4th Rucio Workshop summary

Martin Barisits
on behalf of the Rucio team

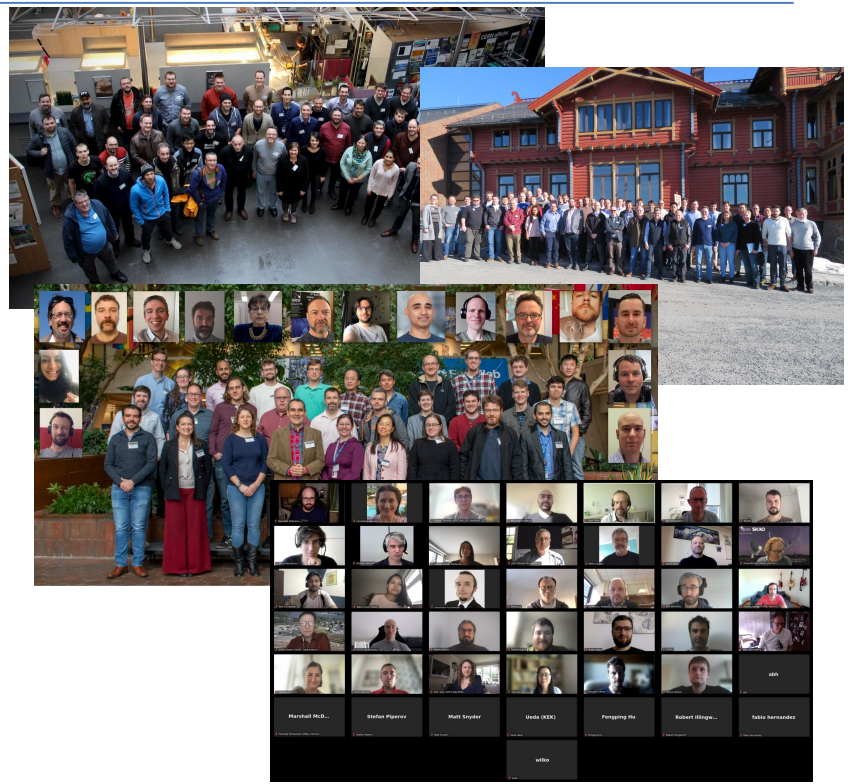
RUCIO
SCIENTIFIC DATA MANAGEMENT

4th Rucio Community Workshop

- **Sep 28 - Sep 30, 2021**
  - Virtual
  - Half-days with sessions allowing participation from all time zones
- https://indico.cern.ch/e/RUCIO2021
- Program committee
  - Martin Barisits, CERN (Chair)
  - Rosie Bolton, SKAO & ESCAPE
  - Alastair Dewhurst, STFC
  - Mario Lassnig, CERN
  - Cedric Serfon, BNL
  - Eric Vaandering, FNAL

4th Rucio Community Workshop

- 110 participants (Biggest Rucio workshop so far!)
- 4th edition of our community workshop!
  - 2018: CERN, Switzerland
  - 2019: University of Oslo, Norway
  - 2020: Fermilab, USA
  - 2021: Virtual
- Community is growing fast!
  - Now seriously established in non-HEP sciences
  - Community contribution model works very well!
  - 350+ users on Slack
  - No coding camp in 2020/2021 due to pandemic, hopefully we can arrange one in 2022



Contributions (to the main repository)

- 2017
  - **540** commits, **33k** LOC, from **12** contributors
- 2018
  - **775** commits, **64k** LOC, from **29** contributors
- 2019
  - **954** commits, **53k** LOC, from **31** contributors
- 2020
  - **559** commits, **54k** LOC, from **29** contributors
- 2021 (so far)
  - **344** commits, **39k** LOC, from **32** contributors
- Commits and LOC not a good indicator to judge complexity of contributions
- Top-contributors produce the majority of the code

Release plan

- 2019
  - **1.19** “Fantastic Donkey” February 2019
  - **1.20 LTS** “Wonder Donkey” June 2019 **EOL 07-2021**
  - **1.21** “Donkeys of the Galaxy” October 2019
- 2020
  - **1.22** “Green Donkey” March 2020
  - **1.23 LTS** “The Incredible Donkey” June 2020 **EOL 07-2022**
  - **1.24** “Aquadonkey” November 2020
- 2021
  - **1.25** “Rat-Donkey” March 2021
  - **1.26 LTS** “Donkey League of La Mancha” July 2021 **EOL at least 07-2023**
  - **1.27** “Batdonkey v Superdonkey” November 2021
- Long Term Support model works well - LTS releases actively used by the community

Agenda

- Keynote
  - “Pipe Dreams (and Nightmares)” Crystal Chua (AARNet)
- 27 community presentations
  - Format: 12’ + 3’
- 3 panels
  - Rucio in a non-grid environment
  - Astronomy & Metadata
  - Transfer & Storage
- 3 topical sessions
  - WFMS
  - Astronomy
  - Long tail of science
- Next slides will show some highlights with a focus on topical sessions & panels

WFMS

- Presentations of WFMS interfaced with Rucio
  - PanDA
  - DIRAC
  - CMS WMCore
- Plans of these projects
- How is Rucio interfaced
- What can be improved on Rucio-side to make WFMS life easier

PanDA

T. Maeno

PanDA and Rucio

- Bidirectional communication between PanDA and Rucio
  - PanDA → Rucio: HTTP requests using python API of rucio-client
  - Rucio → PanDA: Messages through ActiveMQ
- No data management component in PanDA
  - Fully rely on Rucio
  - PanDA system has its own dataset/file catalog
    - Used internally e.g. for data caching
- Most workloads read input files and upload output files only from/to local storages via LAN
  - To avoid chaotic data traffic over WAN

Brookhaven National Laboratory P. Nilsson 3

T. Maeno

PanDA and Rucio: Usages

- Dataset and file location discovery
  - To schedule workloads to computing resources that have input data locally
- File distribution
  - To send input files to computing resources that are idle but don't have the input files locally
  - Possible to distribute workloads more widely
- Preplacement of input files before running workloads on actual computing resources
- File aggregation
  - To aggregate output files produced by workloads
  - Asynchronous stage-out of output files to the final destination
- Additional hints for workload brokerage
  - Status of data transfer channels
  - Storage and free space sizes
  - Blocklist of storages
- User information lookup such as email address
- Lifetime-based data cleanup

Brookhaven National Laboratory P. Nilsson 4

PanDA - Paul Nilsson

DIRAC

[DMS] File Catalogs

- DIRAC namespaces
- Several catalogs can live in parallel
  - e.g. : DFC, LFC, **Rucio**, DIRAC TS, LHCb Bookkeeping, Belle2 AMGA (?)
  - there's always a "master" catalog
  - you can have conditional catalogs
- A new plugin for Rucio catalog was implemented by Belle2, ported to DIRAC
  - some details might be Belle2 specific
  - a mapping policy was needed

Multi-VO DIRAC+Rucio

- DIRAC server is multi-VO
- Multi-VO Rucio prepared by GridPP
  - see pres tomorrow morning
- DIRAC is almost ready to talk to multi-VO Rucio
  - will be part of a future 7.3 patch

10

Some DIRAC developments

- Well advanced (basically done): Python 3
  - py3 clients supported since version 7.2 (pip installable)
  - py3 server supported since version 7.3 (production)
    - rucio full support needs these 2 issues solved -- 1st is done
  - py2 support ends with 8.0 (in preparation)
    - with some obvious exceptions of part of pilots code
- Ongoing: `dips://` → `https://`
  - `dips`: DIRAC proprietary protocol for RPC calls
  - `http`: based on tornado
  - several DIRAC services already available using HTTP, and adding more
- Ongoing: token support, and IdP (IaM, Check-in)
  - will be in 8.0
- Started: running on kubernetes
- Planned: ES/kibana/grafana dashboards

we already store some data in ES

11

WMCore

High level view of Rucio usage

Rest APIs from Rucio through Rucio Client

We have our own python module, working as a wrapper on top of the RucioClient. Reasons for that are:

- Code maintenance - we develop/adapt only needed APIs
- Use of the caching modules provided by WMCore
- Error handling and logging provided by WMCore
- Disentangled from the mainstream RucioClient version (as long as Rucio keeps backwards compatibility)

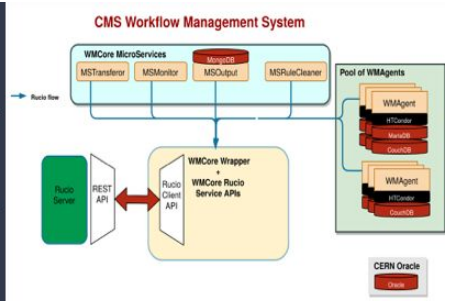
Data abstraction granularity

A general misalignment between Rucio and CMS due to inheritance from the previous data management system.

- **CMS dataset** corresponds to a **Rucio container**
- **CMS block** corresponds to a **Rucio dataset**
- **CMS file** corresponds to a **Rucio file**

Difficult to alleviate. We need to constantly keep track of inherited terminology, namings and references in the code and different parts of the system as a whole.

5



Possible Improvements

For WMAgent:

- Provide a single API that performs an atomic operation for adding a DID and attaching it to another DID
- A better way of handling DuplicateRule Exception when inserting items in bulk. We would like to avoid manual and expensive resolution on the client side - Either a clear message, pointing to which are the duplicates or an additional API giving the returning only the duplicates in a single call: https://github.com/rucio/rucio/issues/3807

For Microservices:

- Support container input type for get_dataset_locks API: https://github.com/rucio/rucio/issues/4807
- Extend get_dataset_locks API to support requests in bulk - thus, client providing a list of datasets to be discovered - we need to find out all the RSEs holding parts of it: https://github.com/rucio/rucio/issues/3982
- Evaluate dataset replication state in the get_dataset_locks API - given a rule on a container, we would want to check which datasets are locked, where, and whether all the files have already been transferred: https://github.com/rucio/rucio/issues/4034
- Properly and on a timely manner update timestamps for Stuck rules: https://github.com/rucio/rucio/issues/4592
- Static locks per RSE for multi-RSE rules - We want to be sure Rucio does not move/transfer the data from an RSE while our agent/htcondor is delaying processing at a single site out of the many covered by that same rule.

9

WMCore - Todor Ivanov

Astronomy

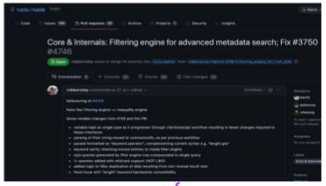
- Many Astronomy communities evaluating Rucio
  - Largely driven by the ESCAPE project
- Challenges
  - Even small Astro communities have considerable data rates/volumes
  - Community not used to (our) large-scale data management ecosystem
    - Authentication, Storage, Third-Party-Copy
    - Introducing our HEP software stack to others is not trivial
      - Projects like ESCAPE help here
- Some data workflows are different to ours
  - Data embargos!
  - Data discovery

SKA

SKA Observatory Data Flow

Metadata improvements

- Dev community very friendly and accommodating.
- Reasonably high entry barrier to core development, not enough thorough documentation about the underlying architecture. Nomenclature is sometimes a little tricky to navigate (think daemon names..)
- Been working on expanding existing metadata search functionality to allow users to build more complex queries that better match prospective SKAO work flows. Work in progress, initial PR: https://github.com/rucio/rucio/pull/4746.
- More work needs to be done to port this functionality to custom metadata searches (specifically not using the specific HEP hardcoded table columns).



Wish list

- Indigo IAM integration/token based authz...we really really need this
- Fine grained authorization is very important to the astronomy community
  - 'View' privileges are not a given!
  - Would be good to see a path forward for this
- Developing a better storage understanding to be able to liaise with astronomy site admins
  - How do we tell someone completely external to the WLCG-world what they need to be onboarded to a Rucio instance?
  - Ceph storage: we would like to explore what it entails to onboard Ceph storage as an RSE

Housekeeping tasks

- Rucio upgrade to next LTS release: 1.26.*
- Flux-ify the k8s setup (Good to hear CMS also has experience here!)

TL;DR

SKA - Rohini Joshi

CTA

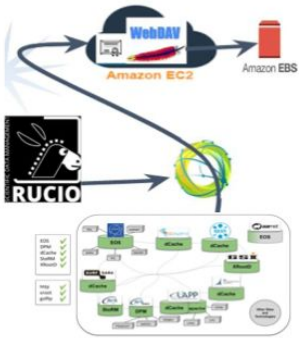
Use Commercial Cloud (CC) to deploy a rse

Infrastructure:
Elastic Compute Cloud (EC2) instance : t3.medium

Instance	vCPU*	CPU Credits/hour	Mem (GiB)	Storage	Network Performance (Gbps)***
t3.medium	2	24	4	EBS-Only	Up to 5

Prérequisit:
1.Set up Domain Name «escapedatalake.com»
2.Activate the domain name service to map domain name with the IP of the EC2 instance
3.Generated a grid certificate for the given Domain name
4.Install the software acting as a rse (apache with WebDav plugin) on the ec2 instance (packaged as a docker image)
5.Create and map a storage capacity using the Elastic Block Store (EBS) service.
6.Configure the rse to use this storage capacity

Results:
- Cloud's rse is managed like any other rse by RUCIO
- Cost
  - Ingress cost is very low and negligible compare to EC2 cost
  - EBS cost is expensive, 80 USD/TB-Month
  - S3 is less expensive (should be tested soon)



1.2 TB of data replicated to the cloud's rse in 4 days

RUCIO/DIRAC interface

Reprocessing CTA's Use case :

- Raw (DL0) data is identified on tape (obsid or time range) via **metadata** e.g. using getMetaData method
- Data is staged from tape (COLD) storage to temporary disk (HOT).
- Data is reprocessed using CTA pipeline software via the workload management system (DIRAC)
- Ingest the resulting final data products (DL2) data into the DataLake.

Infrastructure:

DIRAC :
- CTA instance used for the generation of monte carlo Data.
- RUCIO catalog developped for Belle2 experiment
- Supports CTA VO only

RUCIO : PIC installation (CTA DataLake)
RSE : dCache in PIC

LOFAR

Dreams and hopes (AAI)

Rucio seems to have lots of potential for our use cases

- AAI granularity
  - e.g.: public user, group of owner, data owner (PI), observatory staff, admin
  - Also granularity in who can grant access to data:
  - The PI can grant (either directly or by proxy) access to data, observatory stadd/admin to PI.
- AAI based on metadata
  - ie: if file is part of a certain project, AAI of that project is followed.

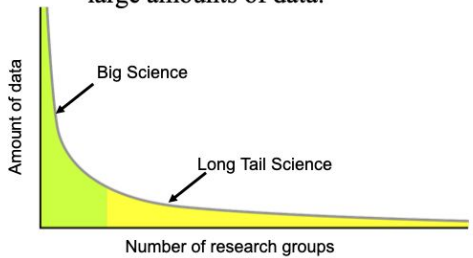
Dreams and hopes (Metadata)

- Custom metadata that is the same in the user-facing queryable database, the Virtual Observatory and the Datalake. Direct connection between Rucio and observation database for shared metadata.
- Advanced metadata rules, including simple math (a**2+b**2 < X as simple cone example). Or setting where dat will go based on projectID.

Long tail of Science

What is the long tail of science?

The long tail of science refers to the large number of individual researchers and small laboratories who do not have access to dedicated computational resources and online services to manage and analyse large amounts of data.



Alastair Dewhurst, 30st September 2021

Alastair Dewhurst

Advantages to use 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

Timothy Noble

LTOS - Views

Rucio @ EGI

- Multi-VO Rucio operated by UKRI-STFC in the context of the **EGI-ACE** project (started in 2021)
- Service provisioning started in June, with some communities interested that have been contacted for demo and got access (e.g. Lattice QCD, ROE)
  - Webinar 6th of October: https://indico.egi.eu/event/5711
- More Communities in EGI-ACE are planning to use the instance
  - LOFAR, EISCAT-3D
- Status
  - Access via X509 using the rucio client docker distribution
  - Connected to grid storages at EGI sites and using the FTS @ UKRI-STFC

Andrea Manzi

Benefits of Multi-VO Rucio for Fermilab

- Fermilab, as a national laboratory may have many smaller entities that wish to utilize Rucio for their data management needs in the future
- Smaller VOs have modest requirements for
  - Metadata support
  - Workflow integration
- A need to execute bulk data curatorial services such as data movement and archival
  - Ad-hoc data movement: Often experiments only need to transfer data between facilities (HPC)
- Utilization of a multi-VO Rucio deployment will reduce support overhead and allow greater agility

Brandon White

Panel - Rucio in a non-grid world

- Two major possibilities
  - **1. Bring the grid-tools**
    - Complexity of the architecture
    - Single-click deployment/configuration/integration
    - Packaging for a multitude of systems
  - **2. Make our software compatible**
    - Large development efforts
    - Potentially a wild goose chase
- How to quickly make available some storage as an RSE?
- Support for personal user cloud storage
- Support for Windows (e.g., integrated IIS, Windows server shares, …)
- **Absolutely needed:** Data management / Rucio for newbies document

Panelists:
Alastair Dewhurst (STFC)
Andrea Manzi (EGI)
David Cameron (UOslo)
Ilija Vukotic (UChicago)
Mario Lassnig (CERN)
Oliver Keeble (CERN)

Panel - Astronomy & Metadata

- Do we mean the same thing by Metadata in Astronomy and in HEP
  - In Astronomy, the use of public data products, with or without custom (planned) data products is quite different to a typical HEP use case
  - We should pay careful attention to which astro metadata types are needed for d**ata management** / **data discovery** / **data analysis**, as three different regimes
- Interplay between filesystem / hierarchical structures and rucio transfers
- Proprietary data / data management and permissions lifecycles
- On-boarding non-HEP storage sites has some barriers - lots of (HEP) ecosystem knowledge necessary

Panelists:
Cedric Serfon (BNL)
Dave Morris (UEdinburgh)
Greg Daues (NCSA)
Pandey Vishambhar (ASTRON)
Rob Barnsley (SKAO)

Panel - Storage & Transfer

- Tokens
  - Lots of progress was done in the last years, lots of work is still ahead of us
    - Most solutions are out there, we need to find the right way to bring them to our infrastructure
  - Clear directives is needed from AuthZ experts
    - Capabilities vs. identity
  - Need to move from fat-tokens to fine-grained auth workflows
  - What does it mean for Rucio to be the sole token issuer?
  - Security audit of our full-stack workflows will be needed
- Transfer protocols
  - WLCG moving to HTTP-TPC and root, what are others doing?
  - Should Rucio have more knowledge of protocol capabilities?

Panelists:
Andrea Ceccanti (INFN)
Andy Hanushevsky (SLAC)
Hannah Short (CERN)
Martin Barisits (CERN)
Mihai Patrascoiu (CERN)
Paul Millar (DESY)

Summary

- Many more interesting community presentations
  - Please have a look on the recordings
- Very productive and interactive workshop
  - Hopefully the next one can be face-2-face
- We need to lower barrier of entry for **new communities**
  - Not only Rucio, but entire HEP software stack
  - Common technologies: OAuth tokens, Containers, Kubernetes, common protocols help!
- **Established communities** further optimize their Rucio usage and make it available to the community
  - Data Lake as a Service, Globus (Online), Fuse plugin, Lightweight Rucio clients, integration with ScienceMesh, and many more!
- For development plans please have a look on State of the Union talk

Community


eneas
Advanced European Network of E-infrastructures
for Astronomy with the SKA


LCLS


Belle II

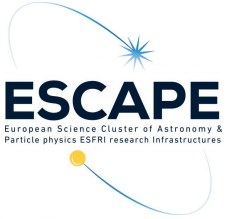

Science & Technology
Facilities Council


COMPASS


CERN

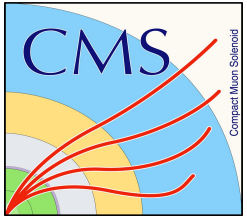

ARC


ATLAS
EXPERIMENT


ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures


Alpha Magnetic Spectrometer
AMS-02
Europe • Asia
North America


Xe
XENON
Dark Matter Project


CMS
Compact Muon Solenoid

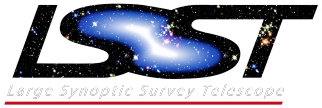

DUNE
DEEP UNDERGROUND
NEUTRINO EXPERIMENT


egi


3D
EISCAT


VIRGO


cta
cherenkov telescope array


LSST
Large Synoptic Survey Telescope


IceCube
SOUTH POLE NEUTRINO OBSERVATORY


FTS


eXtreme DataCloud


Fermilab


BROOKHAVEN
NATIONAL LABORATORY


PAWSEY
supercomputing centre


SKA
SQUARE KILOMETRE ARRAY


LIGO


Open Science Grid


EUXDAT

More information

Website	http://rucio.cern.ch
Documentation	https://rucio.cern.ch/documentation
Repository	https://github.com/rucio/
Images	https://hub.docker.com/r/rucio/
Online support	https://rucio.slack.com/messages/#support/
Developer contact	rucio-dev@cern.ch
Publications	https://rucio.cern.ch/publications.html
Twitter	https://twitter.com/RucioData