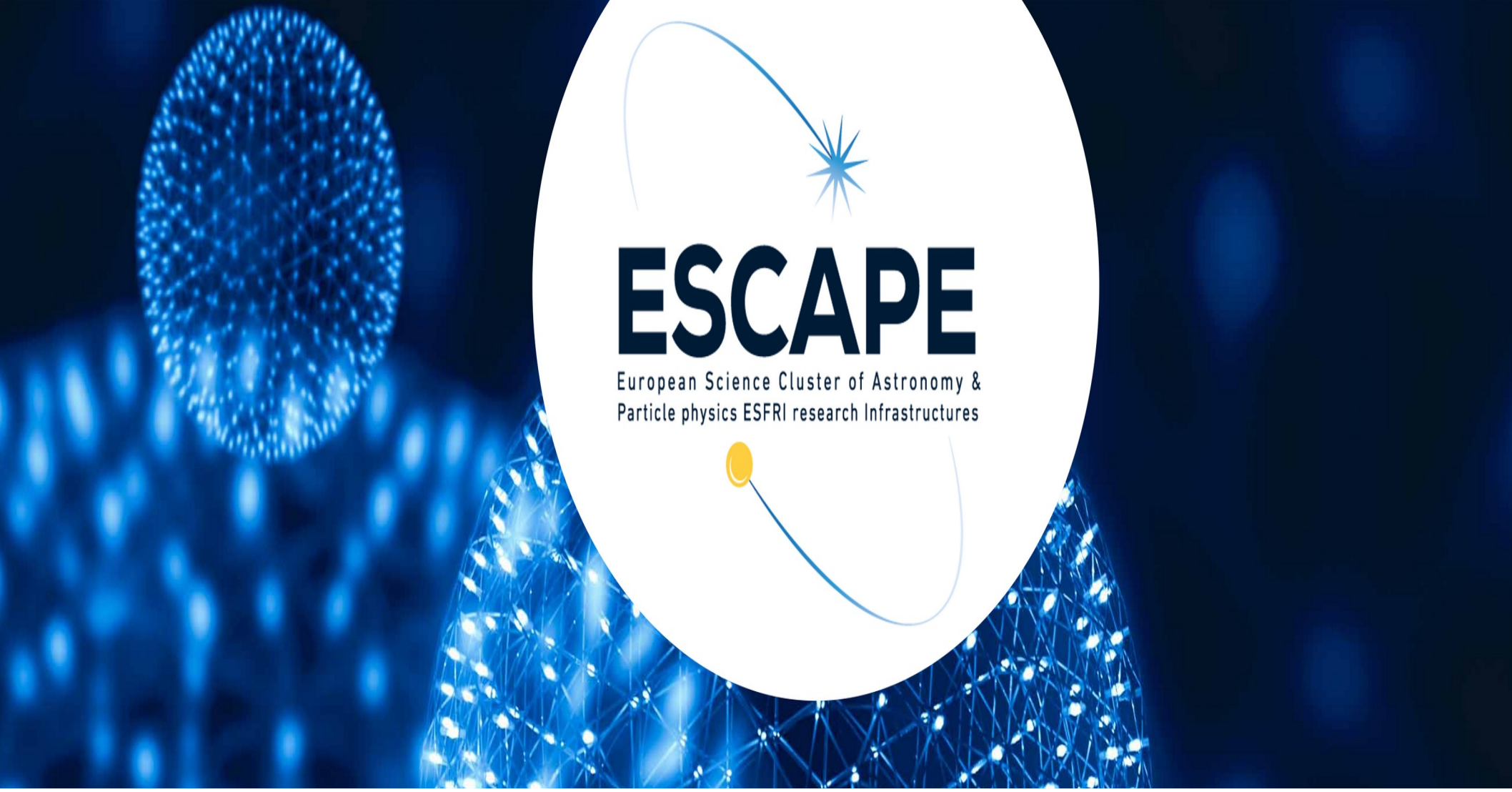


ESCAPE WP2

Agustin Bruzzese – PIC
bruzzese@pic.es

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

Institut de Física d'Altes Energies
EXCELENCIA SEVERO OCHOA
BIST Barcelona Institute of Science and Technology
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
cfp CIEMAT física de partículas

Port d' Informació Científica (PIC)

- **Background :**
  - PIC is a data center of excellence for scientific-data processing supporting scientific groups working in projects which require strong computing resources for the analysis of massive sets of distributed data.

- **General Aims :**

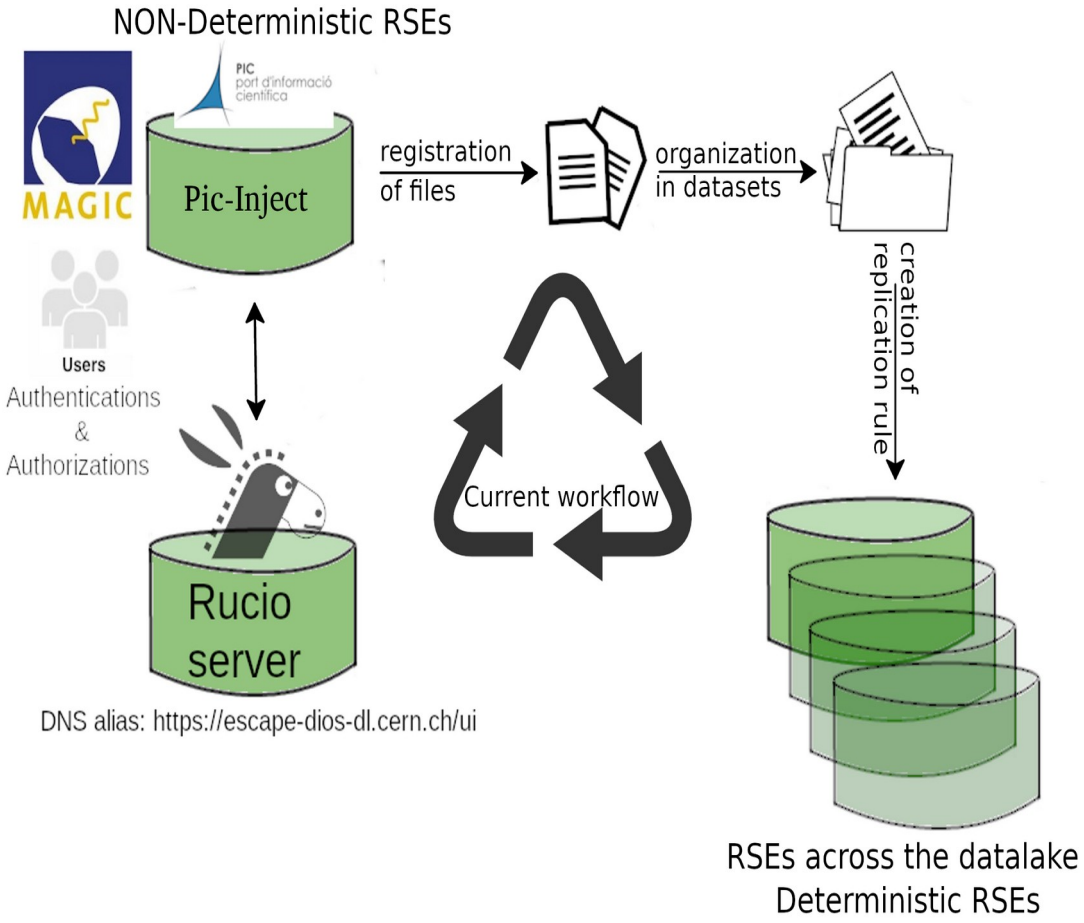
1) Develop solutions to handle the large data sets produced by Gamma ray telescopes :
2) Files streamed from the telescope to the Data Lake for permanent storage and access.

PIC
port d'informació
científica

PIC Workflow

Data Registration :

- Data is injected from the storage system located at the source (ORM La Palma in the future) :
- Our script looks for new incomming data
- Then it triggers the orchestration of these files replication among the datalake.



Rucio Storage Elements (RSEs)

1. **Deterministic RSE :** Rucio automatically manage storage, and creation of a path, only by knowing the scope and name of a data identifier (Ignoring the static prefix of the storage endpoint).
   - PIC-DCACHE

2. **NON-deterministic RSE :** requires full path on the storage to the file. It offers more flexibility in placing the files on storage.
   - PIC-INJECT

- Currently testing a **NON-deterministic RSE** configuration at PIC that allows to register files with their original path in the detector. Mimicking the future config of an RSE at ORM.

Current goals and aims

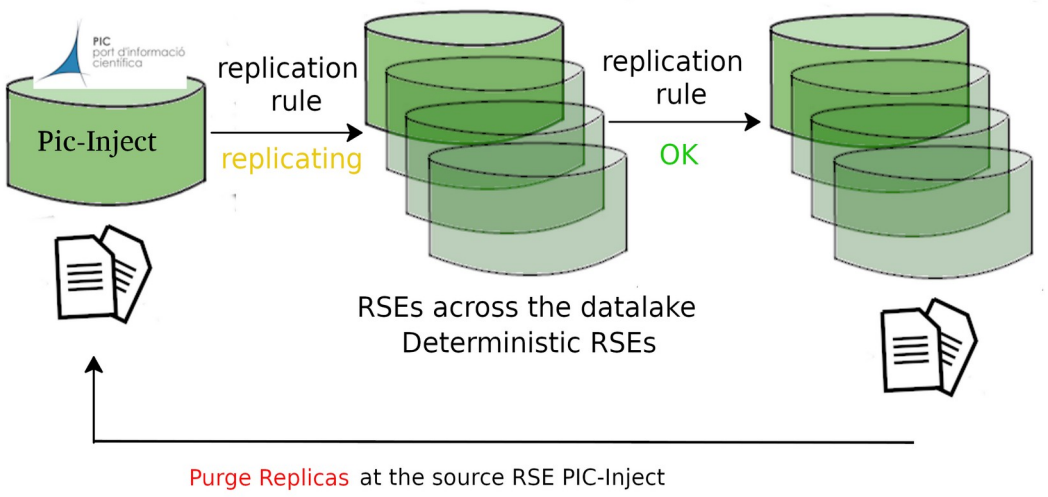
- **By means of RUCIO :**
  - Once a file has been successfully replicated to the destination RSE, delete it from the source.
  - Fully automate with Rucio the MAGIC ORM - PIC data flow, including origin replica deletion when the replica to the destination RSE is done.

PIC
port d'informació
científica

Update on the PIC Workflow

Replication Rules :

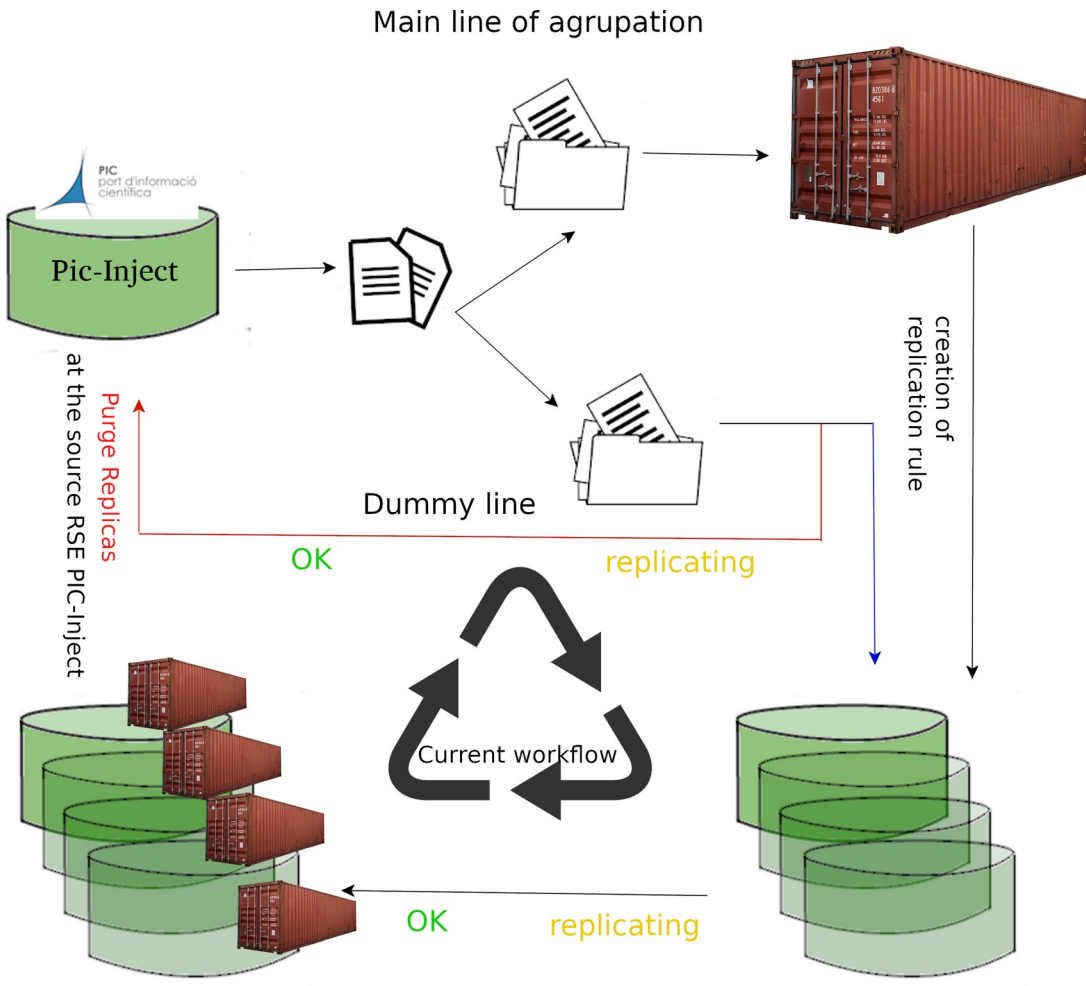
1. Sites – ESCAPE testbed:
   - PIC-INJECT (NON-deterministic) - PIC-DCACHE (deterministic)
2. Files are protected from deletion if they are related with an existing rule.
3. Delete the files once they have been replicated to its destination RSE



Deletion of original source RSE

Deletion of Files :

- Creation of additional rules :
  - Carrier Rules
- Functions :
  1. Just replicate new files
  2. Automatically deleted the rule
  3. Automatically purge the files from the source RSE
     - update_replication_rules



- **Conclusions :**
  By means of the RUCIO API, we automate the orchestration of the data:

  a) Files are replicated through different research facilities among the Data Lake.
  b) Once replicated, source files are removed.

PIC
port d'informació científica

Institut de Física d'Altes Energies

EXCELENCIA SEVERO OCHOA

BIST Barcelona Institute of Science and Technology

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

cfp CIEMAT física de partículas

Further Steps

Next Steps :

1. Configure INJECT RSE for gridftp endpoint in La Palma Observatory.

2. To implement ‘/’ in the naming of file, in order to archive more informative naming files and supporting pre-existing naming convention.
   - i.e. : std:RAW/M1/CrabNebula/2020_03_13/20200313_M1_05089889.008_D_CrabNebula-W0.40+215.raw.gz

3. Develop an algorithm that builds the folder path from the name-space of the file at the destination RSE.

PIC
port d'informació científica

Supplementary information

- **The scripts :**
  - Generate_random_folders.py :

    I. https://github.com/BruzzeseAgustin/Rucio_client_scripts/blob/master/Merge_Rucio_Scripts-v7.py

PIC
port d'informació
científica

- **Acknowledgements :**
  - Special thanks and congratulations to Aristeidis Fkiaras and his team.

- **Take Home Message :**
  - NON-Determinsitc RSEs are useful to register locally placed files.
  - Rules protect files from deletion