3rd ESCAPE DIOS Workshop - MAGIC Use Cases

A. Bruzzese (bruzzese@pic.es), J. Delgado, A. Vendrell, G. Merino

ESCAPE-The European Science Cluster of Astronomy & Particle Physics Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n1 824064

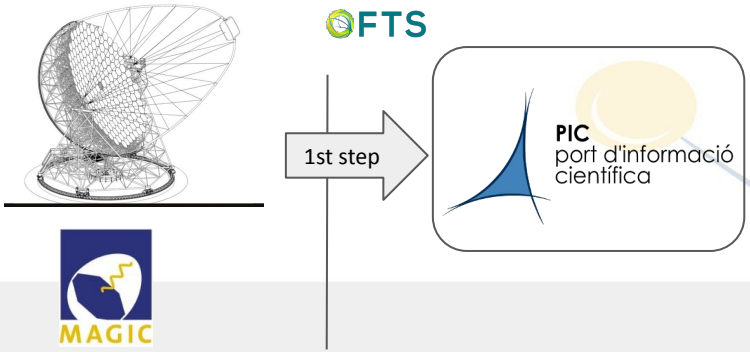
Contents

1) **MAGIC - Workflow**:
   a) Overview
   b) Use case 1 - CTA/PIC-Rucio
   c) Use case 2 - ESCAPE-Rucio
   d) Lessons Learned
   e) Future Work

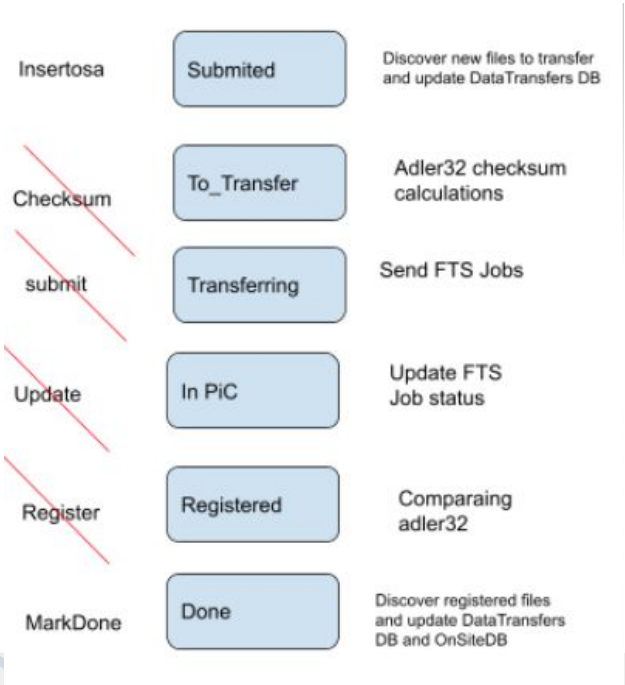
Overview: Magic telescope

Description:

- The MAGIC experiment started with a single telescope in October 2003
- Upgraded with a second twin telescope in 2009
- The MAGIC Telescopes are dedicated to the observation of gamma rays from galactic and extragalactic sources in the very high energy range.

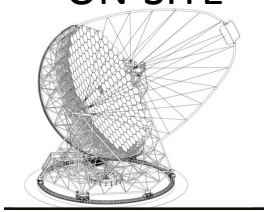


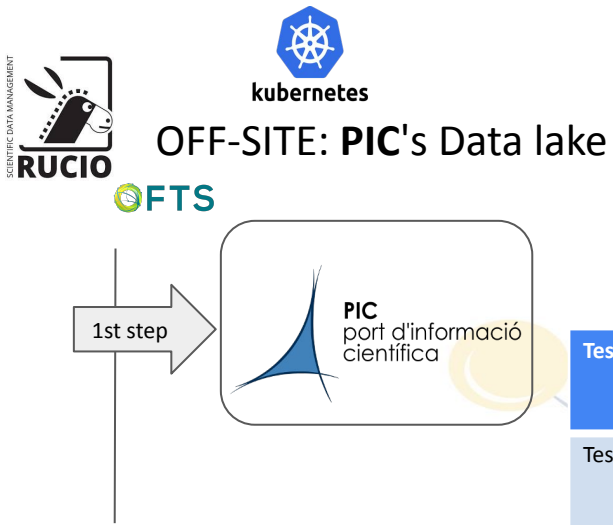
CURRENT TRANSFER WORKFLOW

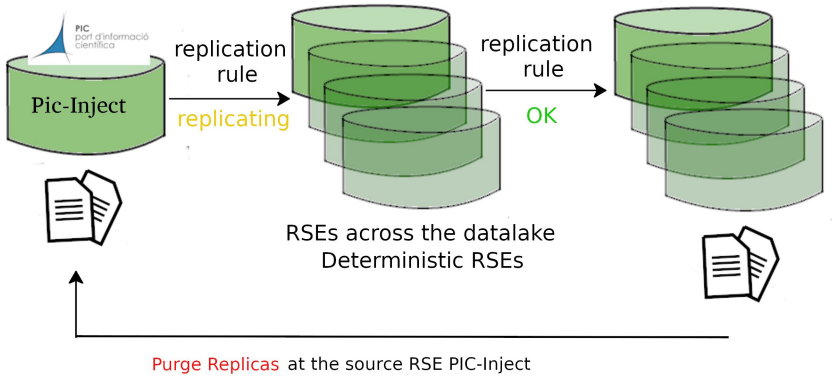


WP2 Use case 1: Long haul ingestion and replication

Description/Goal: Automatic detection and transfers of MAGIC data from a remote site (RSE non deterministic, ‘on-site’) produced from observation at la Palma, transfer and replication in off-site RSEs (PIC) and after replication deletion of the data at origin. Once the FTS transfers are successfully done, Rucio triggers the deletion of the files from the origin (La Palma).







Test name	Protocol	Estimated #Files, data volume	STORAGE
Test 1	GSIFTP	631 GB	Disk
Test 2	GSIFTP	3155 GB	Disk

WP2 Use case: Long haul ingestion and replication - SETUP

- **Base dataset**
  - **Real** Magic Dataset
  - **Number** of files: 2763
  - **Total** Size: 631 GB, with mixed file size(1GB, MB)

- **MAGIC side**

| - Network configurations:
| Link: 1Gbps
| Host: grid.magic.iac
| Ports: according the endpoint configurations
| GridFTP
| IPs: disk doors and servers at PIC
| - Transfer endpoint configurations:
| Local User transfer
| GSIFTP
| - File System:
| Data folder /data/Other/rucio_tmp/ESCAPE-DC21/Dataset/

- **PIC side**

| - Transfer interface machine
| Host RUCIO client code “PIC Inject”
| Connections to CTA side
| User certificate
| - RUCIO server
| Kubernetes deploy
| Trusted certificates
| Orchestration rules
| - FTS server: fts01.pic.es & fts02.pic.es
| - Transfer endpoint configuration:
| gridftp.pic.es
| - File System:
| dCache pools: tokens for disk & tape
| - Grafana monitoring dashboard

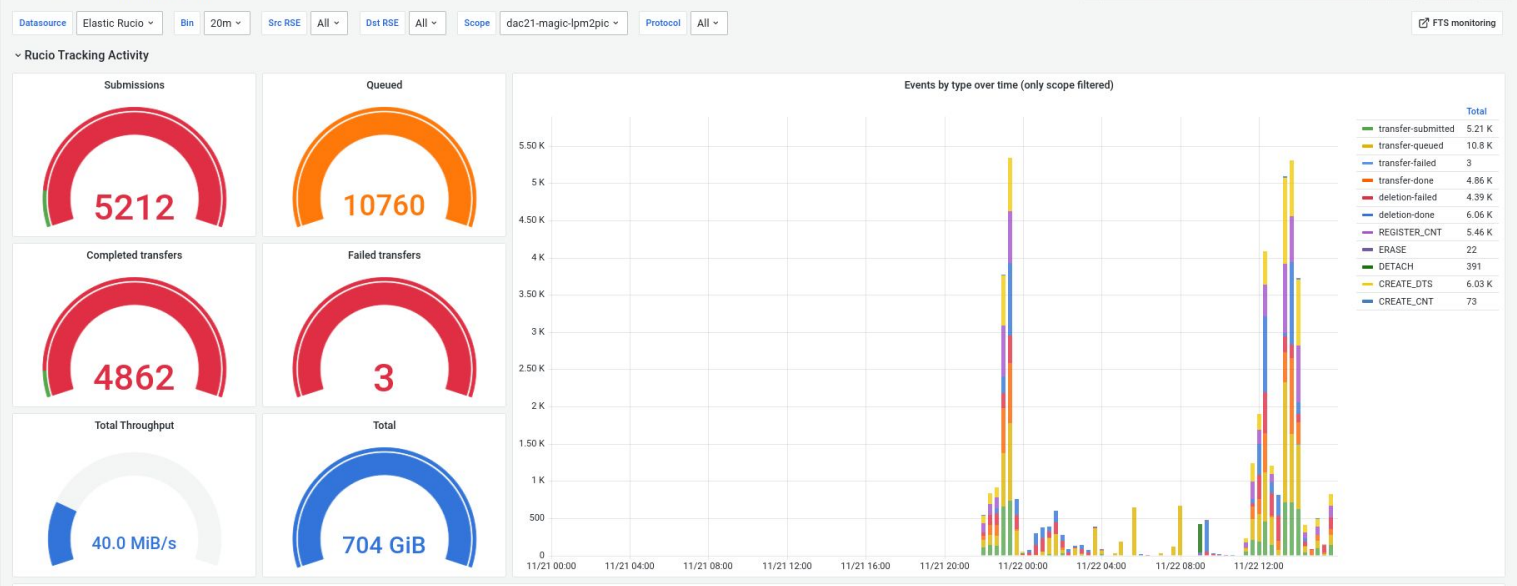
WP2 Use case: Long haul ingestion and replication

TEST1 - MAGIC endpoint GridFTP, Dataset 651GB, Destination endpoint: disk storage, CTA/PIC Rucio context

Experiment and point of contact	Usecase	Test name	Protocol	Method	Anticipated timeline Success metric	Estimated #Files, data volume	STORAGE	Replicas/RSEs
Agustin B, Jordi D, Gonzalo M	MAGIC01:	Test 1	GSIFTP	Python Cron	Data is successfully transferred, replicated, and file deleted on the origin RSE. Data transfer was monitored. Data can be discovered using the CTA-RUCIO instance	651 GB	Disk	CTA/PIC-RUCIO: non-deterministic and deterministic RSEs

WP2 Use case 1: Long haul ingestion and replication

TEST1 - MAGIC endpoint GridFTP, Dataset 651GB, Destination endpoint: disk storage, CTA/PIC Rucio context



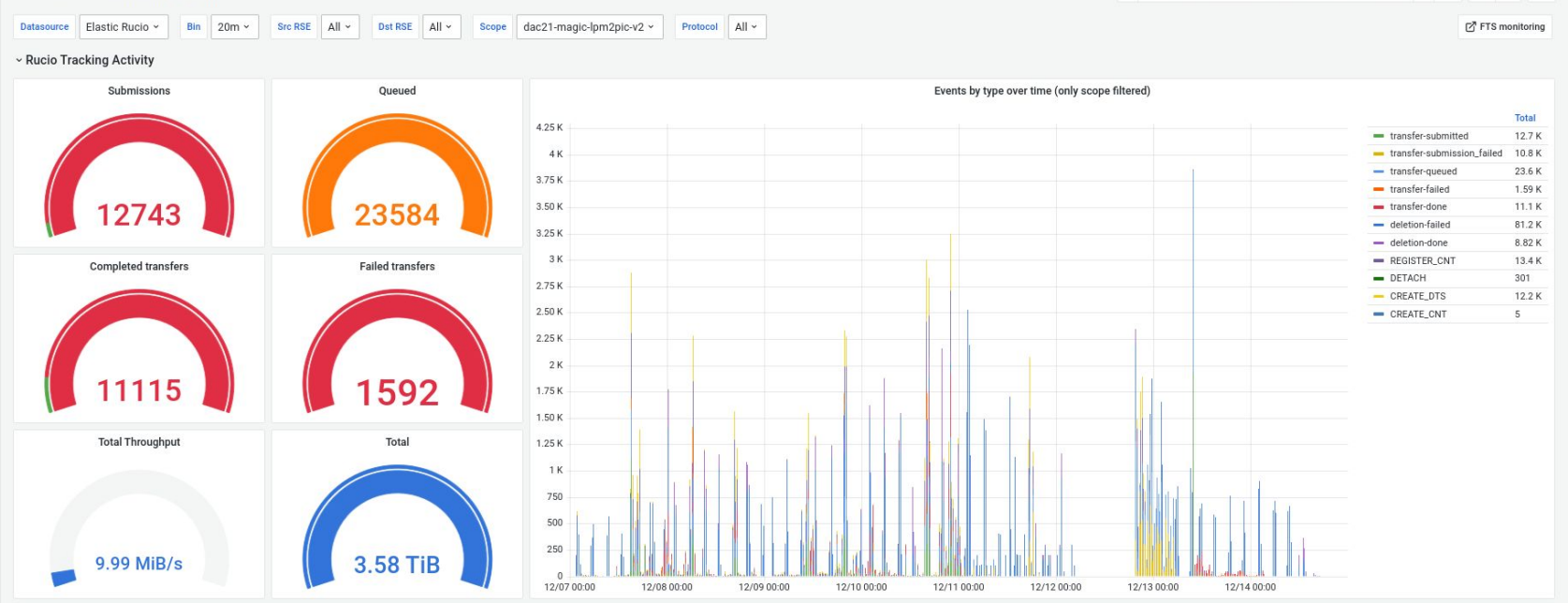
Test1 executed on 23rd of November, completed within the expected time frame (< 8h)

WP2 Use case: Long haul ingestion and replication

Experiment and point of contact	Usecase	Test name	Protocol	Method	Anticipated timeline Success metric	Estimated #Files, data volume	STORAGE	Replicas/RSEs
Agustin B, Jordi D, Gonzalo M	MAGIC01	Test 2	GSIFTP	Python Cron	Data is successfully transferred, replicated, and file deleted on the origin RSE. Data transfer was monitored. Data can be discovered using the CTA/PIC-RUCIO instance	3155 GB	Disk	CTA/PIC-RUCIO: non-deterministic and deterministic RSEs

WP2 Use case 1: Long haul ingestion and replication

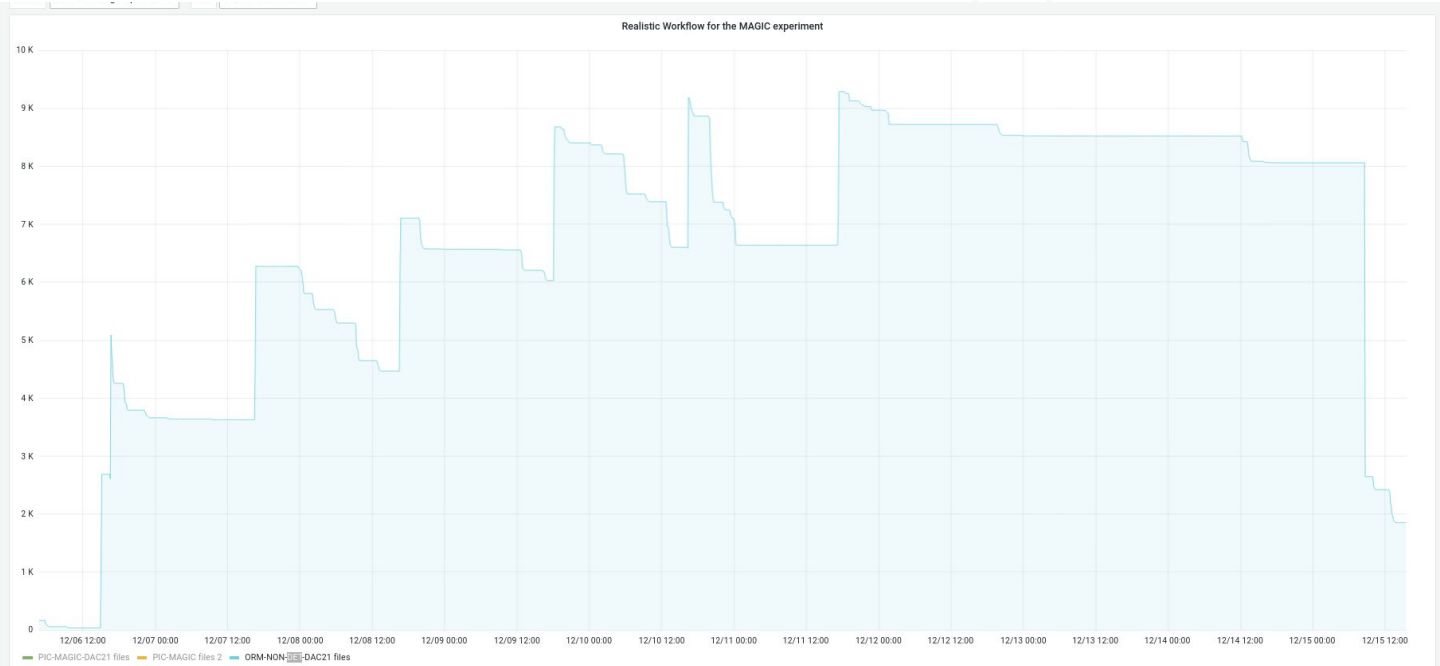
TEST2 - MAGIC endpoint GridFTP, Dataset 3155GB, Destination endpoint: disk storage, CTA/PIC Rucio context



Executed on 7th December, completed within the expected time frame (5 days)

WP2 Use case 1: Long haul ingestion and replication

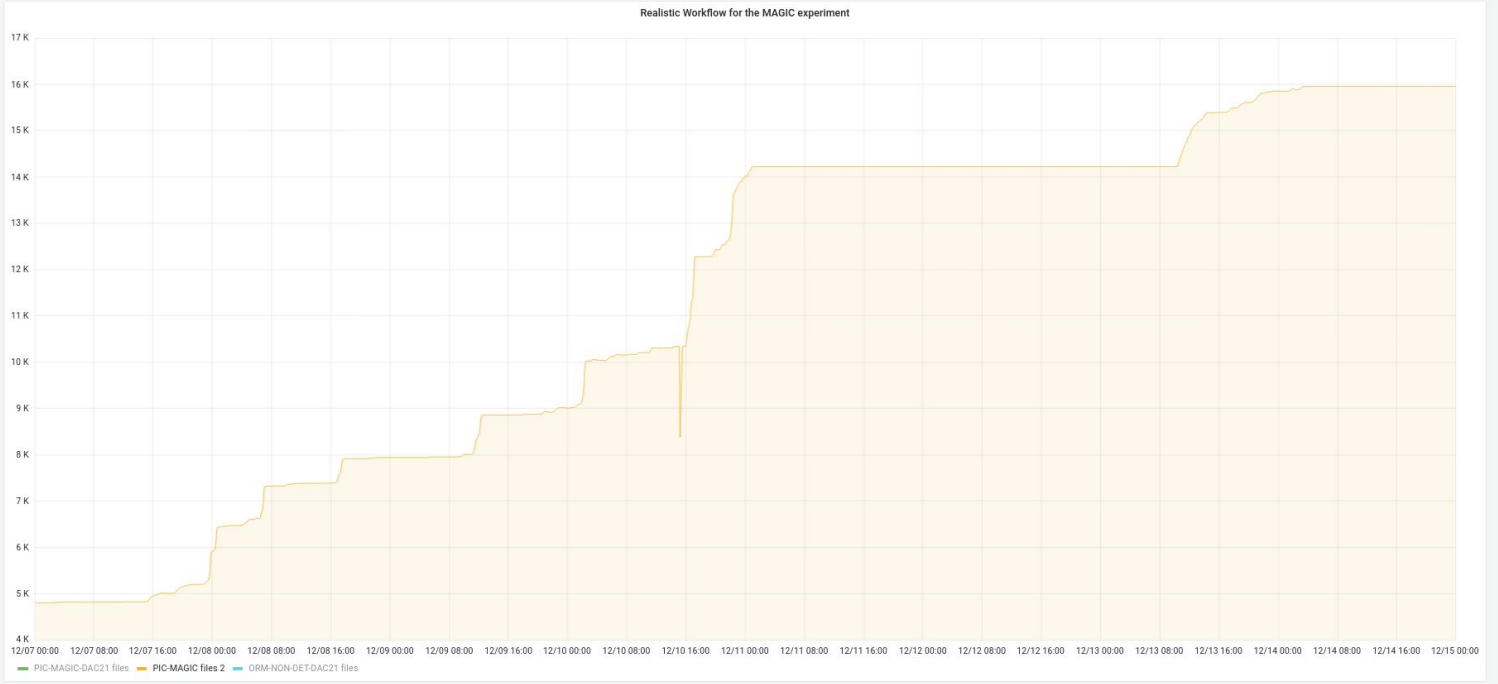
TEST2 - MAGIC endpoint GridFTP, Dataset 3155GB, Destination endpoint: disk storage, CTA/PIC Rucio context



Test1 executed on 7th December, completed within the expected time frame (< 5 days)

WP2 Use case 1: Long haul ingestion and replication

TEST2 - MAGIC endpoint GridFTP, Dataset 3155GB, Destination endpoint: disk storage, CTA/PIC Rucio context



Test1 executed on 7th December, completed within the expected time frame (< 5 days)

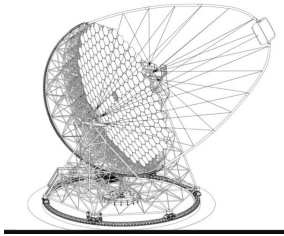
WP2 CTA Use case 2 : Data reprocessing

Description/Goal: We wanted implement a use case associated with GammaHub, an initiative that IFAE-PIC is developing in the context of WP5 and that aims to offer interactive analysis tools using the Gammapy package and the emerging standard in Gamma-ray astronomy, the DL3 format files.

DL3 (FITS file)

- Events Table
- Header (Observation Info)
- Instrument Response Function (IRF)

ON-SITE



MAGIC

RUCIO

kubernetes

OFF-SITE: **ESCAPE**'s Data lake

FTS

1st step

PIC port d'informació científica

2nd step

jupyter

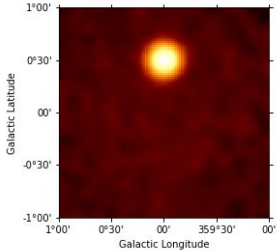
WP2 Use case 2: Data reprocessing

TEST1 - MAGIC endpoint, Dataset 675MB (200 DL3 files), Destination endpoint: disk storage, ESCAPE Rucio context

```
[37]: # let's check just the first observations
counts = analysis.datasets["stacked"].counts
counts.smooth("0.05 deg").plot_interactive()
```

Select energy: 1.00e+02 GeV - 1.58e+02

Select stretch: linear / sqrt / log

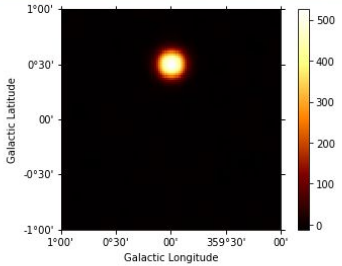



```
[38]: # we can also compute the significance of our source
analysis.get_excess_map()
analysis.excess_map["sqrt_ts"].plot(add_cbar=True);
```

```
Computing excess maps.
Position <SkyCoord (Galactic): (l, b) in deg
    (0., 0.)> is outside valid IRF map range, using nearest IRF defined within
```

```
[38]: # we can also compute the significance of our source
analysis.get_excess_map()
analysis.excess_map["sqrt_ts"].plot(add_cbar=True);
```

```
Computing excess maps.
Position <SkyCoord (Galactic): (l, b) in deg
    (0., 0.)> is outside valid IRF map range, using nearest IRF defined within
```



perform the fit

As a final step we fit the spectrum of the source, and we compare to the one we actually used for simulation

```
[43]: # let us load the model we used for the simulation
models = Models.read("./data/models/point-source-pwl.yaml")
# let us create a copy of the spectral model for later comparison
original_spectral_model = models[0].spectral_model.copy()
```

Test1 executed on 30rd of November, completed within the expected time frame (< 24h)

WP2 MAGIC Use case: Long haul ingestion and replication and Data reprocessing Summary

Use Case	Test name	Context	Estimated #Files, data volume	STORAGE	Replicas/RSEs	Results	Observations
MAGIC01	Test 1	CTA/PIC-RUCIO	631 GB	Disk	non-deterministic and deterministic RSEs	Completed successfully.	
	Test 2	CTA/PIC-RUCIO	3155 GB	Disk		Completed successfully.	We had to repeat the test due to some deletions in the target RSE.
MAGIC02	Test 1	ESCAPE-RUCIO	675MB			Completed successfully.	In order to have the gammapy library on DLaaS, gcc has to be installed.

- Results of tests: (notebook)

Lessons Learned

Things that worked and have been identified to help ESFRI/RIs on Data Management and Data Access:

1. RUCIO can efficiently orchestrate the data transfer from ORM onsite to offsite, for long haul distances. The orchestration can work automatically to discover and transfer new files
2. Both RSEs/endpoints on a RUCIO transfer need to work with the same protocol
3. Data replication to N RSEs can be configured and deletion policies adjusted
4. Addition of complex metadata
5. Deletion at source

Lessons Learned

Did you identified security issues? Any specific security related worries to name (present or future) ?

1. We have not encountered any security problems. However, we are interested embargo data protection between experiments.

Identified barriers to adopt the DIOS model, services or tools

1. Apart from the great complexity involved in the deployment of the architecture presented in DIOS (K8s, RUCIO, ES, FTS), we think that functionalities such as scopes elimination can be very useful.

Lessons Learned

Is your ESFRI/RI Interested in a longer term existence of an ESCAPE or an ESCAPE-like infrastructure?

1. We believe that this kind of collaboration would be very positive in order to reach best practices in data manipulation and in the development of new functionalities.

Is your ESFRI/RI interested establishing standing collaborations, channels, joint efforts? On which specific topics

1. Support channels such as Rucio's slack as well as the rocket chat channel have proven to be an efficient way to detect problems as well as to provide possible solutions, so we believe it would be very positive to maintain/create a channel with the ESFRI participants in WP2 .
2. MultiVo, Dirac, Token authentication

Future Work

1. Your ESFRI/RI plans regarding the technologies exposed in DIOS, will you consider continue exploring or adopting?

Deploy a new k8s cluster to provide a production environment for RUCIO

Test a real case transfer replica to other physical site, in order to test and measure the performance

Introduce an external DB to manage the file discovery

Update the DB with the times of occurrence of a change of transfer status

Acknowledgements

This success of this work was possible with help of many people

PIC Operations Team
- Vanessa Acín - Network
- Elena Planas - dCache
- Esther Acción - Enstore
- Ricard Cruz - Kubernetes
- Carles Acosta - xrootd

And also other people from ESCAPE partners
- Alba Vendrell
- Rizart Dona
- Riccardo di Maria
- Xavier Espinal

Thanks for listening!
Questions?

ESCAPE-The European Science Cluster of Astronomy & Particle Physics Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n1 824064