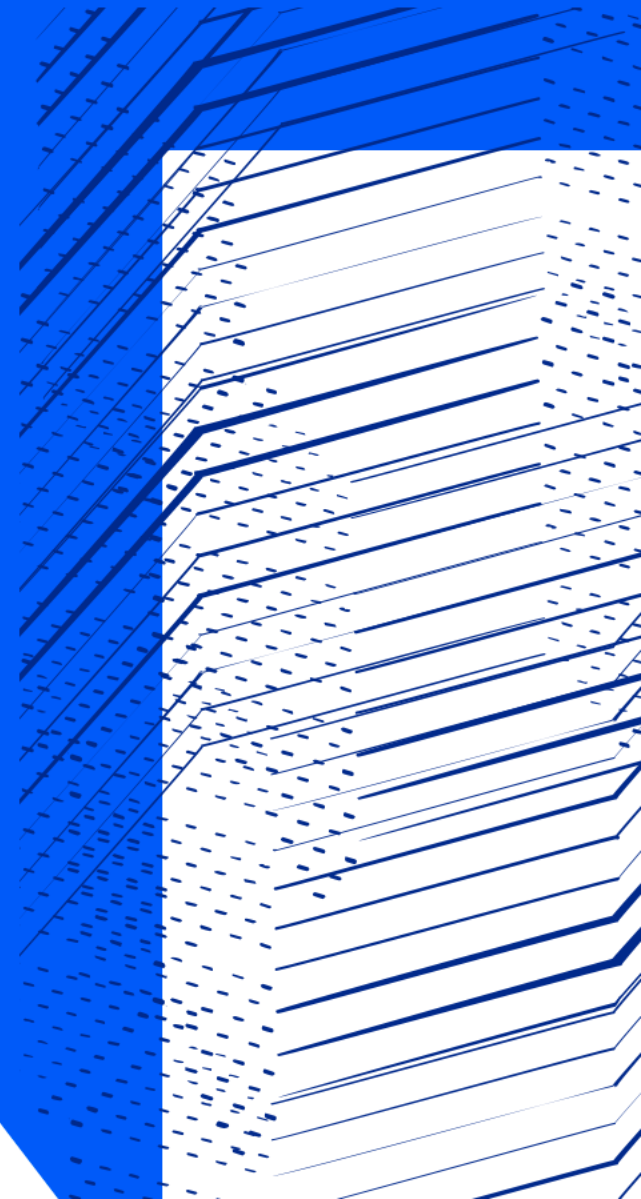




Science and  
Technology  
Facilities Council

# Rucio: Multi-VO support and latest developments

Alastair Dewhurst, with thanks to the  
Rucio Devs

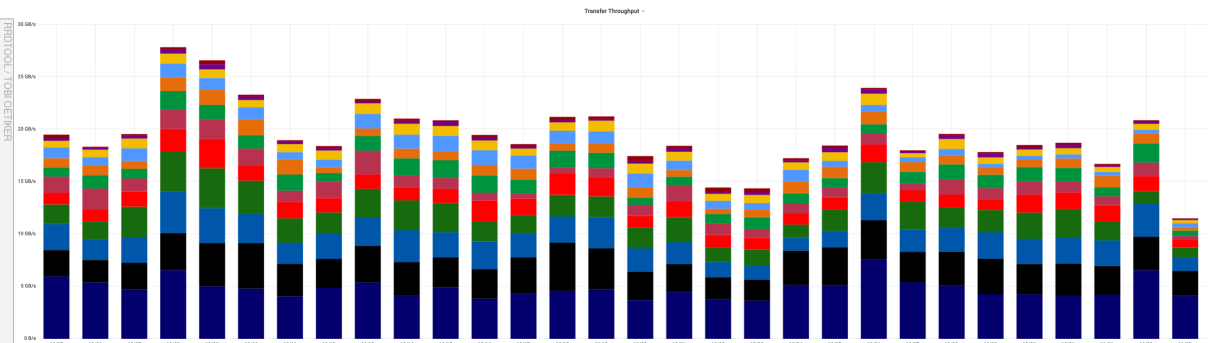
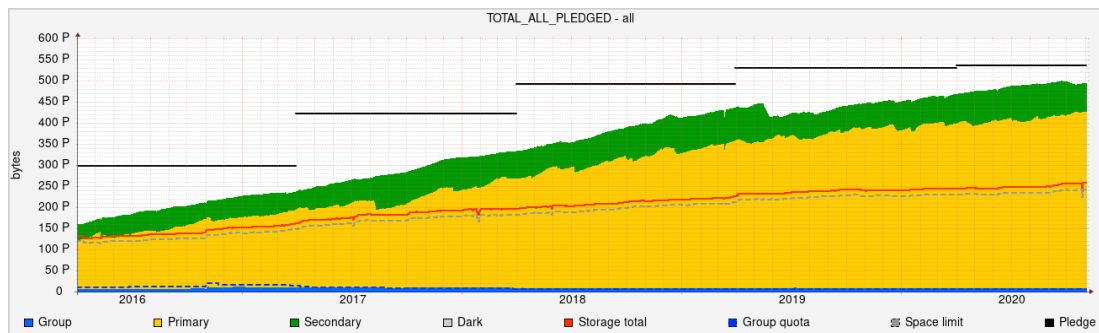


# What is Rucio?

- Rucio provides a complete and generic scientific data management service
  - Designed with more than 10 years of operational experience in large-scale data management!
- Rucio manages multi-location data in a heterogeneous distributed environment
  - Creation, location, transfer, and deletion of replicas of data
  - Orchestration according to both low-level and high-level driven data management policies (usage policies, access control, and data lifetime)
  - Interfaces with workflow management systems
  - Supports a rich set of advanced features, use cases, and requirements
  - Large-scale and repetitive operational tasks can be automated

# Who uses Rucio?

- Rucio was developed by ATLAS.
  - Largest user managing 500PB of data and 20GB/s transfers between 72 official sites.
- Rapidly growing [community of users](#).
  - It has recently been adopted by CMS.
- Currently each Experiment runs its own instance of Rucio.
  - Non-trivial overhead to get started.



# Multi-VO Rucio

- RAL developed the code to allow a single Rucio instance to support multiple VOs.
  - VOs would need to be kept completely isolated from each other.
- Objectives:
  - Reduce the initial overhead for VOs to try Rucio.
  - Allow RAL to provide a scalable solution to smaller communities.
  - Simplify configuration for both users and sites.
- STFC has received funding from multiple sources over the next 3 years (including EGI) to fund a multi-VO Rucio instance.
  - We are in the process of recruiting an additional fulltime developer.

# Rucio instance at RAL

- Multi-VO support was added to Rucio 1.23 which was released in July 2020.
  - RAL have been running this since August for SKA and dteam.
- We are currently fixing various bugs that have been identified by the current test users.
  - We are not actively seeking more users currently, but we can discuss on a case-by-case basis if you would like access sooner.

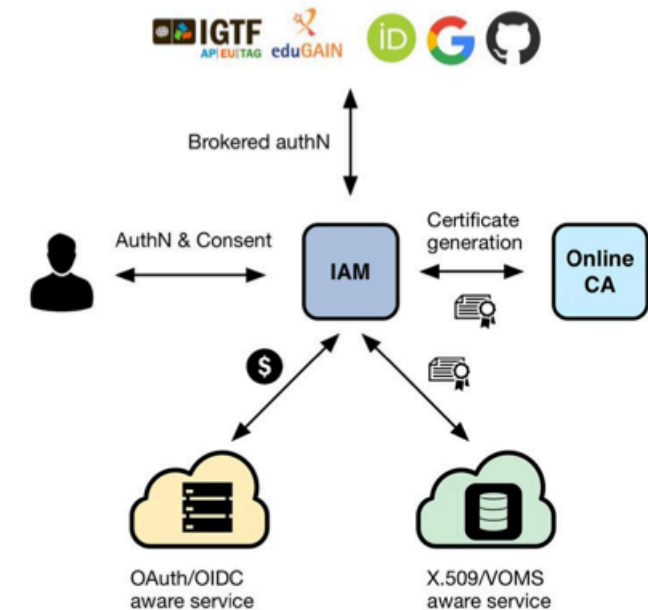


# Rucio at RAL Vision

- The service is running but it is still very much in development.
- We want to focus on making it as easy for the end user as possible.
  - We should be able to create new VOs within a working day.
  - VOs would only need to configure a policy package for their specific requirements.
  - Existing sites (e.g. those that support dteam) could be pre-configured. We would simply need a VO specific ACL/directory.
- We would continue to incorporate the many new features being developed by the Rucio team.

# OIDC & OAuth

- AAI integration with OIDC, OAuth2, JWT available since Rucio 1.22.0
- Currently running continuous tests on Rucio WLCG DOMA instance to test the full AAI capabilities of the involved infrastructure
  - Rucio → FTS → Storage(s)
- Different modes for transfers supported
  - Orchestrate all transfers under admin token
  - Orchestrate transfers under individual user tokens



# Quality of Service

- Integration of storage quality of service in Rucio
  - Replication rule functionality will allow users/workflows to express QoS needs
  - Rucio communicates QoS needs to storage
  - Rule evaluation engine will issue QoS transitions when possible/necessary
    - Otherwise conventional third-party-copy requests are being made
- Alignment with [DOMA QoS workgroup](#)
- Lots of interest from different communities
  - Rucio will deal with QoS in a generic way to allow communities to express their individual QoS policies
- Development ongoing
  - First, non-functional, changes already included in 1.23.0
  - 1.24.0 (November release) will include the first functional features



# Metadata

- Rucio supports different kinds of metadata
  - File internal metadata, e.g., *size, checksum, creation time, status*
  - Fixed physics metadata, e.g., *number of events, lumiblock, cross section, ...*
  - Generic metadata that can be set by the users
- A metadata plugin system was introduced with Rucio 1.23.0
  - All metadata interactions with Rucio are done via a single interface
  - The backend relays metadata request to the matching metadata plugin
    - Current available plugins: *table-column-metadata* and *json-column-metadata*
    - Communities can develop & integrate their own metadata plugins to Rucio
- Current developments expand the metadata query system to allow users more complex metadata queries (integrated in plugin system)
  - Comparisons, Inequality, compound queries, ...

# Summary

- A large amount of active development work is going on with Rucio.
- STFC will provide a Multi-VO Rucio instance for the EGI community to use which should be opened up to more users in (early) 2021.



Science and  
Technology  
Facilities Council

# Questions?