



# Multi-VO Rucio for EGI

Timothy Noble

# Rucio

- Intelligent data movement
- High Capacity
- Heterogeneous network and storage infrastructure
- Adaptive replication
- Data recovery
- Seamless FTS integration

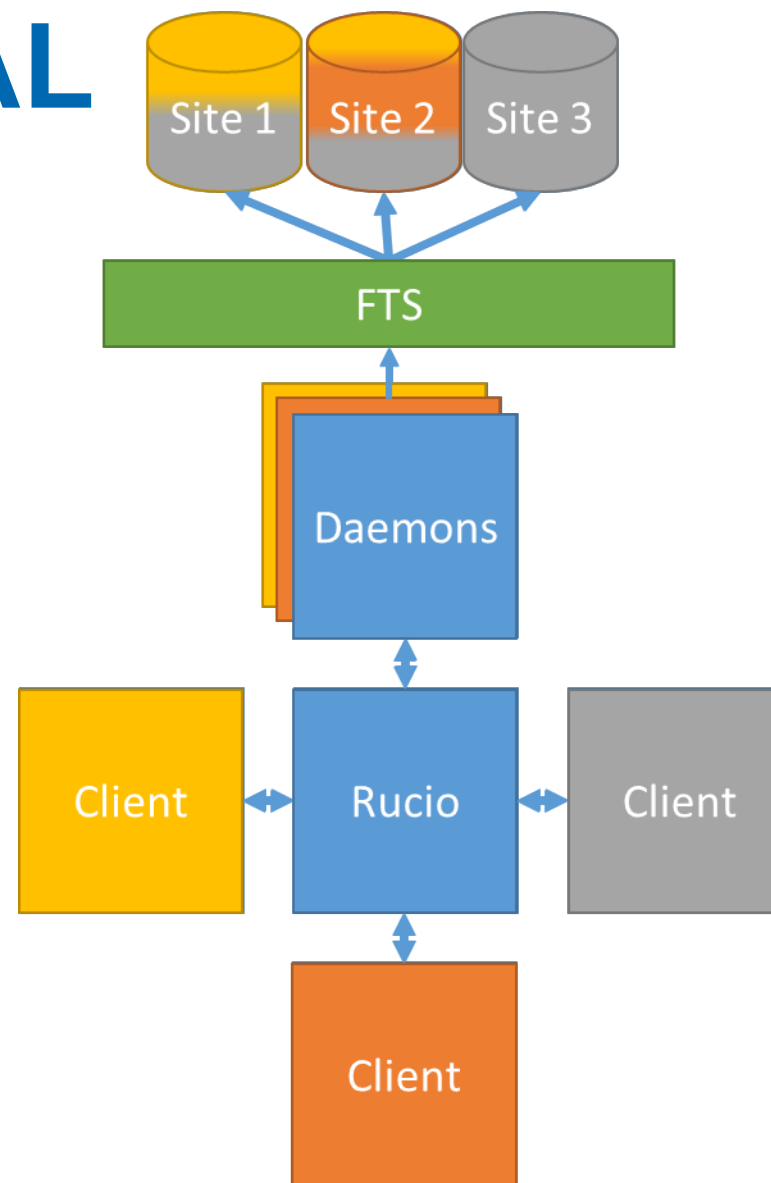


SCIENTIFIC DATA MANAGEMENT



# Multi-VO Rucio at RAL

- **6 VMs on SCD OpenStack cloud**
  - VM for access (Bastion)
  - Server
  - ELK
  - 3 daemon servers – one per VO
- **2 machines running the databases**
- **Benefits from existing RAL Tier-1 services**
  - Load balancer for the Rucio server provided by HA Proxy
  - Data transfers mediated by the RAL FTS
- **Providing regular functional tests (every hour)**
  - Submits data transfers between RSEs in the UK



# Advantages of Multi-VO Rucio

- **Running a Rucio instance that supports multiple VOs is beneficial for small experiments:**
  - Maintained by RAL not by smaller experiments
  - Low levels of load from smaller experiments.
  - One instance to support and maintain.
  - Shared RSE configuration
  - New VOs are quick to add – work with VO admin to setup their environment
- **More contact with Developers and larger communities using Rucio to know how best to utilise**

# Moving data with Rucio

- **EGI Data Transfer Service - FTS**

- **Replication Rules**

- A file from a data source may be uploaded to Rucio to an RSE
- This may not be where it is needed (wrong site)
- Add a replication rule

```
$ rucio add-rule scope:first_dataset scope:second_dataset 2 'country=uk'
```

- This would add the two datasets to the RSEs in the UK, and have 2 copies of this dataset within the UK

```
$ rucio add-rule scope:first_dataset scope:second_dataset 2 'country=uk\site=RAL-LCG2'
```

- This would add the same two copies across the UK but excludes RAL as one of the potential sites to keep this data

- **RSE expressions**

# Policy Packages

- **Allows you to customise access to Rucio**
- **Python Packages to better integrate Rucio with your experiment**
  - Generic files available for customisation – and we will work with you to develop these
  - Maintained by the experiment (with our support)
  - Permissions – Who can read, write, and delete, and where
  - Placement algorithms – How the data is located within the RSEs
    - Hashes
      - user.jdoe:test.file.1  
077c8119053bebb168d125034bff64ac  
/07/7c/user.jdoe/test.file.1
      - Deterministic – based on file names  
/test/user.jdoe/file.1
      - Non-deterministic – uses all parts of file  
/data/dataset1234/user.jdoe/test.file.1
- <http://rucio.cern.ch/documentation/policy-packages/>

# Future plans

- **EGI-ACE Developments**
  - WebUI for Multi-VO
  - Integration with IAM and EGI check-in
  - Invite new experiments to use Multi-VO Rucio
  - Improve documentation
- **Internal Developments**
  - Multi-VO selection of certificates for daemons
  - Containerisation
  - Improve monitoring
  - Gain experience with on boarding a variety of communities and requirements



Website: [www.egi.eu/projects/egi-ace](http://www.egi.eu/projects/egi-ace)

 [EGI Foundation](https://www.linkedin.com/company/egi-foundation)  [@EGI\\_elnfra](https://twitter.com/EGI_elnfra)

