

SKAO

Datalake testing plans
Feb 24th, 2021



SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Rohini Joshi

On behalf of SKAO members

Plans (1/2)

- Running our own prototype Rucio instance
- Bigger, faster, better data injection
 - Simulate data injection possibly via registering data at a non-deterministic RSE
 - While we likely won't have individual file sizes that are $O(60\text{GB})$, we'd like to follow the issue of timeouts from the LOFAR use case for large files.
- Embargoed data
- Long-haul data transfers and how it can be made possible within the Rucio framework

Plans (2/2)

- Metadata (ESCAPE)
 - Concrete metadata use cases for SKA may not exist but we would like to look into this functionality alongside other ESFRIs like LOFAR and CTA.
- QoS lifecycle
Better monitoring of QoS in the data lake
- Data centric view
Could start by tweaking the indices stored in our ES DB and build a matching dashboard but this could perhaps also be done via improvements to Rucio Web UI.

Helpful enablers

Rucio deployment

- Daemon-level monitoring to better understand the inner workings
- Enabling metadata on the instance
 - This will further enable subscriptions

Rucio development

- Uploading to QoS (in progress)
 - Timestamps to define rules
 - Expiration timestamp
 - Creation timestamp
 - Policies
- https://rucio.readthedocs.io/en/latest/policy_packages.html

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope



Thank you

www.skatelescope.org