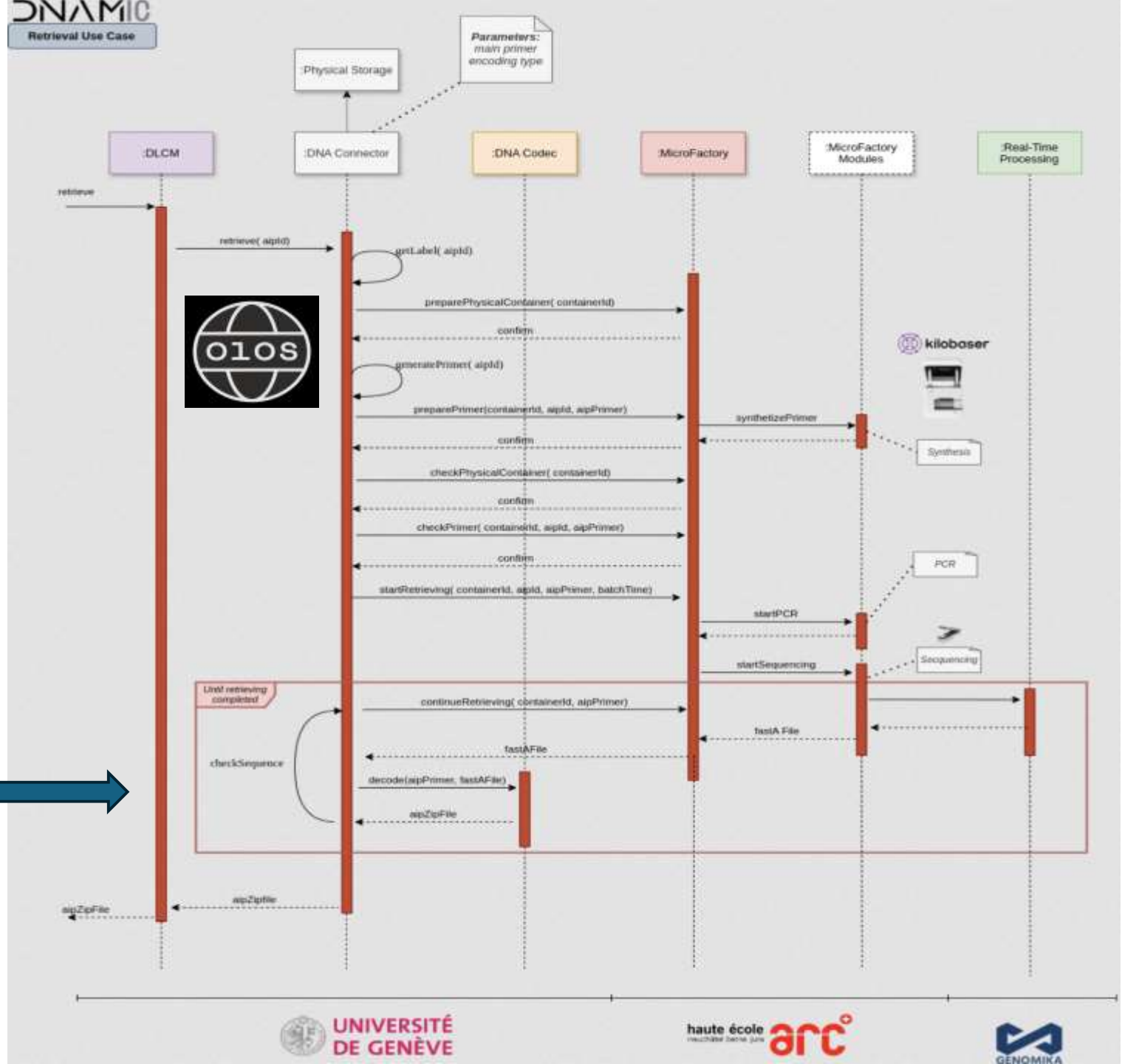
# Supervisor system and autonomization

- Control the tech stations and robots according to one or several flows
- Choose which tech station or robot must execute each operation
- Choose when an operation must be done according to flow and priority
- React to the requests of the tech stations (brin consumable, change tool, etc.)
- Synchronize actions with interaction between a tech station and robot



# to the DNA micro-factory with feedback control



# The whole cycle

DNAMIC

6 use cases

Archivist

OAIS

OIOS

For a total of
1'075'828 bytes
31'642 strands
7'689'024 nts

Files

Recognized standard

[10011001]

[10011001]

[TCGTCTAC]

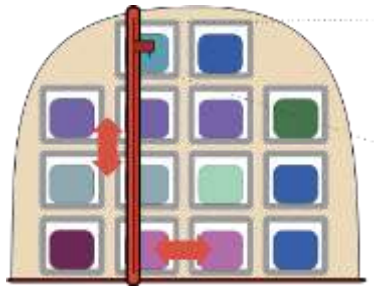
Tested

Twist Bioscience

Segmentation

Encoding

Synthesis



Micro-Factory

Primer Synthesis

PCR Sequencing

Grouping Alignement

Reconstruction Decoding

[TCGTCTAC]

Groups with noisy strands

[TCGTCTAC]

Multiplied, mixed and noisy

# Challenges and Next steps

- Synthesis: currently limited to primers
- Waste management : Reusable pipettes ? Echo liquid handler ?
- Disaster recovery needs a sound self-description – the approach has not yet been finalized but the CODEC complexity is kept low
- Optimizing CODEC to target TB scale

# Thanks for your attention !

DNAMIC

Contact us: https://dnamic.org/contacts