ESCAPE Extended Discussion Day

# WP2 + WP5 discussion

Yan Grange
28 September 2021

# DataLake as a Service (DLaaS)

- By now I will assume WP5 people know what a Datalake is (more or less), WP2 people what the ESAP is.
- Mainly as access layer to the Datalake for the normal user.
- Jupyter lab implementation (I'm sure WP5, you saw a demo, if not we should definitley schedule one; WP2: you saw at least two).
- DLaaS notebook provides access to data in the Datalake, staging and access to and of files via FUSE mounted file system.

# DLaaS integration extension

- Currently, the shopping basket can not be read from the DLaaS notebook. Working on an integration
  - Getting the JSON from the shopping basket in the notebook is straight forward, getting it from the JSON in the Rucio running in the notebook.
- Connected to that: DLaaS notebook provides access to the OIDC data, working on demoing that in a different notebook for ESAP (this is not part of this discussion)

# DAC21

- November 22-26 2021
- Experiment-led tests determined by the use cases relevant for our partners
- Focus on ensuring sustainability of the expertise and testing technical advances made during the past year
- DAC21 is our focussed assessment time for the pilot data lake
- Opportunity to explore the use cases and perspectives of different ESFRI partners

# DAC21 use cases for DLaaS

- DAC21 use cases, using DLaaS, and/or ESAP for data location (I hope I didn't miss any)
  - LOFAR processing use case. Taking data from Datalake, making an image, uploading image back to the datalake.
    - Ambition to extend this with combining the image with VO data to make it a multiwavelength image, analysis
  - CTA: processing data in the Datalake, use case on combining CTA and KM3Net data
- Two that use the DLaaS
  - ATLAS open data replication and analysis
  - CMS: Downloading, processing and putting data back in the data lake using the notebook

# Discussion and questions

- Several ESAP entities talking to Rucio?
    - Would be nice to demonstrate ESAP can be easily deployed, and that we can use the Rucio integration from multiple locations.
- Would be useful to have several Jupyter notebooks with the DLaaS and instrument specific software in them.
    - Would be nice if we could somehow mix and match the ESCAPE support tooling (DLaaS, ESAP shopping basket, ESAP Zenodo, etc) and the processing tooling (e.g. LOFAR software).
- Metadata for queries.