

The logo graphic for ESCAPE features a stylized blue starburst at the top, a blue orbital path with a yellow dot at the bottom, and the word "ESCAPE" in large, bold, dark blue capital letters in the center.

ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

ESCAPE - A dive into a Datalake for Open Science

Xavier Espinal (CERN) - ESCAPE WP2 leader



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

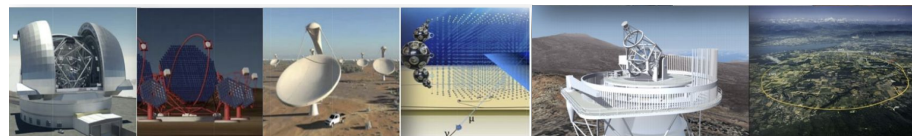
ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.



Science Projects



- 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

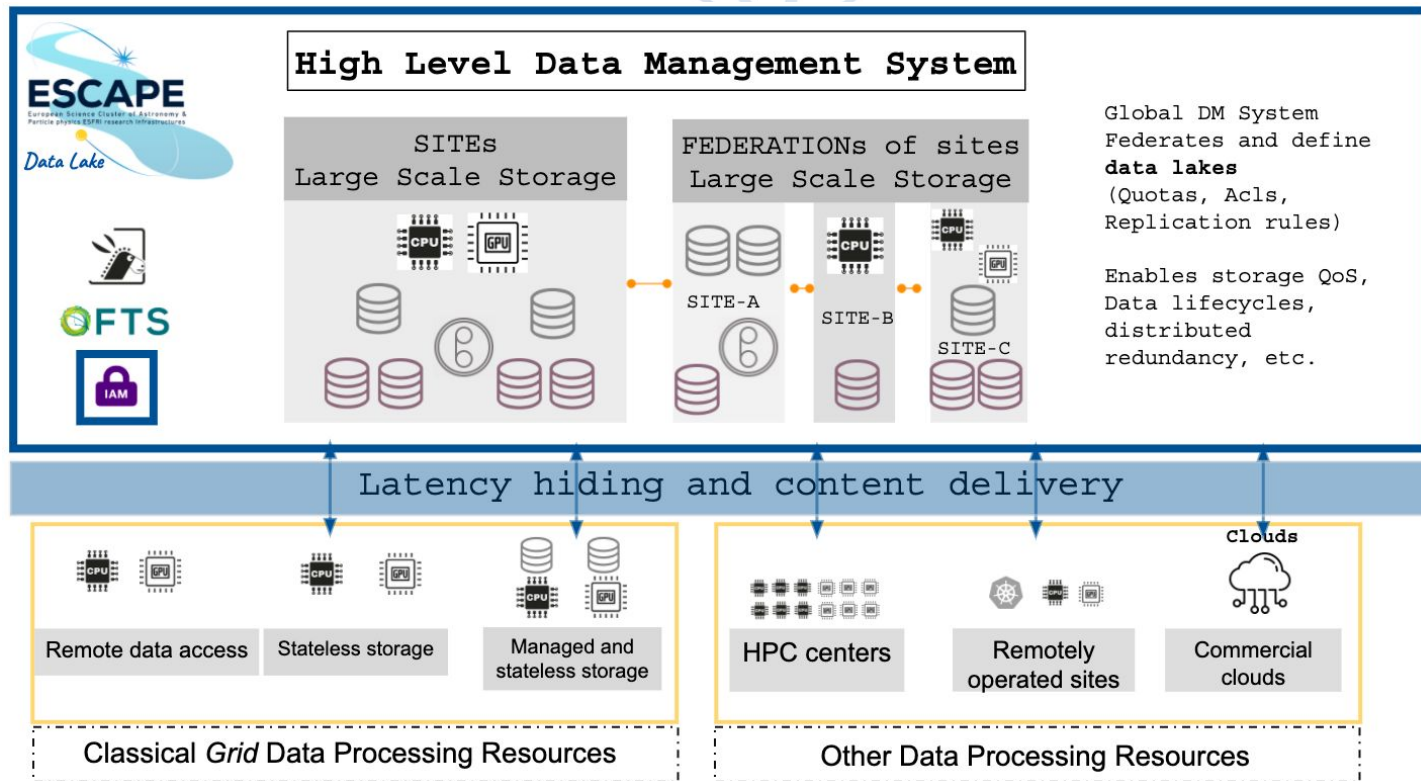


rijksuniversiteit
groningen



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 Lake

Orchestrator



Rucio Server

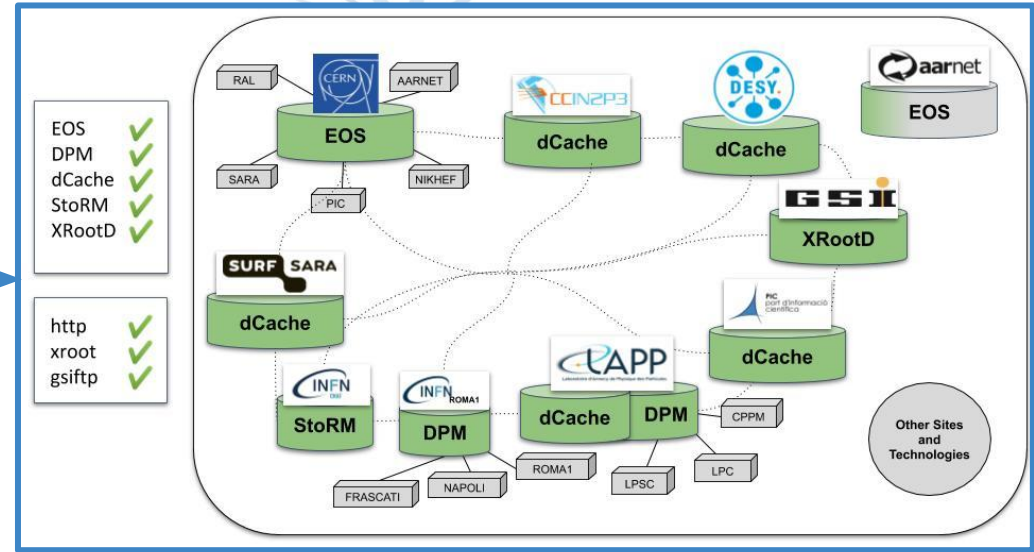
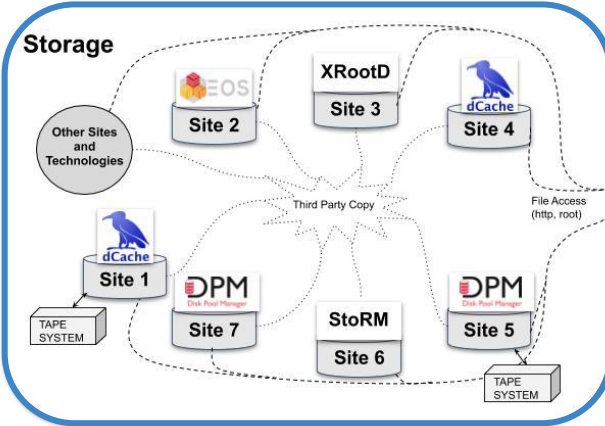
Middleware



Storage

CACHING
SOLUTION

NETWORK
OPTIMIZATION



EOS ✓
DPM ✓
dCache ✓
StoRM ✓
XRootD ✓

http ✓
xroot ✓
gsiftp ✓

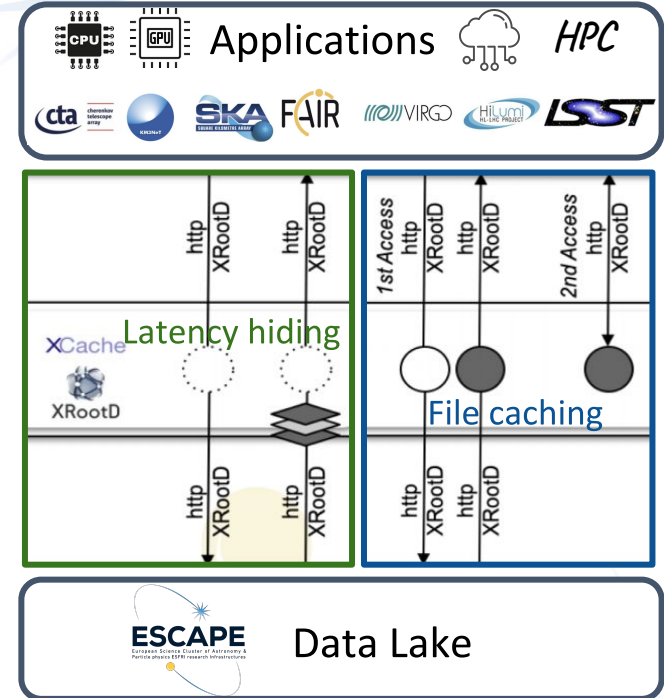
- Hiding complexity and providing transparent access to data
- Heterogeneous federated storage and operations model
- Some centers joining even if not funded by ESCAPE

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



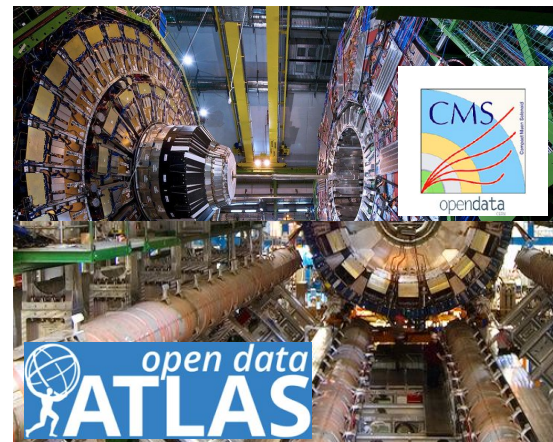
A word on Content Delivery and Caching

- 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 facilitate ingress/egress of data with heterogeneous computing resources: Commercial Clouds and HPCs



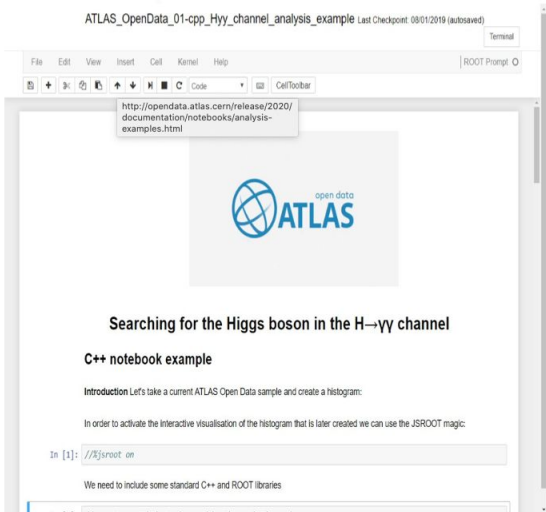
Data and Data access in the ESCAPE Data Lake

- 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)



Data and Data access in the ESCAPE Data Lake

ATLAS analysis demo



Credits: Stephane jezequel (LAPP)

```
// Connect to the ATLAS Open Data website repository */
// BEFORE ESCAPE : TString path = "https://atlas-opendata.web.cern.ch/atlas-opendata/samples/2020/GamGam/";
TString path = " root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/";

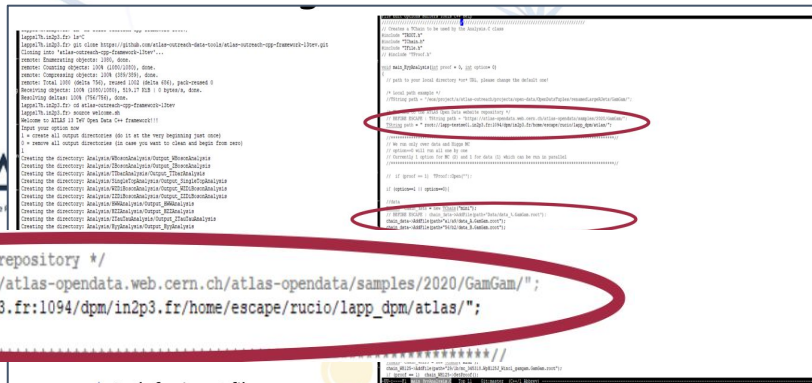
//*****
```



► Produce $H \rightarrow \gamma\gamma$ plot

★ ESCAPE exercise :

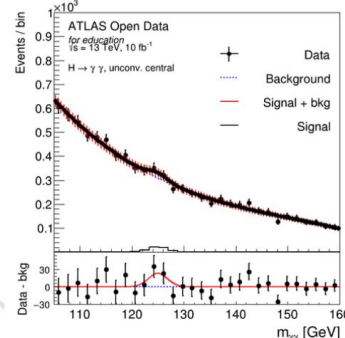
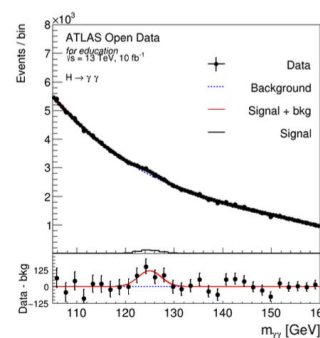
- Upload files to ESCAPE datalake with rucio client
- Adapt file access to ESCAPE datalake
- Produce plots



★ Path for input files :

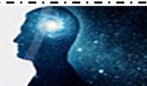
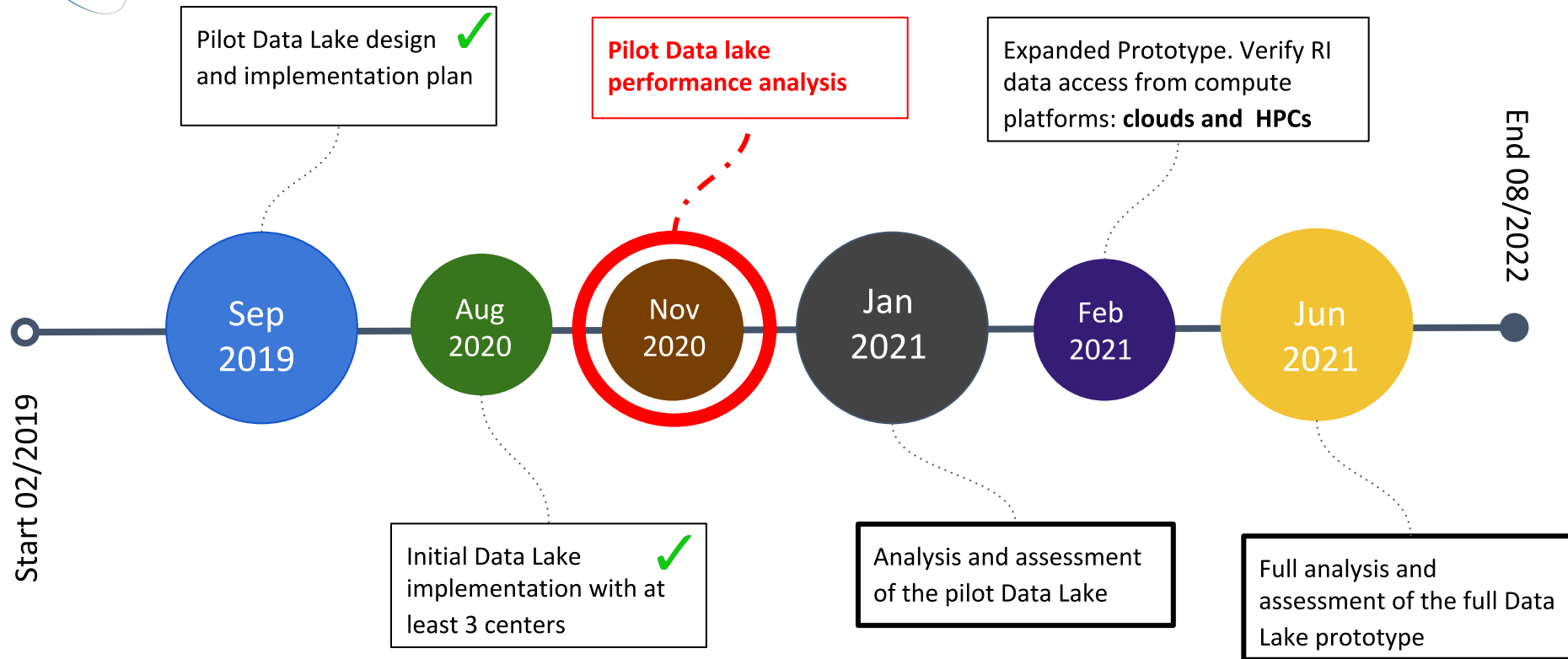
- Can be automatised decoded by just providing dataset and location
- Pattern can nbe built to include xcache in the path

★ Post processing root code to make plots



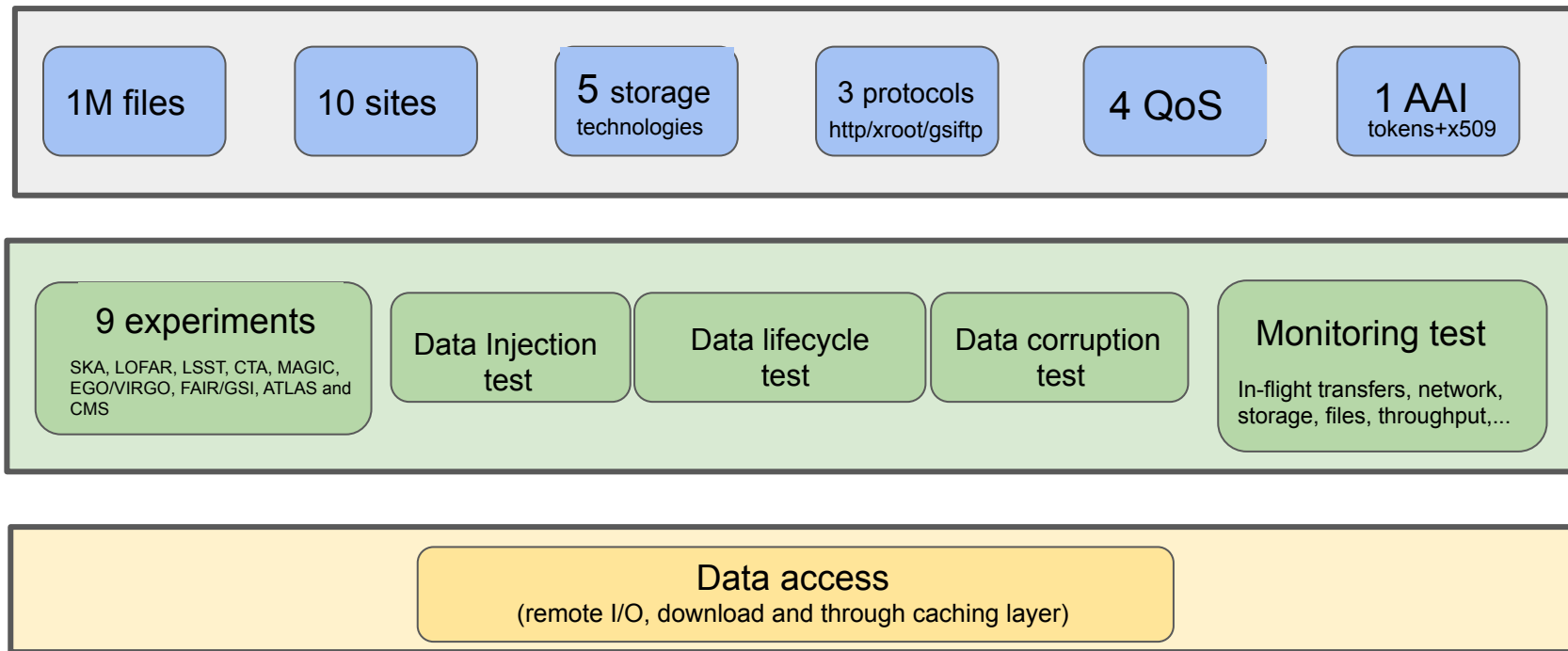


Some important milestones



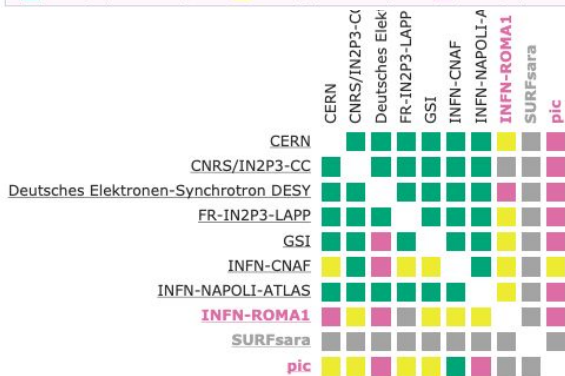
Pilot Data Lake Full Dress Rehearsal

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 Mesh Config - ESCAPE IPv4 Bandwidth - Throughput

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



Successful Transfers Percentage (days)

Source \ Destination	codaltest10.in2p3.fr	dcache-door-doma11.desy.de	dclwep2bda1.gsi.de	door05.pic.es	eessulake.cern.ch	lapp-dcache01.in2p3.fr	lapp-esc02.in2p3.fr	lapp-testest01.in2p3.fr	webdav.grid.sara.nl	xfer.craef.info.it
codaltest10.in2p3.fr	-	0%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	14.65%
dcache-door-doma11.desy.de	100.00%	-	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	18.18%
dclwep2bda1.gsi.de	100.00%	94.43%	-	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%
eessulake.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-testest01.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	codaltest10.in2p3.fr	dcache-door-doma11.desy.de	dclwep2bda1.gsi.de	eessulake.cern.ch	lapp-testest01.in2p3.fr	lubber10.grid.surfsara.nl	l2-dpm-dome.na.infn.it	xsroot.pic.es
codaltest10.in2p3.fr	-	100.00%	100.00%	100.00%	95.00%	100.00%	70.18%	100.00%
dcache-door-doma11.desy.de	100.00%	-	100.00%	100.00%	96.43%	100.00%	77.50%	0%
dclwep2bda1.gsi.de	100.00%	100.00%	-	100.00%	92.86%	100.00%	100.00%	100.00%
eessulake.cern.ch	0%	100.00%	0%	-	100.00%	82.58%	82.22%	0%
lapp-testest01.in2p3.fr	100.00%	100.00%	100.00%	100.00%	-	100.00%	82.22%	100.00%
lubber10.grid.surfsara.nl	100.00%	80.00%	100.00%	100.00%	83.33%	-	75.00%	100.00%
l2-dpm-dome.na.infn.it	100.00%	66.67%	100.00%	100.00%	100.00%	100.00%	-	100.00%
xsroot.pic.es	100.00%	100.00%	0%	100.00%	84.09%	100.00%	72.59%	-





Thanks for listening!

