ESCAPE - A dive into a Datalake for Open Science

Xavier Espinal (CERN) - ESCAPE WP2 leader

PHIDIAS

Webinar - Steps forward in detection and identification of anomalous atmospheric events 13 Oct 2020

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

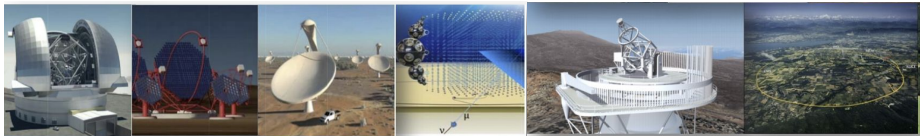
Science Projects

HiLumi HL-LHC PROJECT
SKA SQUARE KILOMETRE ARRAY
VIRGO
FAIR
KM3NeT
JIVE Joint Institute for VLBI ERIC
cta cherenkov telescope array
LSST
EST european solar telescope
ESO
CERN

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

EUROPEAN OPEN SCIENCE CLOUD

- Prototype an infrastructure adapted to **Exabyte-scale** needs of large science projects
- **Common** data infrastructure for Astro-particle, Radio-astronomy, Gravitational Waves, Cosmology and Particle Physics
- Ensure the **sciences** drive the development of the EOSC
- Address **FAIR** data management principles



Data centres

CERN
INFN
GSI
DESY
Nikhef
SURF SARA
rijksuniversiteit groningen
CCIN2P3
cnrs
PIC port d'informació científica
LAPP Laboratoire d'Annecy de Physique des Particules
INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

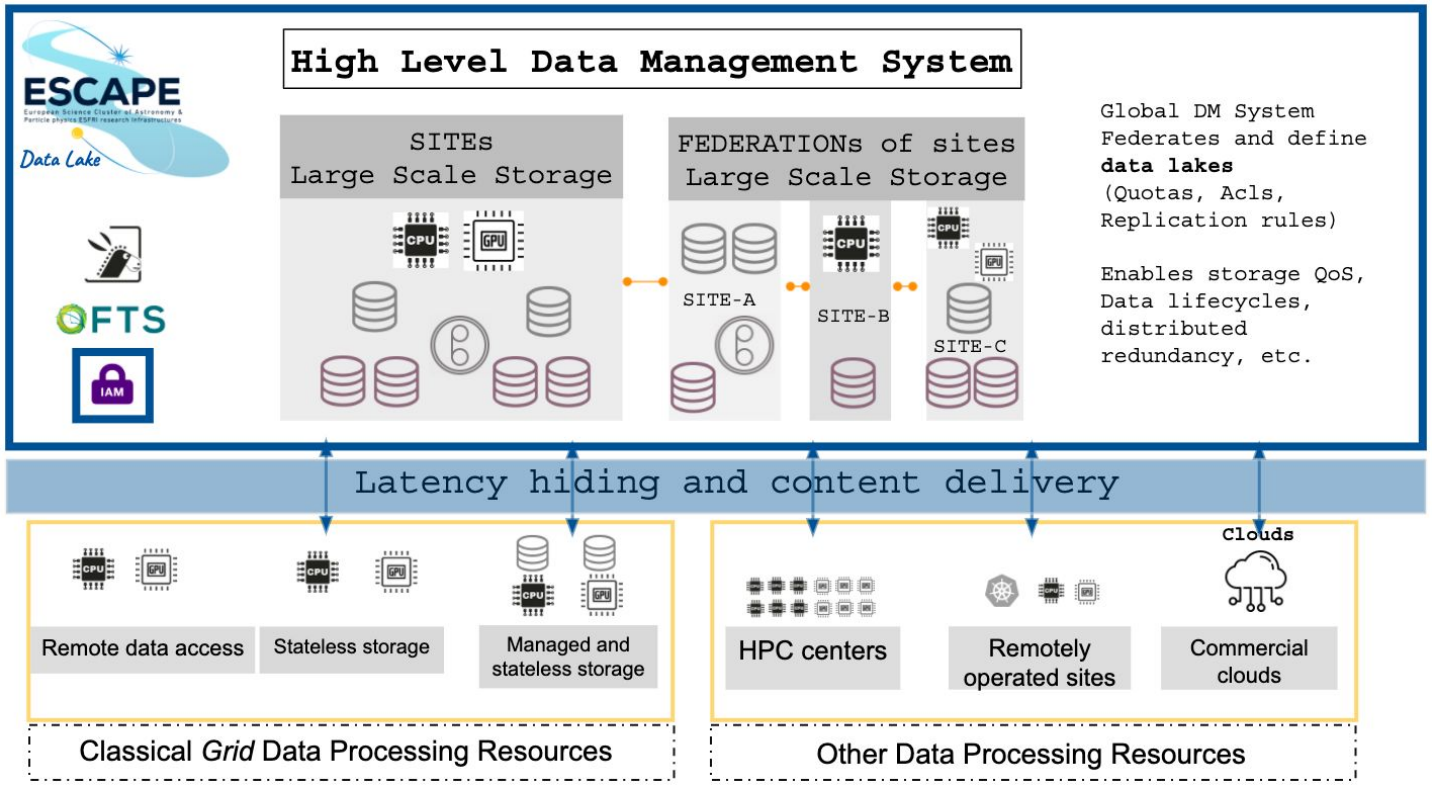
Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Infrastructure for Open Science

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

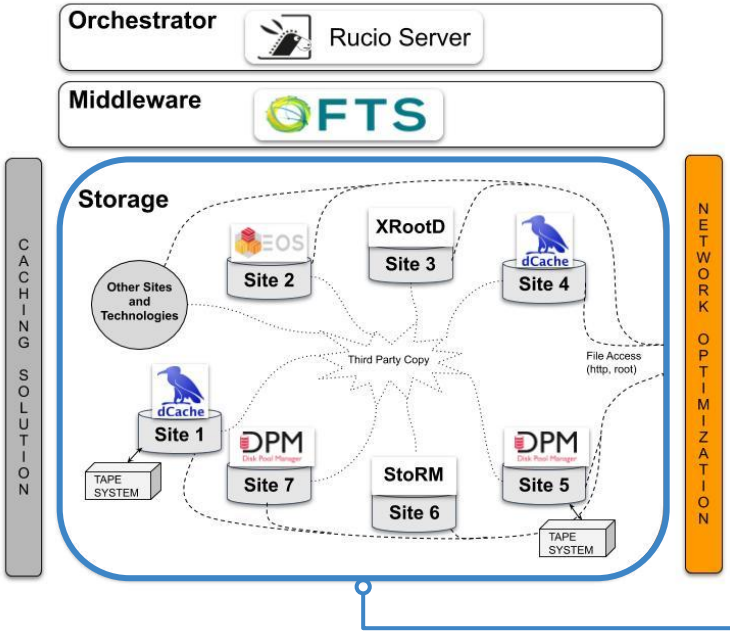
- Contributes to deliver **Open Access and FAIR data services**: trustable data repositories; enable data management policies; transparent data access layer

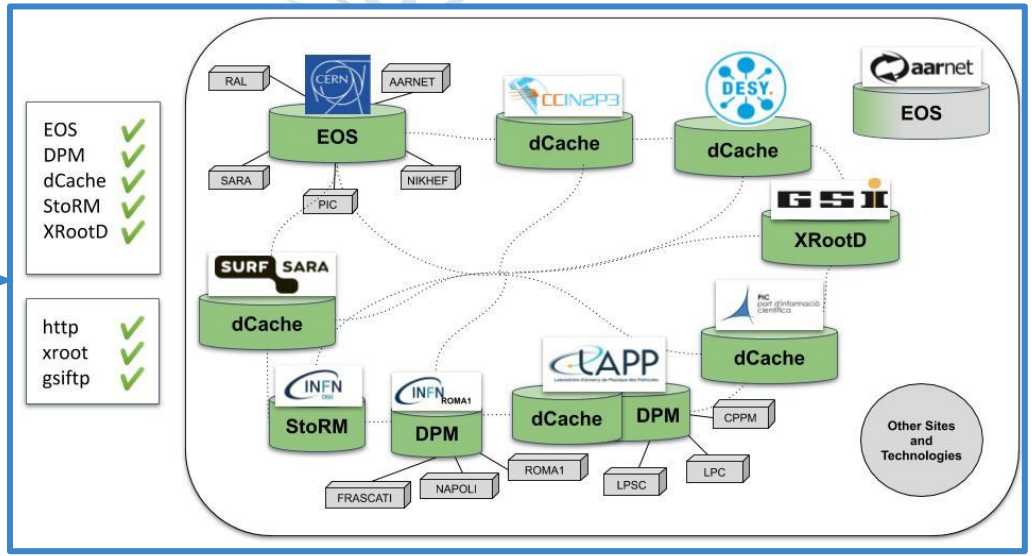
- Science **projects to drive** the services requirements most suitable to their needs



Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Lake





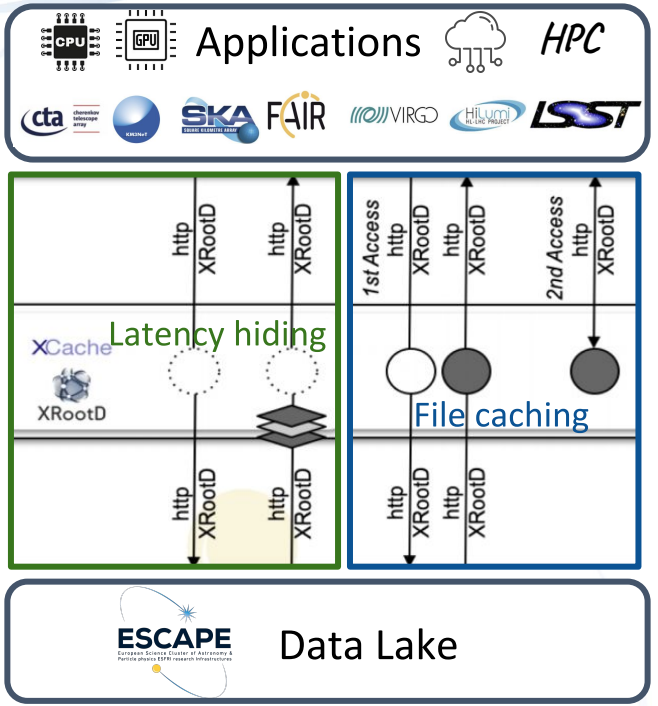
- Hiding complexity and providing transparent access to data
- Heterogeneous federated storage and operations model
- Some centers joining even if not funded by ESCAPE

Further info: https://wiki.escape2020.de/index.php/WP2_-_DIOS#Datalake_Status

Funded by the European Union's Horizon 2020 - Grant N° 824064

A word on Content Delivery and Caching

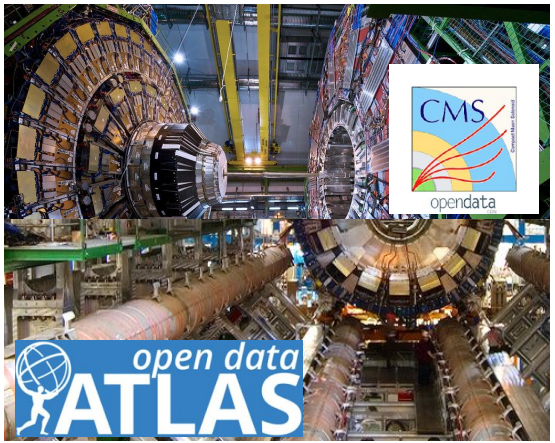
- Streaming caches demonstrate potential on latency hiding and file re-usability in Particle Physics workflows
- Investigating and understanding whether caching can also help on non-event based formats, e.g. images, data-cubes,...
- Caches can facilitate ingress/egress of data with heterogeneous computing resources: Commercial Clouds and HPCs



Funded by the European Union's Horizon 2020 - Grant N° 824064

Data and Data access in the ESCAPE Data Lake

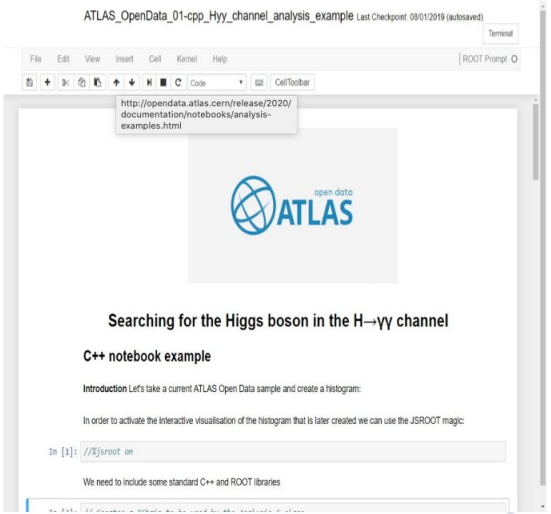
- Pilot Data Lake performance evaluation ongoing with the engagement of:
  - Radio-astronomy (LOFAR, SKA)
  - Astro-particle (CTA and MAGIC)
  - Cosmology (LSST)
  - Gravitational waves (EGO/VIRGO)
  - Particle physics communities (FAIR, ATLAS and CMS)



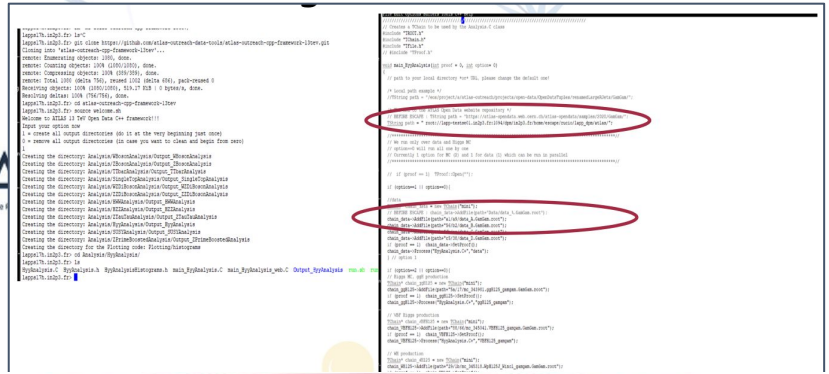
Funded by the European Union's Horizon 2020 - Grant N° 824064

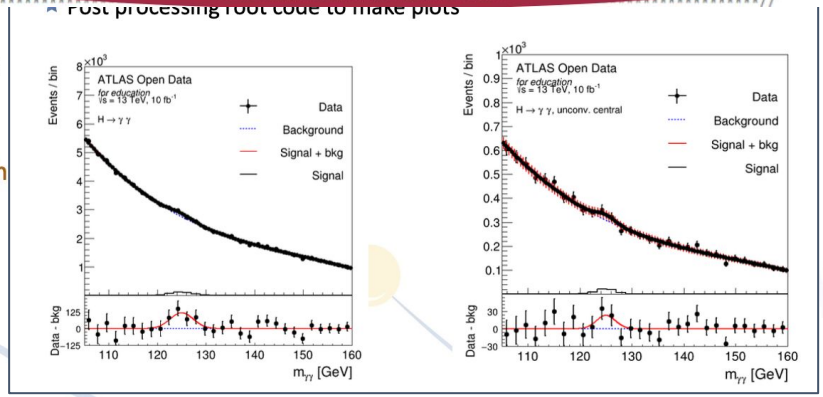
Data and Data access in the ESCAPE Data Lake

ATLAS analysis demo



- ATLAS Exercise
  - Mockup of analysis activity using public ATLAS files
  - C++ code
  - Produce H → yy plot
- ESCAPE exercise :
  - Upload files to ESCAPE datalake with rucio client
  - Adapt file access to ESCAPE datalake
  - Produce plots

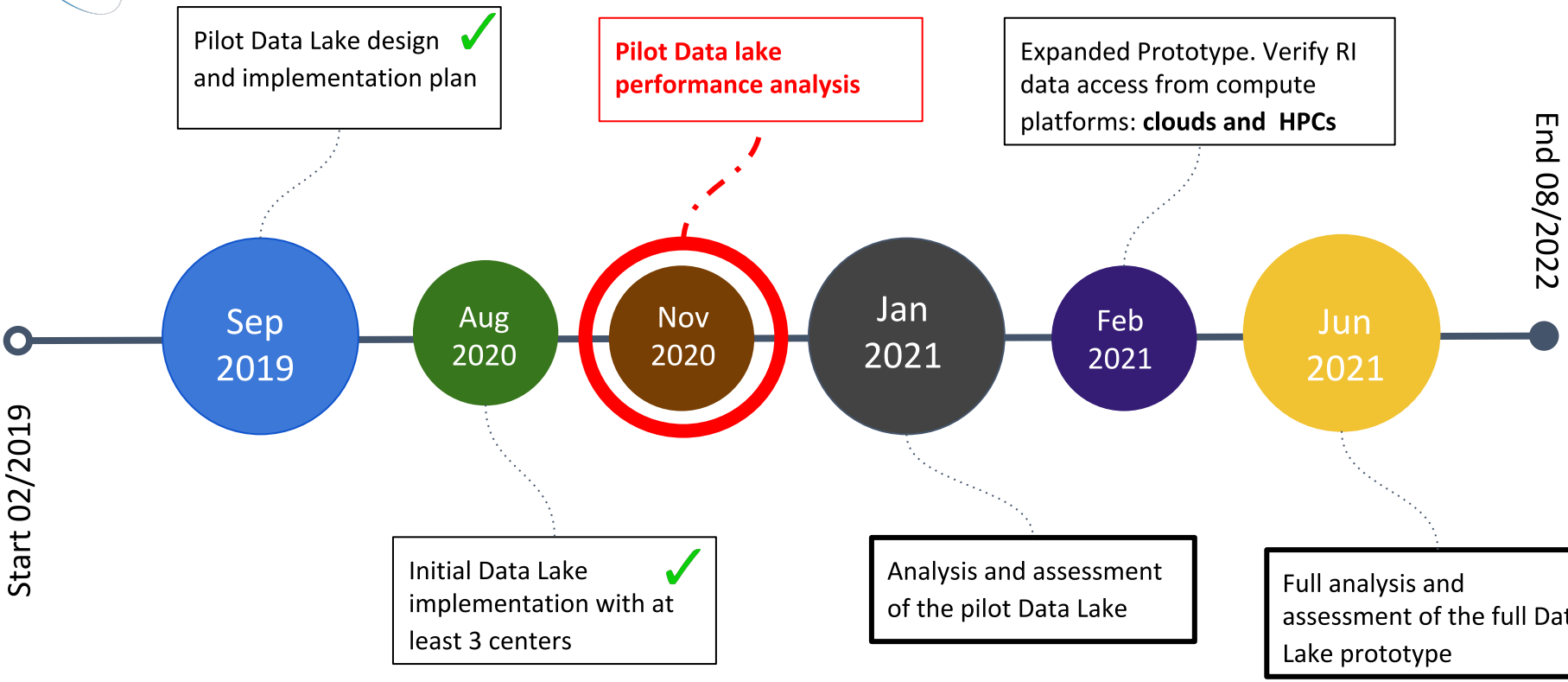




Credits: Stephane jezequel (LAPP)

Funded by the European Union's Horizon 2020 - Grant N° 824064

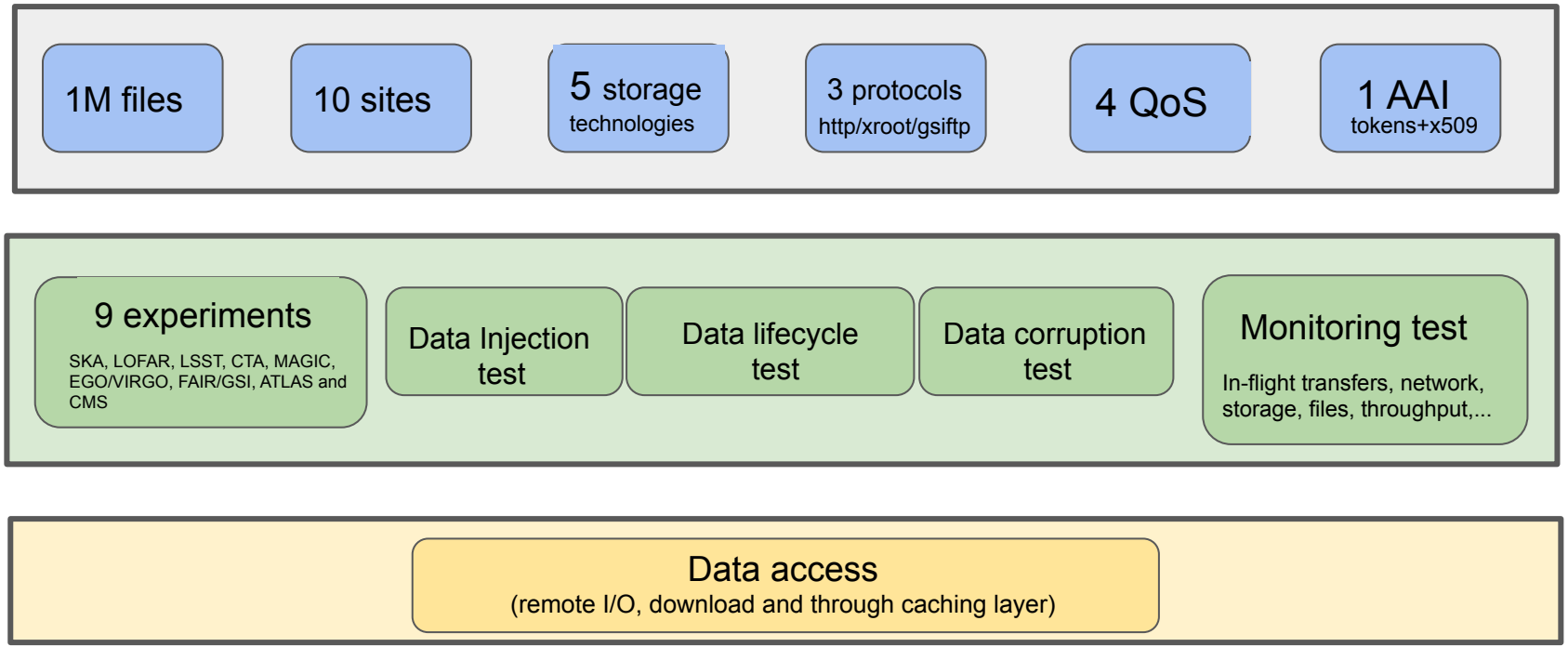
Some important milestones



Funded by the European Union's Horizon 2020 - Grant N° 824064

Pilot Data Lake Full Dress Rehearsal

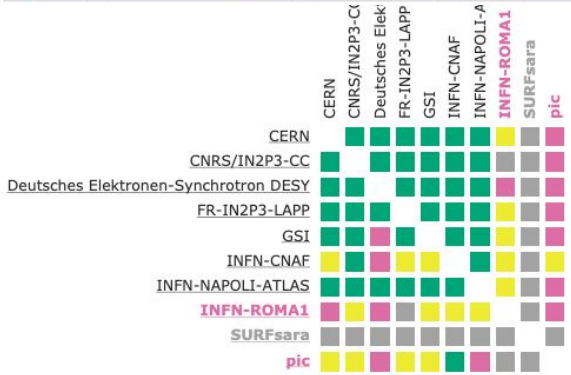
Goal: Exercise covering **experiment data workflow** needs on a single day. From data injection, to data replication and data access. Three fold goal: perspective from **scientists**, perspective from **sites**, and the assessment of the **ESCAPE datalake tools and services** under **pseudo-prod conditions**: RUCIO, FTS, CRIC, IAM, PerfSONAR, monitoring, QoS, clients, etc. **First exercise: 24 November**



ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ESCAPE Mesh Config - ESCAPE IPv4 Bandwidth - Throughput

Throughput >= 1Gbps | Throughput < 1Gbps | Throughput <= .5Gbps | Unable to find test data

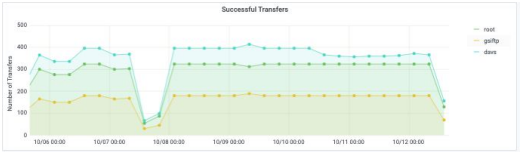


Successful Transfers Percentage (davs)

Source \ Destination	ccdcalitest10.in2p3.fr	dcache-door-doma01.desy.de	dclxwp2dlds1.gsi.de	door05.pic.es	eoseulake.cern.ch	lapp-dcache01.in2p3.fr	lapp-esc02.in2p3.fr	lapp-testse01.in2p3.fr	webdav.grid.sara.nl	xfer.cr.cnaf.infn.it
ccdcalitest10.in2p3.fr	-	0%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	14.65%
dcache-door-doma01.desy.de	100.00%	-	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	18.18%
dclxwp2dlds1.gsi.de	100.00%	84.62%	-	100.00%	100.00%	96.15%	100.00%	100.00%	100.00%	9.09%
door05.pic.es	100.00%	75.00%	100.00%	-	100.00%	100.00%	100.00%	100.00%	100.00%	23.86%
eoseulake.cern.ch	100.00%	0%	100.00%	100.00%	-	100.00%	100.00%	100.00%	100.00%	13.29%
lapp-dcache01.in2p3.fr	100.00%	0%	100.00%	100.00%	100.00%	-	100.00%	100.00%	100.00%	8.81%
lapp-esc02.in2p3.fr	100.00%	0%	100.00%	100.00%	100.00%	100.00%	-	100.00%	100.00%	10.97%
lapp-testse01.in2p3.fr	100.00%	0%	100.00%	100.00%	100.00%	100.00%	100.00%	-	100.00%	10.61%
webdav.grid.sara.nl	100.00%	0%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	-	5.00%

Successful Transfers Percentage (root)

Source \ Destination	ccdcalitest10.in2p3.fr	dcache-se-doma.desy.de	dclxwp2dlds1.gsi.de	eoseulake.cern.ch	lapp-testse01.in2p3.fr	lobster10.grid.surfsara.nl	t2-dpm-dome.na.infn.it	xrootd.pic.es
ccdcalitest10.in2p3.fr	-	100.00%	100.00%	100.00%	95.00%	100.00%	70.18%	100.00%
dcache-se-doma.desy.de	100.00%	-	100.00%	100.00%	96.43%	100.00%	77.50%	0%
dclxwp2dlds1.gsi.de	100.00%	100.00%	-	100.00%	92.86%	100.00%	100.00%	100.00%
eoseulake.cern.ch	0%	100.00%	0%	-	100.00%	0%	82.50%	0%
lapp-testse01.in2p3.fr	100.00%	100.00%	100.00%	100.00%	-	100.00%	62.22%	100.00%
lobster10.grid.surfsara.nl	100.00%	50.00%	100.00%	100.00%	83.33%	-	75.00%	100.00%
t2-dpm-dome.na.infn.it	100.00%	66.67%	100.00%	100.00%	100.00%	100.00%	-	100.00%
xrootd.pic.es	100.00%	100.00%	0%	100.00%	84.09%	100.00%	72.50%	-



Funded by the European Union's Horizon 2020 - Grant N° 824064

Thanks for listening!

Funded by the European Union's
Horizon 2020 - Grant N° 824064