



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

ESCAPE Data Lake 2021

Xavier Espinal (CERN) - ESCAPE WP2 lead

WP2/WP2 joint workshop - 6th and 7th of April 2021

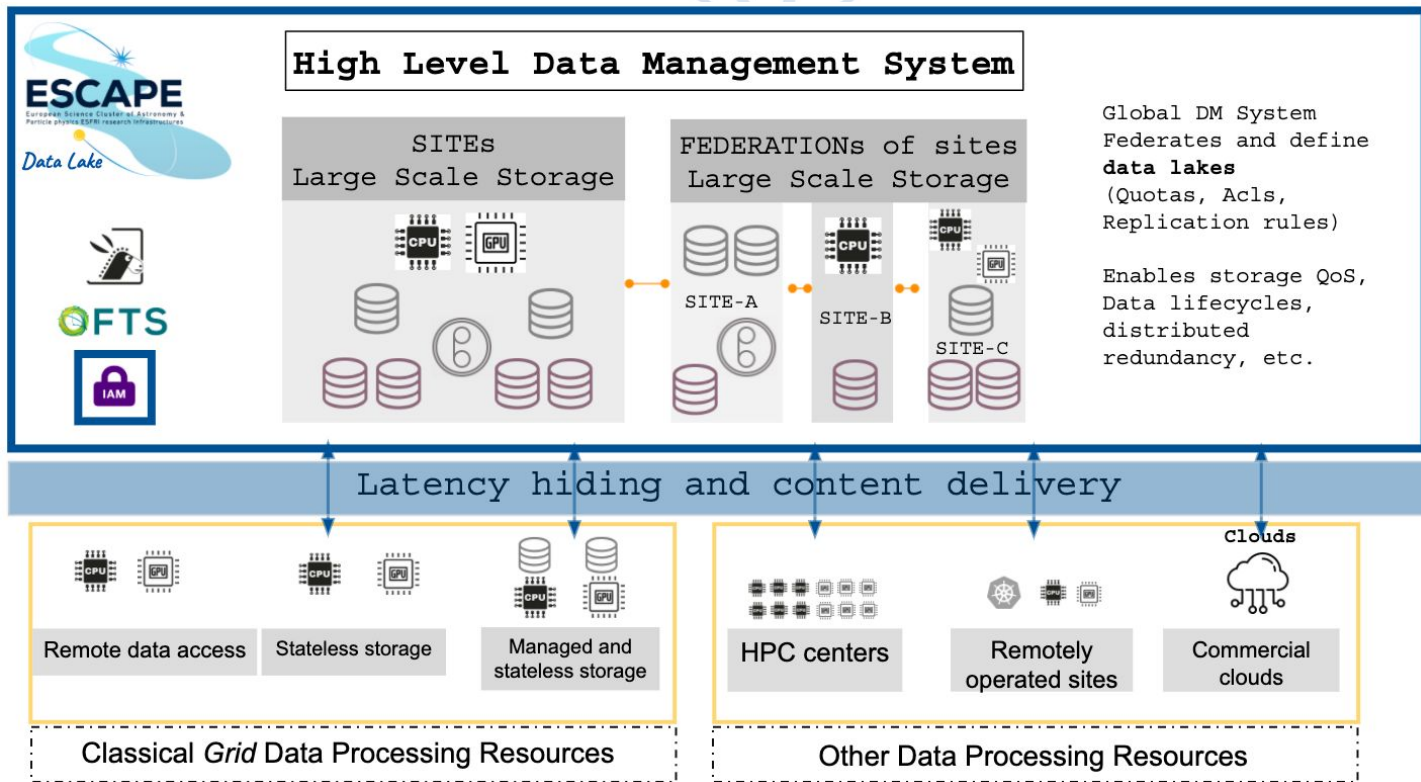


The ESCAPE Data Infrastructure for Open Science

- Define, integrate and commission an ecosystem of tools and services to build a data lake

- Contributes to deliver **Open Access and FAIR data services**: trustable data repositories; enable data management policies; transparent data access layer

- Science **projects to drive** the services requirements most suitable to their needs



The ESCAPE Data Infrastructure for Open Science

- Define, integrate and

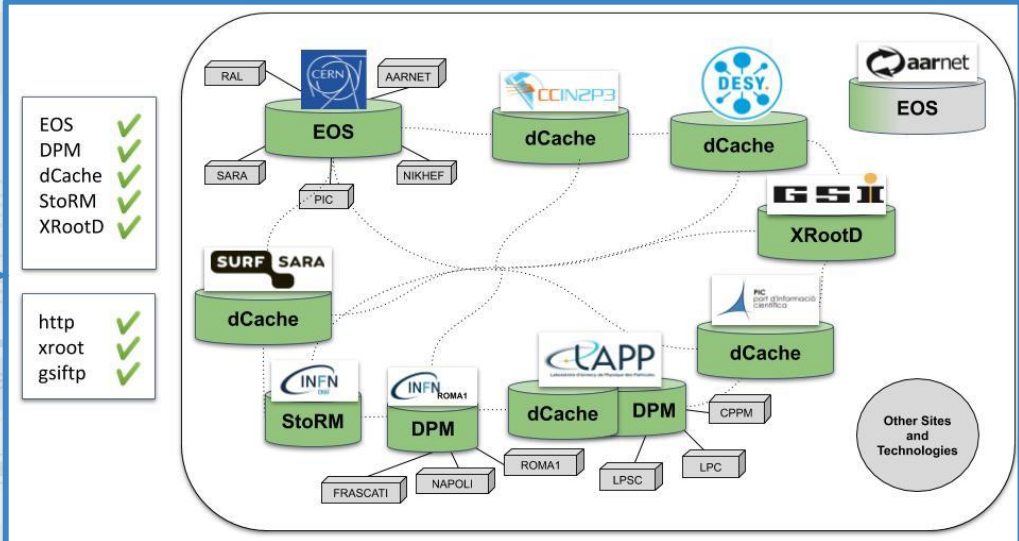
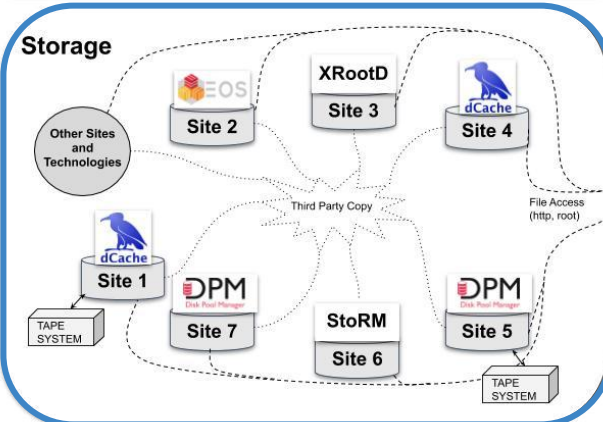
High Level Data Management System

Orchestrator  Rucio Server

Middleware  OFTS



CACHING SOLUTION



- EOS ✓
- DPM ✓
- dCache ✓
- StoRM ✓
- XRootD ✓

- http ✓
- xroot ✓
- gsiftp ✓

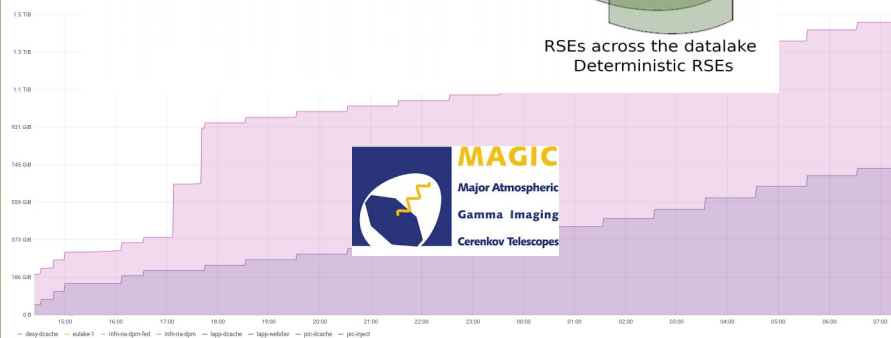
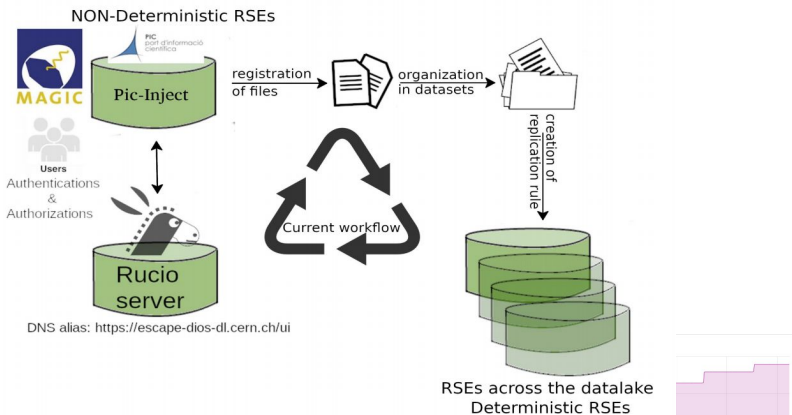
Science projects to drive the services requirements most suitable to their needs

Classical Grid Data Processing Resources

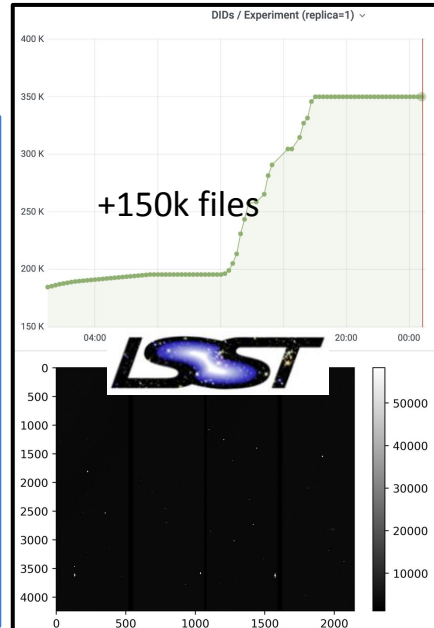
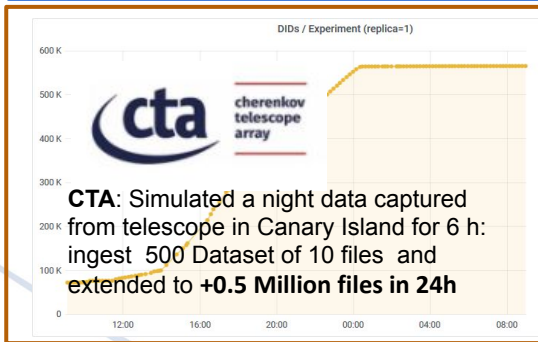
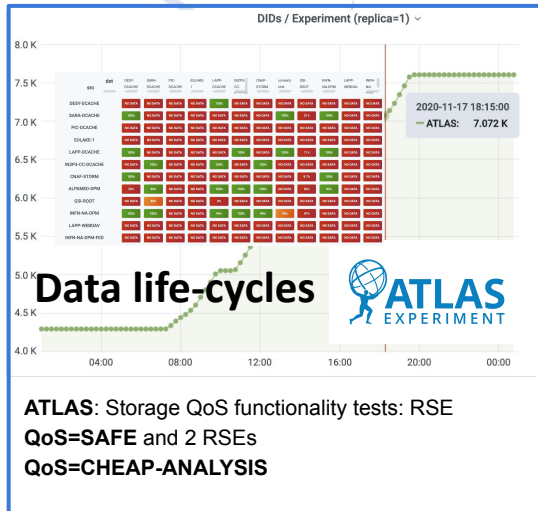
Other Data Processing Resources



Data Lake 24-hour Dress Rehearsal



MAGIC: Mimics a real MAGIC observation use case. Remote storage (Data Lake aware) **next to the telescope** acts as a buffer for subsequent data injection to the ESCAPE Data Lake (and local deletion after success)



LSST: Simulate production conditions: ingest the HSC RC2 dataset from CC-IN2P3 local storage to the Data Lake, **at a realistic LSST data rate (20TB/24h)**. Then **confirm integrity and accessibility of the data via a notebook**.

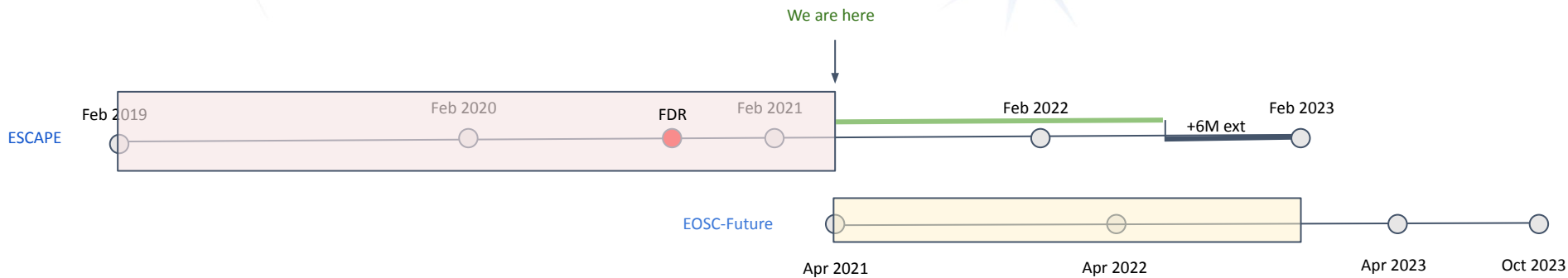
→ The image is a reconstruction drawn within a Jupyter Notebook accessing the data used in the Full Dress Rehearsal.

So far...

- Functional Pilot Data Lake integrating 10 labs with a variety of storage technologies EOS, dCache, DPM, STORM, xrootd, http
- Data Lake/Data Management orchestration layer **consolidated**: AAI/IAM, RUCIO, FTS, CRIC, early QoS and Data lifecycles
- **Strong involvement** from experiments: data workflows tested, Data Lake operations, technology transfer, etc.
- ESCAPE Data Lake **successfully integrated** with notebook platforms. This is a multi-purpose goal: User Analysis and Open Data
- Full Dress Rehearsal exercise Nov 2020 [[+info](#)]



ESCAPE 2021



ESCAPE 2021

- Evolution from the pilot to the prototype phase of the ESCAPE Data Lake
 - Data management and Data Processing: injection, preparation and analysis
 - End-to-End AAI, Advance with token based integration, fine-grained AuthN/Z
 - External resource integration: clouds and HPC, expand data life-cycles
- Demonstrate Data Lake orchestration tools sustainability after ESCAPE (towards the **EOSC**)
 - Leverage, integrate and use the new experiment and site's installations
- Keep the **strong** involvement from the RIs and experiments
 - We are covering a wide range of disciplines: Astro-Particle, Radio-Astronomy, Gravitational Waves, Cosmology, Particle Physics and Nuclear Physics.
- X-Synergies :
 - PaN (photon+neutron), fellow EC-funded projects: ARCHIVER, CS3MESH4EOSC and National Initiatives: PUNCH4NFDI (Germany), etc.
- Program of Work for the 2nd phase focused on a Full scale exercise by end November, codename **DAC21** (Data and Analysis Challenge [[+draft](#)])

