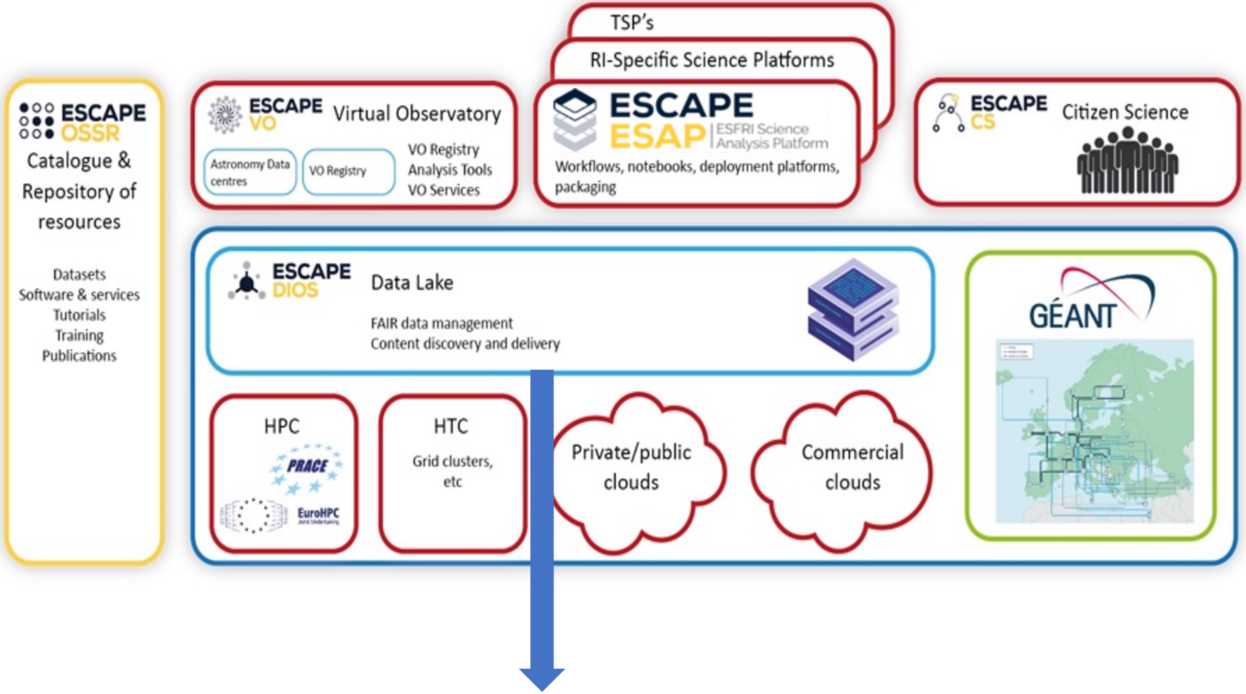
ESCAPE
"Large science experiments and networking futures"

Yan Grange, Xavier Espinal, on behalf of ESCAPE WP2

11th GEANT SIG NGN, Prague, 20th April 2023

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.



The ESCAPE Scientific Data Lake is a reliable, policy-driven, distributed data infrastructure. Capable of managing Exabyte-scale data sets, and able to deliver data on-demand at low latency to all types of processing facilities

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 Grant Agreement no. 824064.

The ESCAPE Scientific Data Lake is a reliable, policy-driven, distributed data infrastructure. Capable of managing Exabyte-scale data sets, and able to deliver data on-demand at low latency to all types of processing facilities

Services operated by the ESCAPE partner institutes

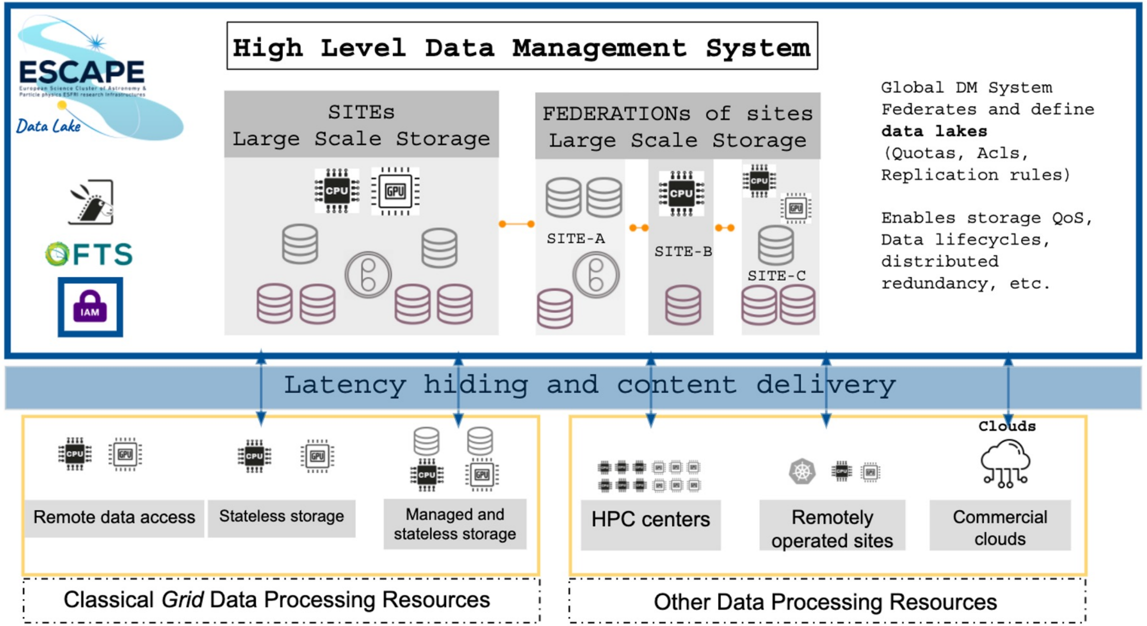
Petabyte scale storage: DESY, SURF-SARA, IN2P3-CC, CERN, IFAE-PIC, LAPP, GSI and INFN (CNAF, ROMA and Napoli)

Data management and storage orchestration (Rucio)

File transfer and data movement services (FTS)

Global Data Lake Information System (CRIC)

ESCAPE IAM: common Auth/Authz/IM (AAI)



Xavier Espinal
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 Grant Agreement no. 824064.

From a Data Lake Pilot to a full Prototype (1/2)

WP2 work plan focused on a continuous assessment and evolution of the pilot Data Lake, with the target to meet ESFRI/RI requirements and resulting in a fully working system

- **Token-based authentication** boosted its integration in the several layers of the Data Lake infrastructure: Rucio, FTS, storages (wip) and integration with other AAI *providers.* Easing user experience with a single and global authentication point
- **Data life-cycle accommodation** ESFRI/RIs users are able to define data replication rules, lifetimes, access policies, data location and storage *quality of service* (adjusting storage cost with data value)
- **Webdav/HTTP** promoted to be the de-facto standard in the Data Lake. The widespread knowledge of HTTP protocols provide a flexible way to interact and integrate with other storage resources, also eases data access from heterogeneous compute platforms and end-user devices
- **Data Management (Rucio) Evolution and Consolidation** channeling feedback from the new scientific communities using Rucio. Discussions on extending metadata capabilities together with ESO. Two extra ESFRI/RI private Rucio instances in operation for SKA and CTA, harmonically using the same global Data Lake storage infrastructure

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 Grant Agreement no. 824064.

From a Data Lake Pilot to a full Prototype (2/2)

WP2 work plan focused on a continuous assessment and evolution of the pilot Data Lake, with the target to meet ESFRI/RI requirements and resulting in a fully working system

- **Enlarged Data Lake monitoring capabilities** providing real time follow up for data transfers, automated test suite results, resources usage
- **Active Deployment and Operations (DepOps) team** early in the project identified need to share expertise, organised via a well-established meeting. Crucial to consolidate the infrastructure, to foster knowledge transfer and to prepare and drive the data challenges
- **Expanded Data Lake capabilities with user environments** the ***Data Lake as a Service*** product provides to the end users increased data browse/download/**upload** capabilities, trigger data movement, integrate with local storage, leverage storage caches, etc. Extending functionalities of Analysis Platforms (in conjunction with WP5), and to leverage computing infrastructures (ie. local batch systems and external resource providers)
- **Integration of heterogeneous resources** has been demonstrated, Data Lake interfacing with commercial clouds, public clouds and HPCs

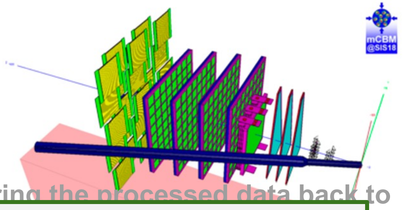
DIOS work plan brought together scientific communities addressing collective goals in a common data infrastructure. The various Data Challenges certified the infrastructure as a fully working system

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 Grant Agreement no. 824064.

Putting the system to work: **Data and Analysis Challenges** (1/3)

FAIR

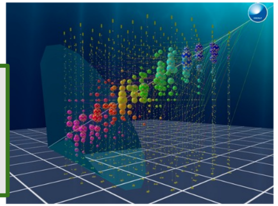
- **Registration of RAW data acquired by the mCBM detector on FAIR-ROOT**
- Ingestion and replication of simulated R3B data
- Ingestion and replication of simulated and digitised raw PANDA fallback data
- Particle-transport and digitisation of Monte-Carlo events
- Live ingestion of simulated data
- **Retrieval of stored RAW data from the data-lake, processing of the data and storing the processed data back to**



Raw data injected, stored and preserved in the DL. Data processed by users, results are stored back in the DL.

KM3NeT

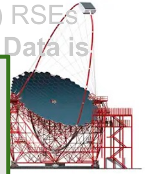
- **Ingestion of raw data from the storage at the KM3Net shore station to the Data Lake, and policy-based data replication across the Data Lake infrastructure**



Offload data from the storage buffer in the coast, replicate across sites, run data calibration, store back. Data product ready for user consumption

cta cherenkov telescope array

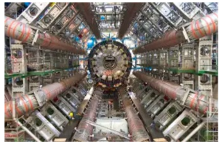
- Long-haul transfer and replication. CTA-RUCIO @PIC: non-deterministic (La Palma) and deterministic (PIC) RSEs
- **Data reprocessing. Primary data stored and findable in the datalake (using the CTA Rucio instance). Data is**



Full data re-processing workflow: data injection and distribution, data processing and results consolidation

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 Grant Agreement no. 824064.

Putting the system to work: **Data and Analysis Challenges** (2/3)

HiLumi HL-LHC PROJECT ATLAS EXPERIMENT	● Exercises (data production, replication and documentation) before and during the DAC21. Include the creation of datasets for real-kind final user analysis examples using current open-access datasets. ~200*10 = 2000 files uploaded in the Datalake. Two copies of such files (rules) into at least two RSE's ● **User analysis pipeline tests on experimental particle physics by using augmented open data** ... a de *Large experiment demonstrating open data capabilities* 
MAGIC Major Atmospheric Gamma Imaging Cerenkov Telescopes	● Long haul raw data ingestion and replication. Data is successfully transferred from the ● telescope station and replicated to the Data Lake, file deleted on the telescope storage ● buffer. ● Data transfer monitored. Data can be discovered using the CTA-RUCIO instance. ... the reading access of the samples via ... ammapy library. *Data management from remote locations* 
LOFAR	● Ingestion of LOFAR data from a remote site to the Data Lake. Data transfer and replication into off-site storage, after successful replication delete data at the source ● Process data in the Data Lake at an external location, combine results with other astronomical data to produce a multiwavelength image. ● Include a read-only RSE to a location outside the data lake. Get data from there into the DL. ● Extending use cases by using larger files and leveraging several QoS, running all processing in the DLaaS, requiring a the availability of specific LOFAR software in the DLaaS.  *Full-cycle scientific data management and data processing*

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 Grant Agreement no. 824064.

Putting the system to work: **Data and Analysis Challenges** (3/3)

SKAO

- Data replication. Data in correct place in timely manner.
- **Long haul data replication. SKAO Rucio (Australia and South-Africa to UK RSEs), using the RUCIO SKA instance.**
- End-to-end proof of concept data lifecycle test, AUS/SA to northern hemisphere sites
- ’s science data challenge pipeline using data

Global-scale Data Management

HiLumi HL-LHC PROJECT / CMS

- **Multi purpose Analysis Facility PoC with data access via DASK (workload orchestrator) leveraging computing at Marconi (HPC) and large batch clusters**
- Access control for embargo data, test in CNAF and DESY
- Content delivery and caching: XCache Protocol Translation: xroot internal vs http External for Data Lake data transfer. Performance comparisons for Analysis workflows

DL interface with local and heterogeneous resources, CDN and caching

LSST / VERA C. RUBIN OBSERVATORY

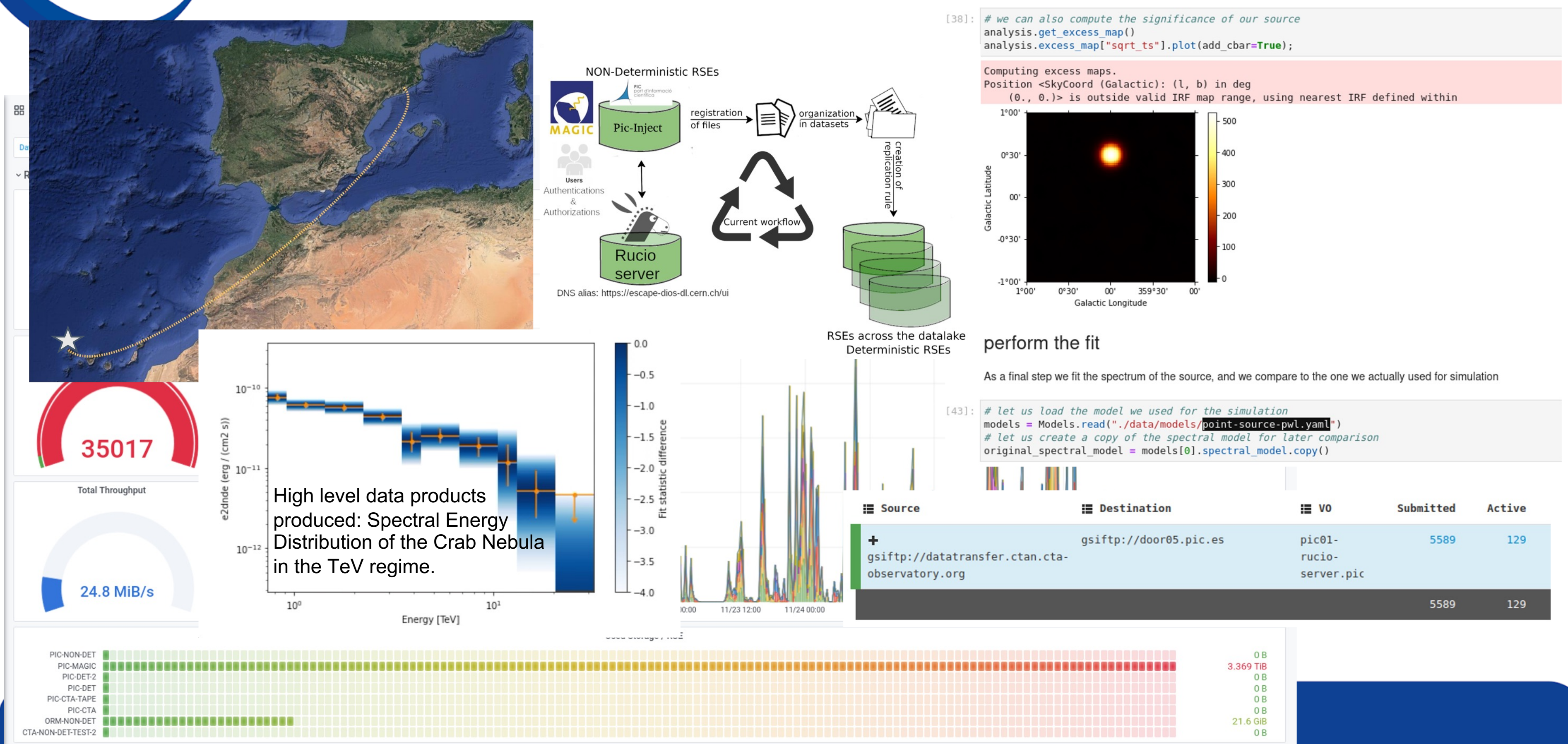
Simulate replication of one night’s worth of raw images data between two Vera C. Rubin data facilities, perform the exercise several times. Each iteration is composed of 15TB, 800k files, ideally to be replicated in 12 hours or less

Incorporate SLAC National Accelerator Laboratory (US) in the data replication chain (postponed)

Leverage telescope local storage data replication to fulfill daily data management cycles

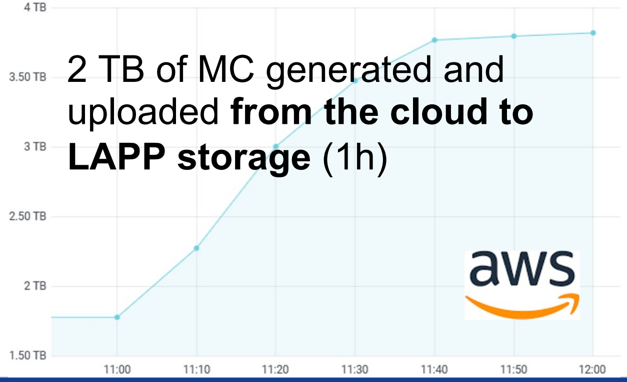
ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement no. 824064.

Example-1:Full-cycle long-range data workflows



Example-2: Integration with commercial and public clouds

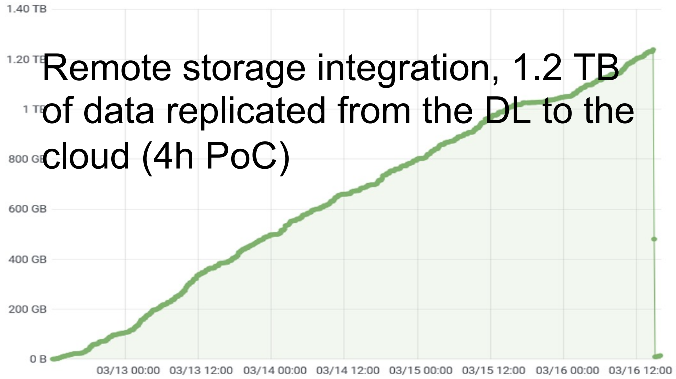
- Goal: assess integration of heterogeneous resources within the ESCAPE DL, Including CPU and storage using industry standards (Swift/S3 protocol)
- Exercise performed with the support of the Cloud Bank EU NGI project with fundings for AWS and Google Cloud Platform
- *Use case 1: Generation of CTA's Monte Carlo and results upload to the Data Lake*
- *Use case 2: Ad-hoc integration of Commercial Cloud storage in the Data Lake*



2 TB of MC generated and uploaded **from the cloud to LAPP storage** (1h)

egress cost for 2 TB

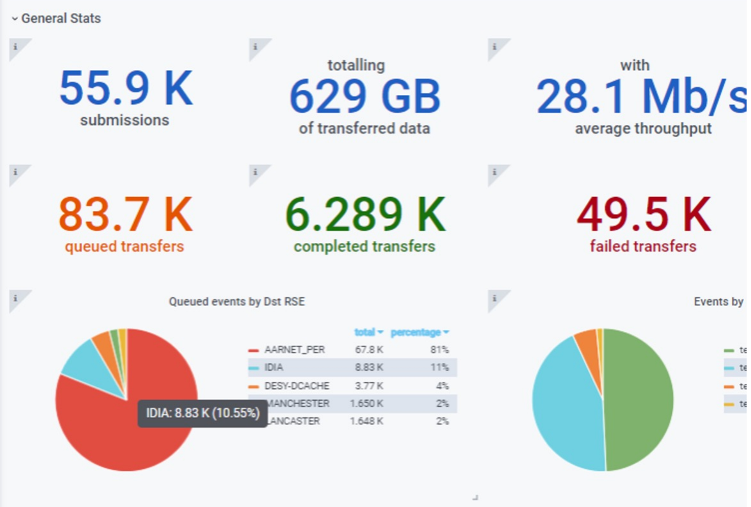
DataTransfer	$173.59
EC2	$21.33
Taxes	$38.99
Total	**$233.91**



Remote storage integration, 1.2 TB of data replicated from the DL to the cloud (4h PoC)

Full report available in the ESCAPE website: [link]

Example-3: Extreme distance data management in SKAO



Transfer Matrix - Replica Creation

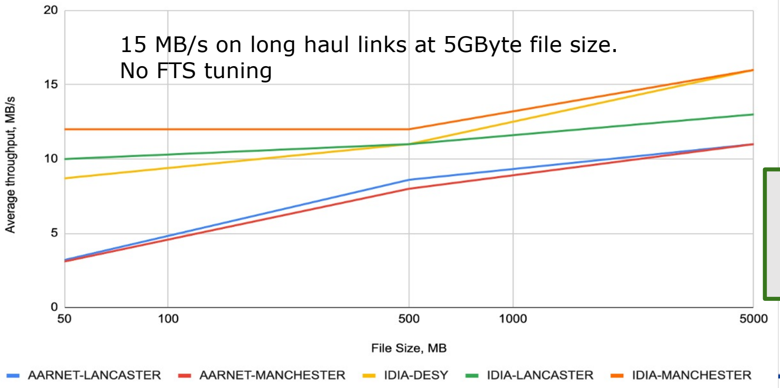
Src\Dst	LANCASTER	IDIA	DESY-DCACHE	AARNET_PER	MANCHESTER
MANCHESTER	100%	100%	100%	41%	NO DATA
LANCASTER	NO DATA	100%	100%	26%	100%
IDIA	100%	NO DATA	100%	NO DATA	100%
DESY-DCACHE	100%	100%	NO DATA	0%	100%
AARNET_PER	100%	NO DATA	7%	NO DATA	99%

Storage Endpoints
Moved SKA Rucio to tokens-only
Site include Canada, Sweden, Switzerland, Spain, Korea, Japan, China, Australia, Italy, UK
(Loss of SA site)

Building our own data lake using the technologies from ESCAPE

Test development within SKA Regional Centre development work - collaboration of teams from across globe.

Long-haul transfers from Australia and South Africa to European locations during DAC21



See talk by Rosie Bolton earlier today

Manual data transfers with Rucio began Feb 2021; automated tests still running, feeding live dashboard.

Conclusions

- Next-generation large-scale scientific experiments in the domains of particle physics and astronomy will have to make use of distributed storage infrastructure.
- Managing data at this scale, in terms of both size and distribution, requires an advanced data management system like the Scientific Data Lake.
- Long-distance network connectivity is an essential component of a distributed Data Lake.
- Especially in the case of multi-messenger astronomy, observations will often happen based on triggers of transient events, causing unplanned swarms of data that will show up as a peak in the network that is hard to predict and plan.
    - Also since this is one of a kind data, it is non-reproducible and especially this data will need to be replicated.