EOSC Activities in the High Energy Physics community

Nuno Castro

nuno.castro@fisica.uminho.pt

thanks to Jorge Gomes
for the discussions and
materials for this talk

11th Iberian Grid Conference
Faro, Portugal, 10th October 2022

Physics projects with relation to EOSC

- **ESCAPE**: Establish a single collaborative cluster of next generation European Strategy Forum on Research Infrastructures (ESFRI) facilities in the area of astronomy- and accelerator-based particle physics in order to implement a functional link between the concerned ESFRI projects and European Open Science Cloud (EOSC)

- **ExPANDS**: collaboration between 10 national Photon and Neutron Research Infrastructures and EGI. The aims to deliver standardised, interoperable, and integrated data sources and data analysis services for Photon and Neutron facilities

- **PANOSC**: Photon and Neutron Open Science Cloud, joins six European research infrastructures (ESRF, CERIC-ERIC, ELI Delivery Consortium, the European Spallation Source, European XFEL, ILL, and the e-infrastructures EGI and GEANT

- **interTwin**: interdisciplinary Digital Twin Engine for modelling and simulation includes fast simulation solution to complement the Monte Carlo approach at the LHC and Lattice QCD simulations to develop a theoretical understanding of matter in the plasma phase (started last September)

- **ARCHIVER**: Pre-Commercial Procurement (PCP) approach to competitively procure R&D services for archiving and digital preservation, lead by CERN

- **EOSC-Future**: Implementing the European Open Science Cloud (EOSC), CERN participates with the consolidation of data lake activities and enabling open access to data

- **FAIR4FUSION**: make European funded data more widely available to the fusion community (finished in May 2022)

ESCAPE partners

- 31 Partners
- 7 ESFRIs & flagship projects
- 2 EIROs (CERN, ESO)
- Budget 15.98M€
- Feb 2019 - Jan 2023

- High Energy Physics
  - HL-LHC
    - High Energy Particle Physics
  - FAIR
    - High Density Exotic Matter Physics

- Multi Messenger Astronomy
 - Not a topic of this presentation

Computing in High Energy Physics

- Particle physics accelerators and detectors are amongst the most complex devices built by the humankind

- Being on the edge of the technology is required
  - Big Data
  - Advanced computing
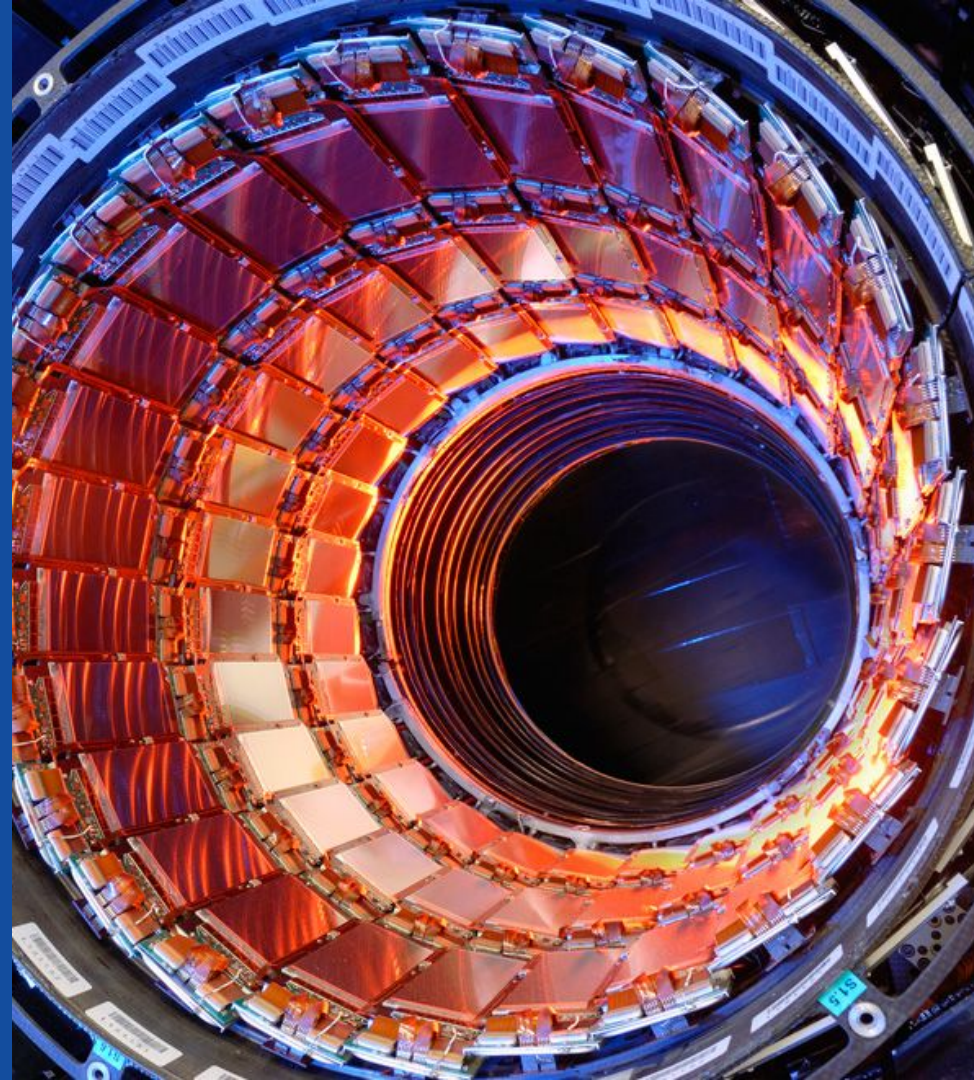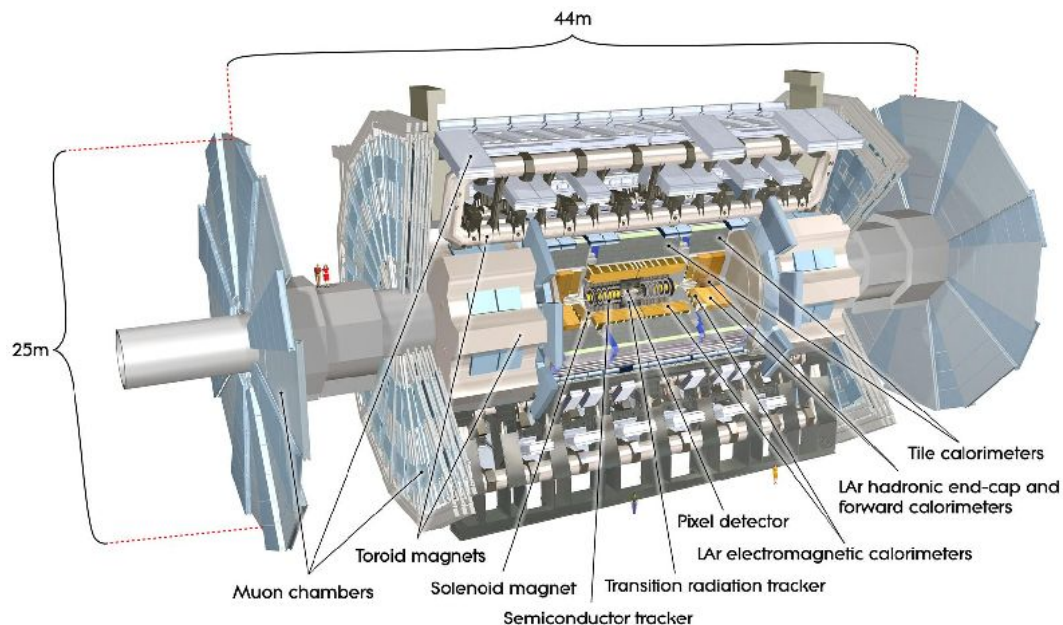
Big Data - High Energy Physics in context

https://cloud.datapane.com/reports/dkjK28A/big-data-2021
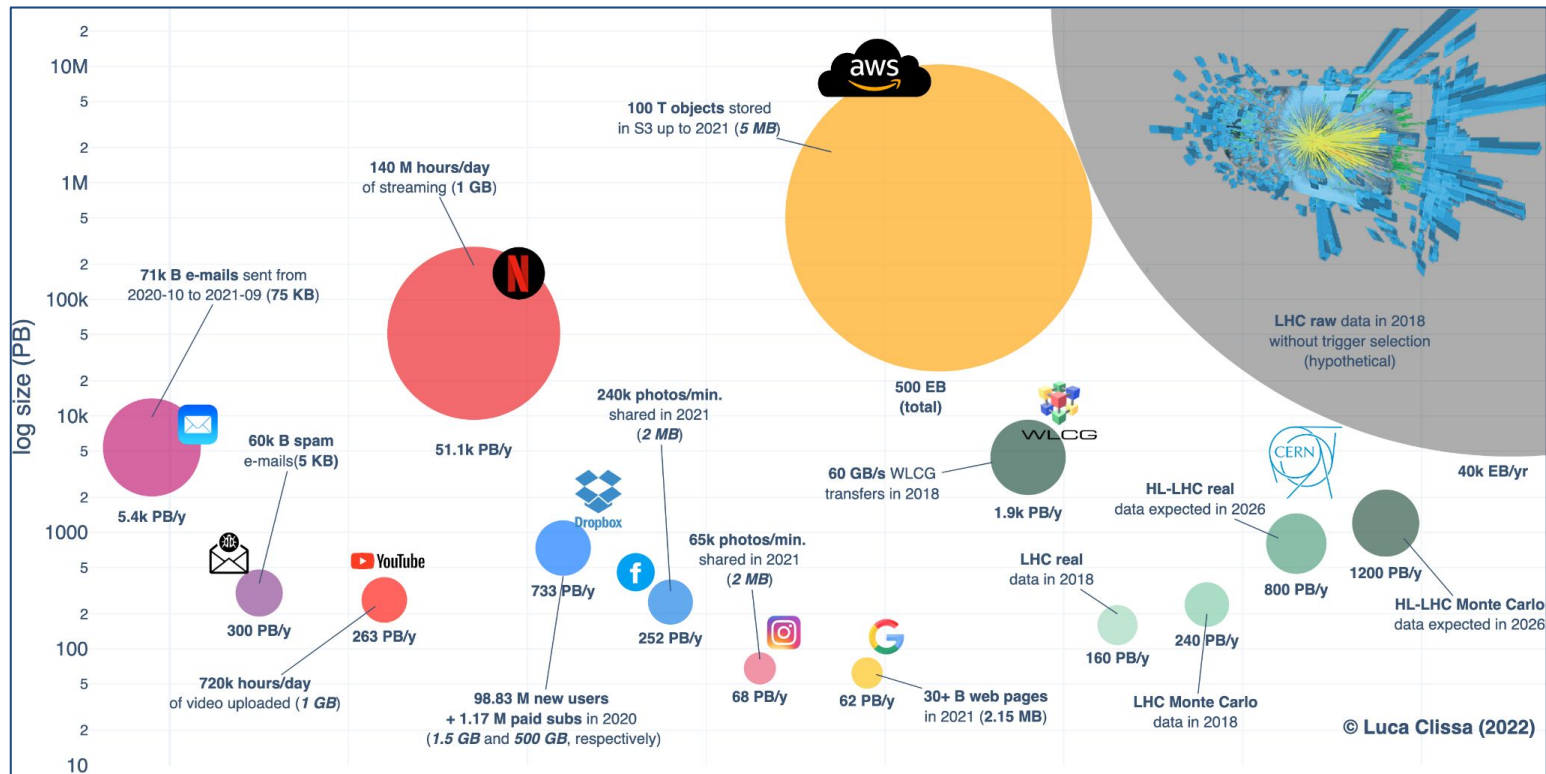
from data to physics at the Large Hadron Collider
a long and complex path
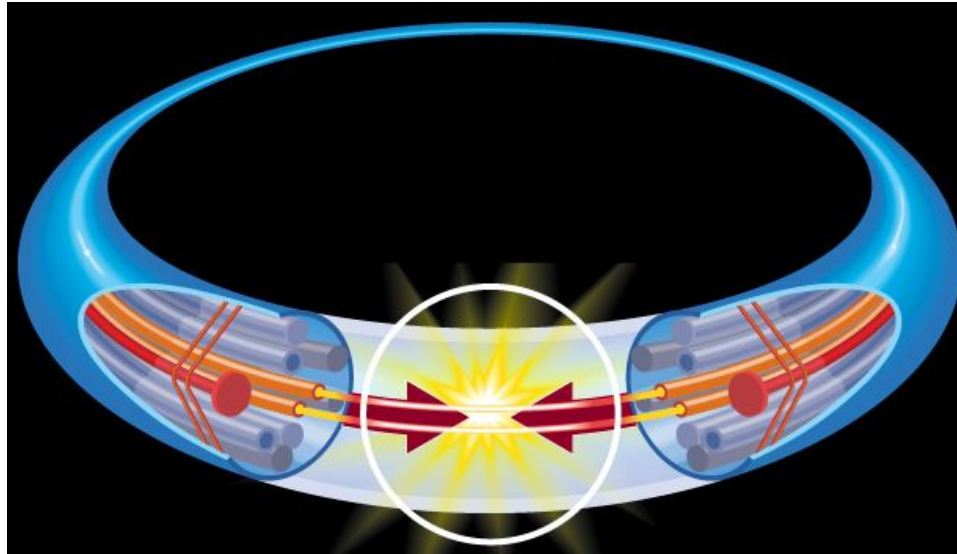
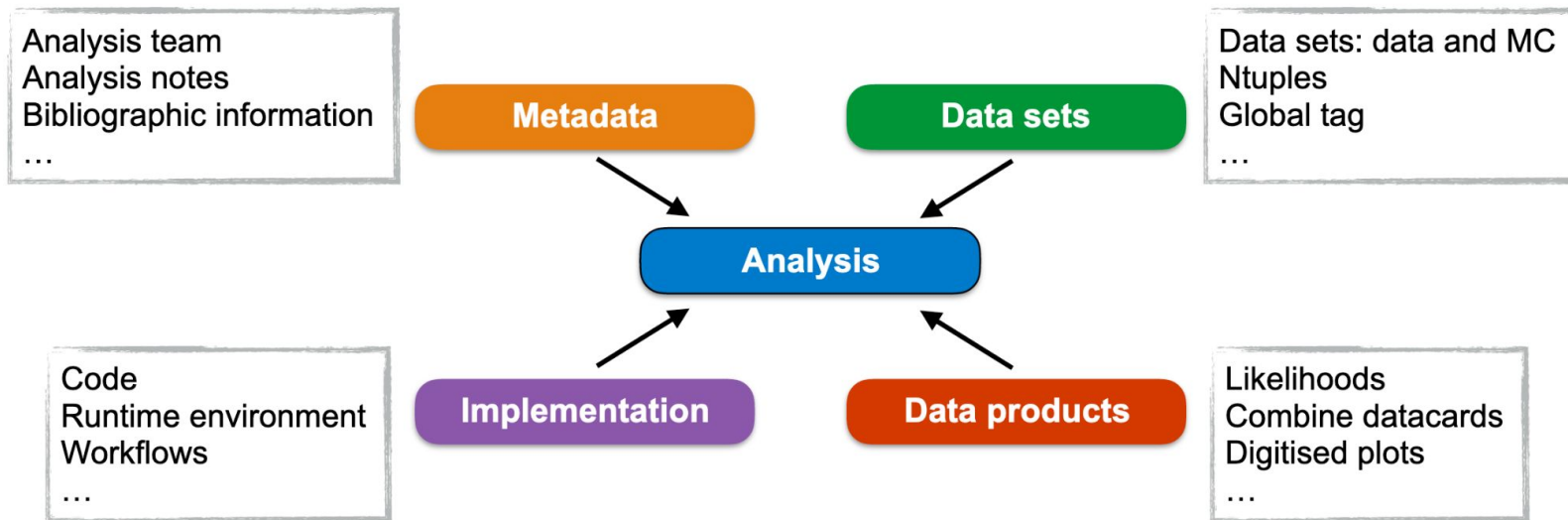- 40 million proton-proton collisions per second

from data to physics at the Large Hadron Collider
a long and complex path

- O(100 million) readout channels

- assuming 1 channel = 1 byte

  40×10^6 ev/s * 100×10^6 byte/ev

  = 4 PB/s

44m

25m

Tile calorimeters

LAr hadronic end-cap and forward calorimeters

Pixel detector

LAr electromagnetic calorimeters

Toroid magnets

Transition radiation tracker

Muon chambers

Solenoid magnet

Semiconductor tracker

from data to physics at the Large Hadron Collider
a long and complex path

Ingredients for analysis preservation

Analysis team
Analysis notes
Bibliographic information
…

Metadata

Data sets

Data sets: data and MC
Ntuples
Global tag
…

Analysis

Code
Runtime environment
Workflows
…

Implementation

Data products

Likelihoods
Combine datacards
Digitised plots
…

[credit: Clemens Lange]

Software preservation

- LHC experiments are investing in reuseable / reproducible analysis

- Technology Choice for software archival:

  - Git

  - Linux Containers
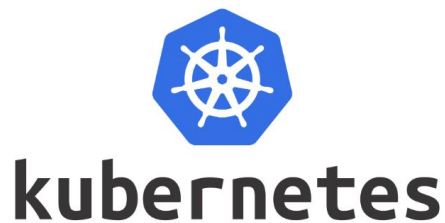
Software preservation - reinterpretation of results

- Analysis Preservation as part of the workflow

- Allow future theories to be tested against current analysis (reinterpretation)

[credit: Lukas Heinrich]
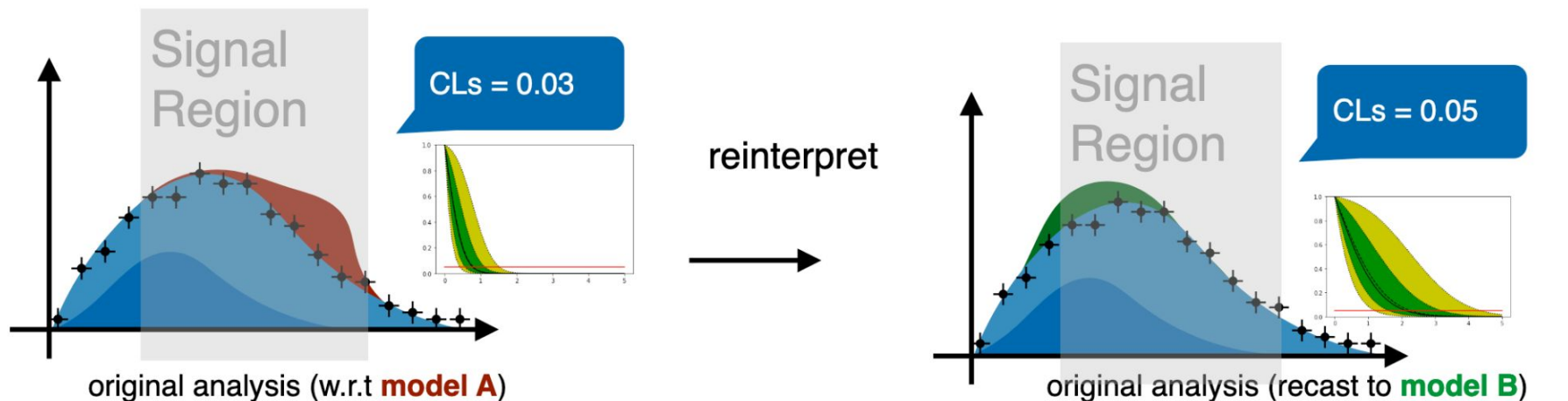
Preservation of Code, Scripts, Workflows

https://reanahub.io
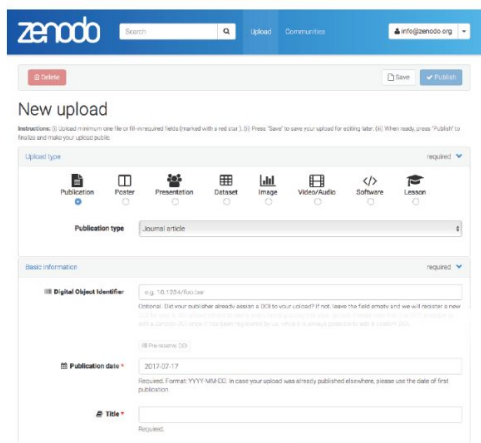
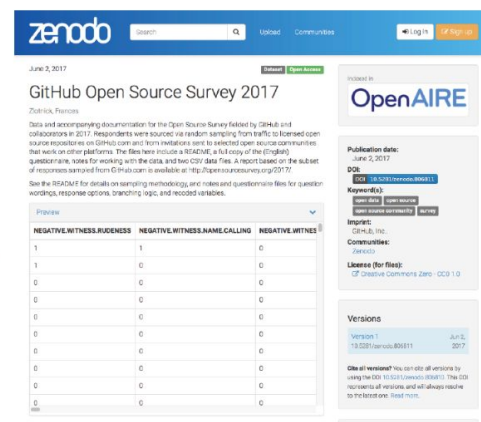# Data repositories
zenodo

https://zenodo.org
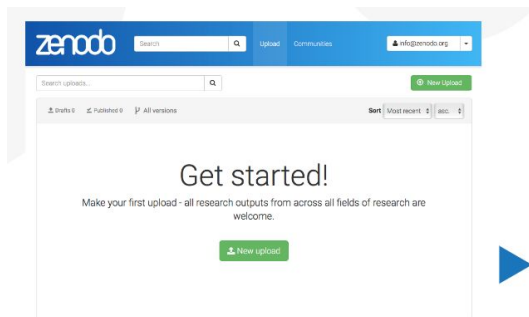
Upload

Any size/format
Any science
Any research output

Describe

Reusable for others
Link to related research
Open, embargoed and closed content

Publish

Instantly available
DOI: Citeable. Discoverable.
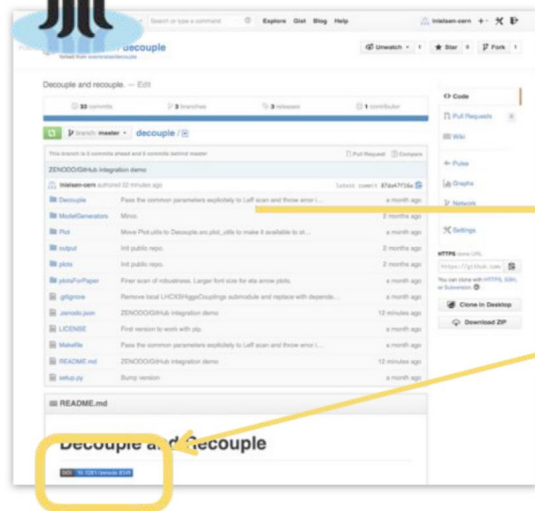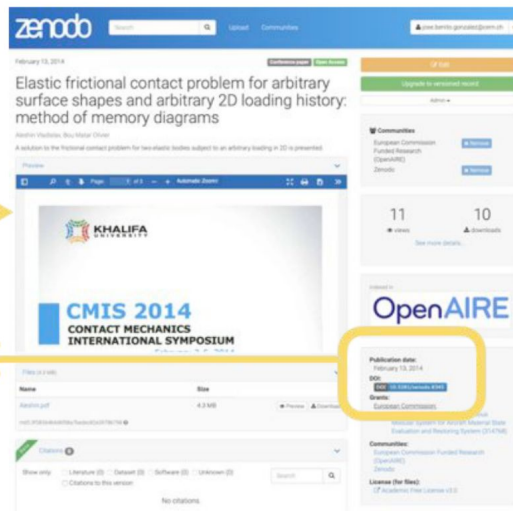Article Level Metrics

# Data repositories
zenodo

Archive: GitHub integration

Data lake model

⇒ Keep the real value from the experiments safe
- (RAW) **data** and a solid baseline of **CPU** in owned and stable sites
- Allow for multiple CPU resources to join, even temporarily
  - Eventually choosing the cheapest at any moment
- Solid networking: use caches / streaming to access data

⇒ Reduce requirements for Computing resources
- Commercial Clouds
- Other sciences' resources
  - SKA, CTA, Dune, Genomics, ...
- HPC systems

[credit: Tommaso Bocali]

Data lake model

[credit: Simone Campana]
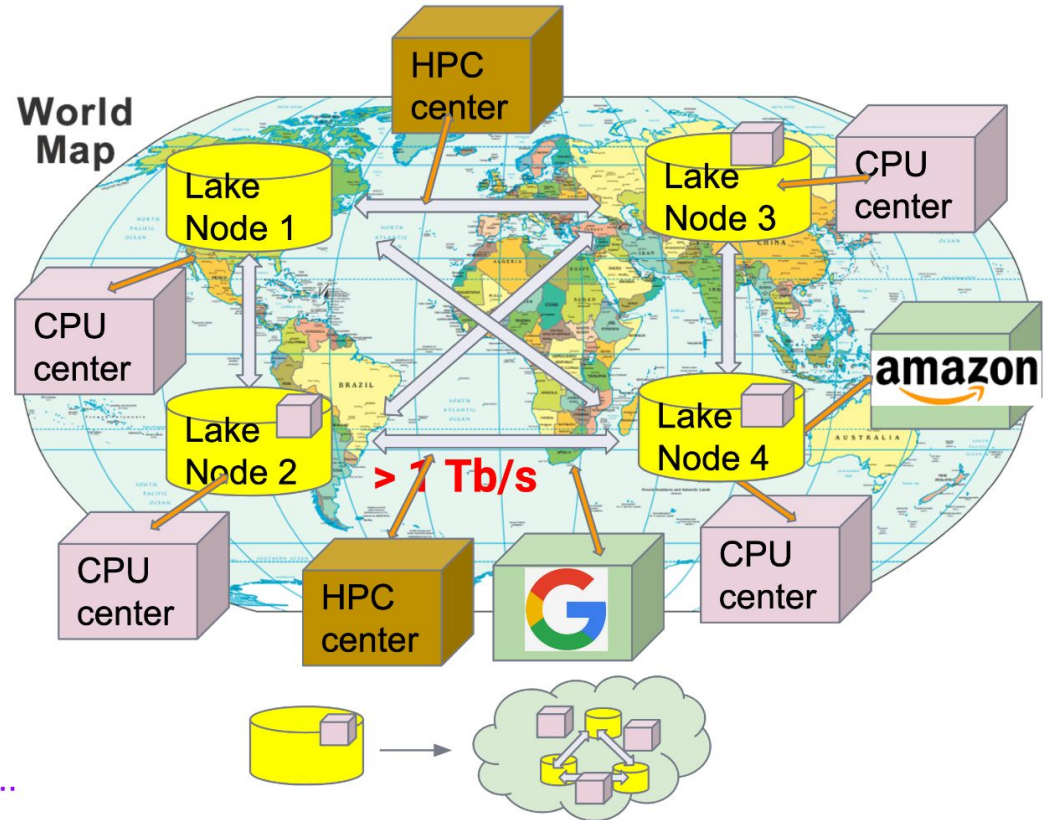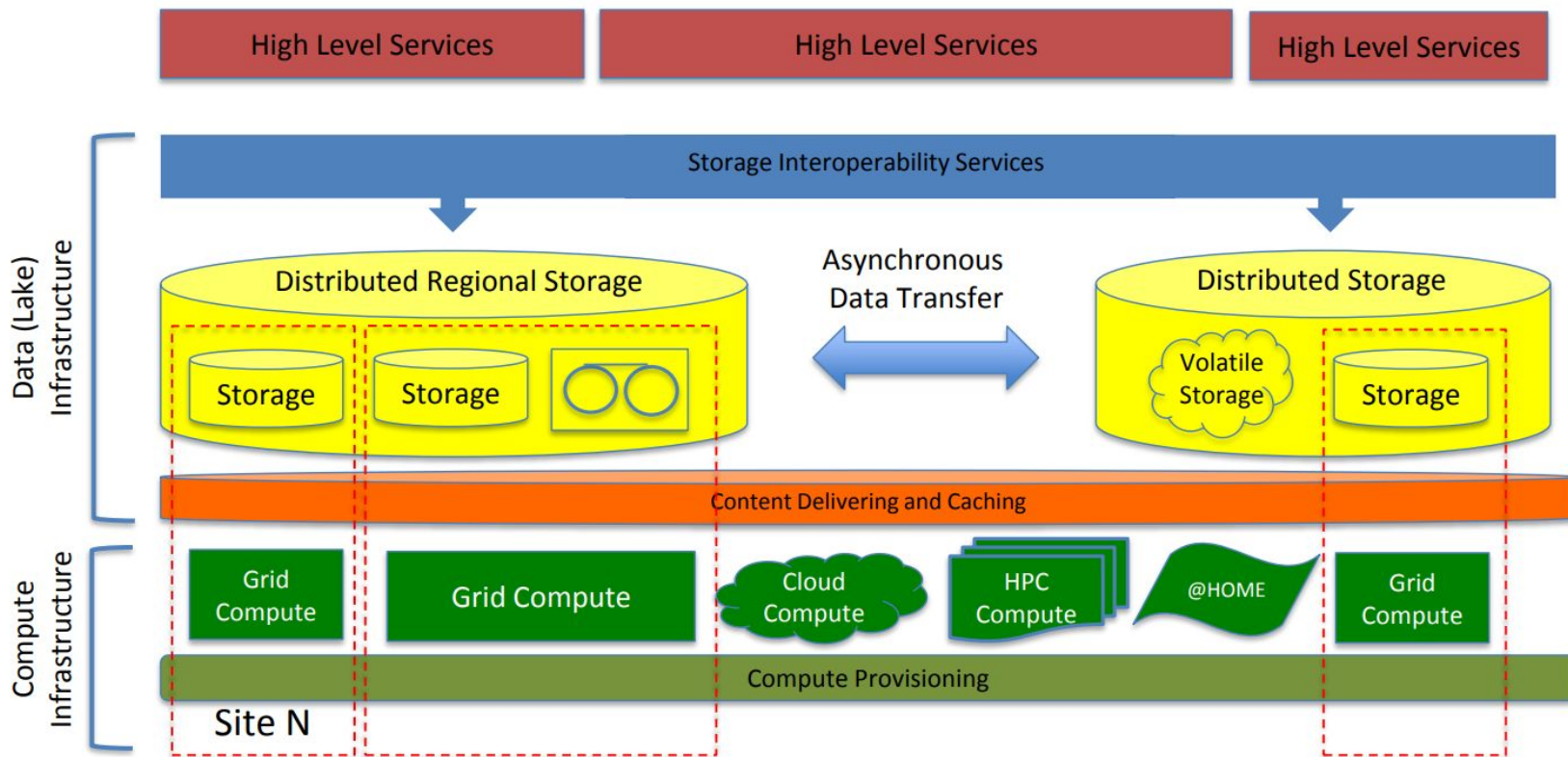
Thanks!

any questions?

you can also find me at nuno.castro@fisica.uminho.pt

https://labs.openai.com/s/aUBB7NIMAjxThBOpuYwoboml