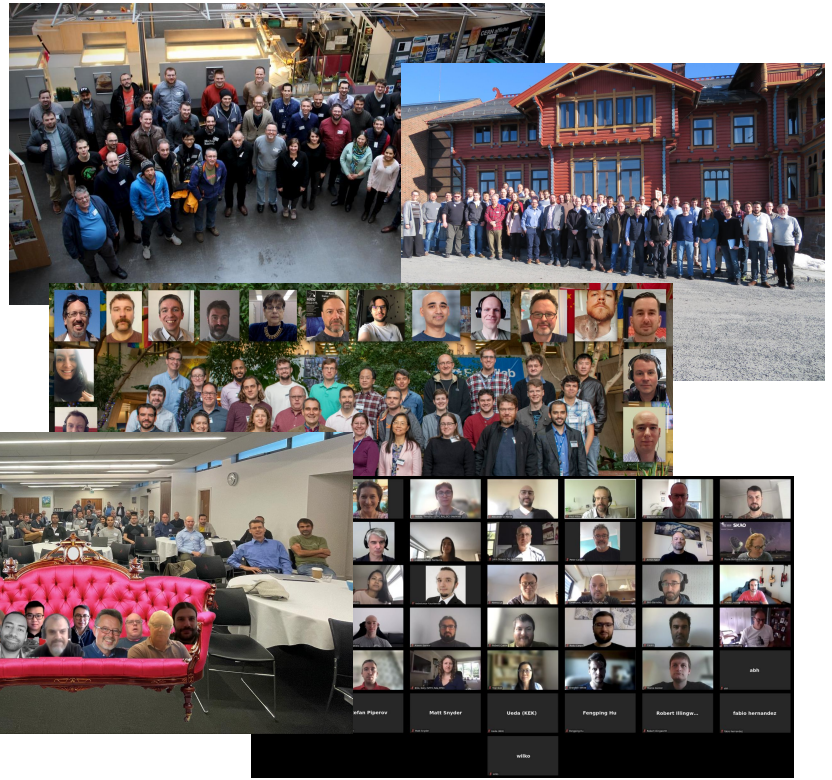
5th Rucio community workshop summary

Martin Barisits (CERN)

CERN

5th Rucio community workshop

- Nov 10 & Nov 11
  - Co-located with the WLCG workshop (Nov 7 - Nov 9)
  - Lancaster University, UK
- https://indico.cern.ch/event/1185600/
- Our 5th community workshop
  - 2018 CERN
  - 2019 University of Oslo, NO
  - 2020 Fermilab, US
  - 2021 Virtual
  - 2022 Lancaster University, UK

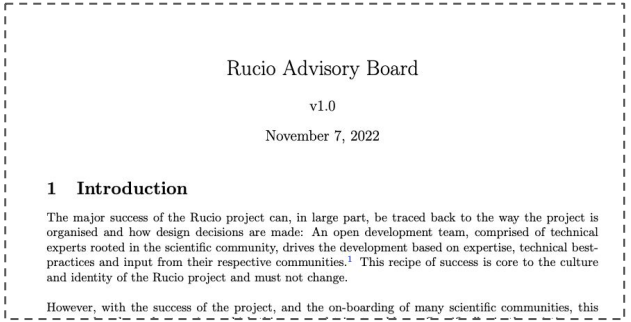


5th Rucio community workshop

- Registrations
  - 46 in-person
  - 40 remote (20-25 in average active)
- Program committee
  - Martin Barisits, CERN (Chair)
  - Rob Barnsley, SKAO
  - Gareth Hughes, CTAO
  - Michael Kirby, FNAL
  - Mario Lassnig, CERN
  - Cédric Serfon, BNL
  - Eric Vaandering, FNAL
- Videos of all talks available online!
- Thanks a lot to Roger Jones and Lancaster University team!!!

Community

- Growing fast
  - 475 users on our Slack (350 at this time in 2021)
  - Various organisations seriously committing to Rucio as the DDM solution for their communities
    - Fermilab running 6 Rucio deployments
- No coding camps in 2021 & 2022
  - Due to COVID
- We will organise a DocAThon and possibly Coding Camp in 2023
- Will introduce a Rucio Advisory Board in 2023
  - Long-term priorities and plans of communities
    - Advise on alignment of Rucio project objectives and plans
    - Identification of common objectives to form common developments
  - Advise on resource and person-power situation within Rucio
  - Advice on collaboration on funded projects
  - Non-binding strategic advice



Rucio Advisory Board

v1.0

November 7, 2022

1 Introduction

The major success of the Rucio project can, in large part, be traced back to the way the project is organised and how design decisions are made: An open development team, comprised of technical experts rooted in the scientific community, drives the development based on expertise, technical best-practices and input from their respective communities.[1] This recipe of success is core to the culture and identity of the Rucio project and must not change.

However, with the success of the project, and the on-boarding of many scientific communities, this

Release plan

- 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
- 2022
  - 1.28 “Teenage Mutant Ninja Donkeys” March 2022
  - **1.29 LTS** **“Into the Donkeyverse”** **July 2022** **EOL at least 07-2024**
  - 1.30 “The Donkeynator” November 2022
- 2023
  - 1.31 “Donkeys of the Caribbean” March 2023
  - **1.32 LTS** **“The Good, The Bad and the Donkey”** **July 2023** **EOL at least 07-2025**
  - 1.33 “Eternal Sunshine of the Donkey's Mind” November 2023

Stats for PRs and Commits in core repository

- 2018
  - 540 commits, 33k LOC, from 12 contributors
- 2019
  - 775 commits, 64k LOC, from 29 contributors
- 2020
  - 954 commits, 53k LOC, from 31 contributors
- 2021
  - 559 commits, 54k LOC, from 29 contributors
- 2022 (so far)
  - 583 (+303) commits, 53k (+230k) LOC, from 30 contributors
- Commits and LOC not a good indicator to judge complexity of contributions
- More and more work is done outside of the core repository (WebUI, documentation, helm-charts, containers, …)

Workshop agenda

- Very tight program
  - Only 1+½ days of time
- Two community report sessions
  - 10 presentation a 20’ (15’+5’)
- Four technical sessions
  - Technology stack
  - Metadata
  - Tokens
  - Transfers
  - 18 presentations a 15’
    - Discussion blocks for each session
- Closing discussion
- Dinner @ The Borough
- No keynotes or panels this year

Rucio and the ESCAPE project experience (Xavi Espinal)

Outlook

Rucio is at the forefront on key activities in future distributed computing activities

Consolidate efforts on Rucio R&D instances?

- Joint **operational and maintenance** efforts in the several R&D activities? Common Rucio infrastructure for DOMA, WLCG and EC-projects? (virtualization platform, CD/CI, Terraform, etc.), joint instances?
- To benefit from development on several areas: metadata, tokens, more robust (and wide) testing on new releases
  - ESCAPE ESFRI RIs also participating actively and perhaps interested in common deploy and operational models

Future projects and joint funding opportunities, e.g. the ESCAPE Open Collaboration Agreement

- Signed by the ESCAPE ESFRs is providing a formal statement to foster scientific collaboration in computing, maintaining the community together and following-up on ESCAPE work program
- Involve computing experts from ESCAPE RIs and other scientific communities together with LHC community
- Identify **common needs and initiatives**, common interest will naturally keep activities active and in shape to apply for funding windows

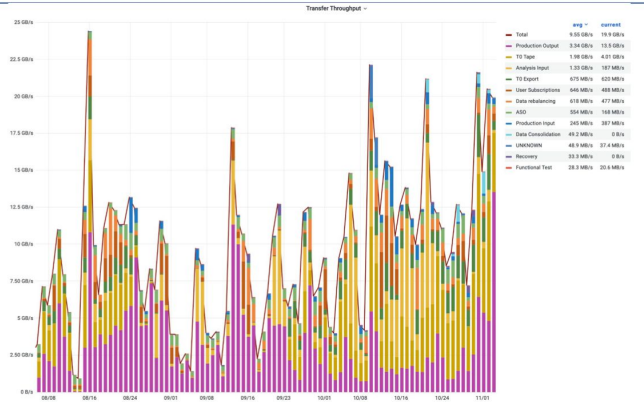
The Rucio Experience: ATLAS & CMS (Mario & Eric)

CMS Data Rates

400k transfers/day

1 PB/day

10 GB/s throughput

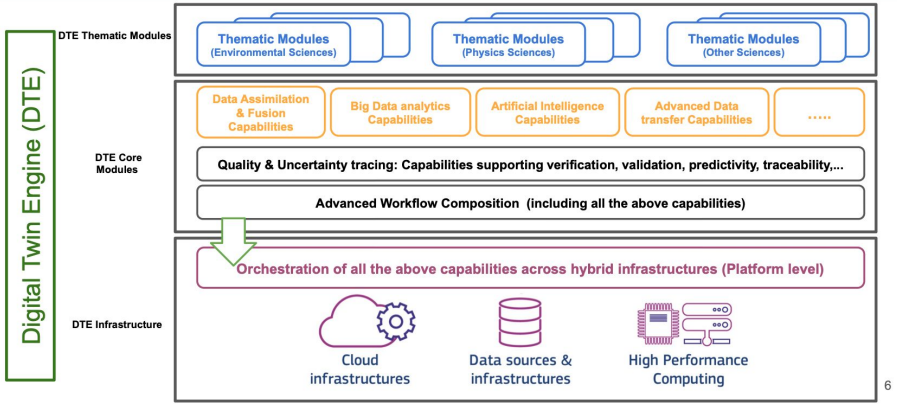


ATLAS Data Rates & Deletion

Production (Monte-Carlo Simulation) is dominating ATLAS transfer volume by a factor 3
Deletion volume always slightly surpasses transfer volume
Recent site operations (storage decommissioning, storage migration) are putting significant load on *Data Consolidation* activity
Data Carousel (dynamic exchange of data between disk and tape) is picking up larger fractions of volume at the Tier-1s

Rucio and the interTwin project (Paul Millar)

Digital Twin Engine - Strawman Concept



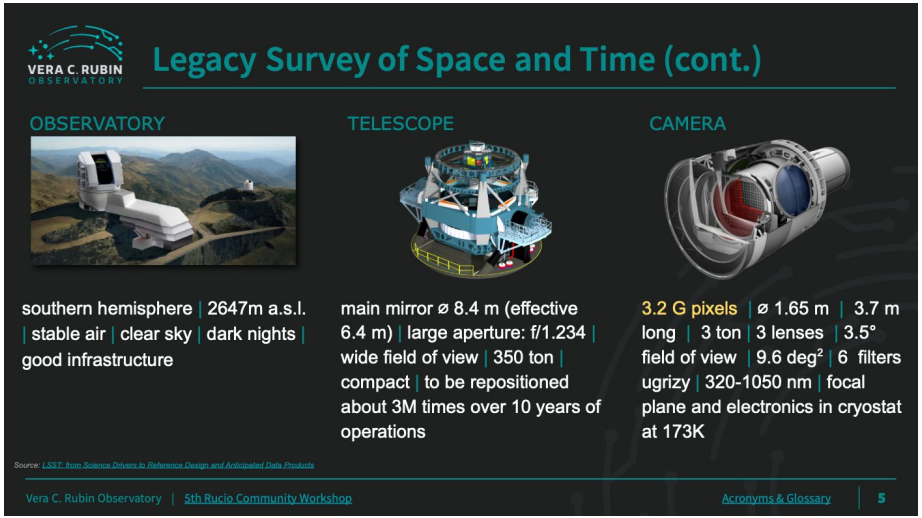
6

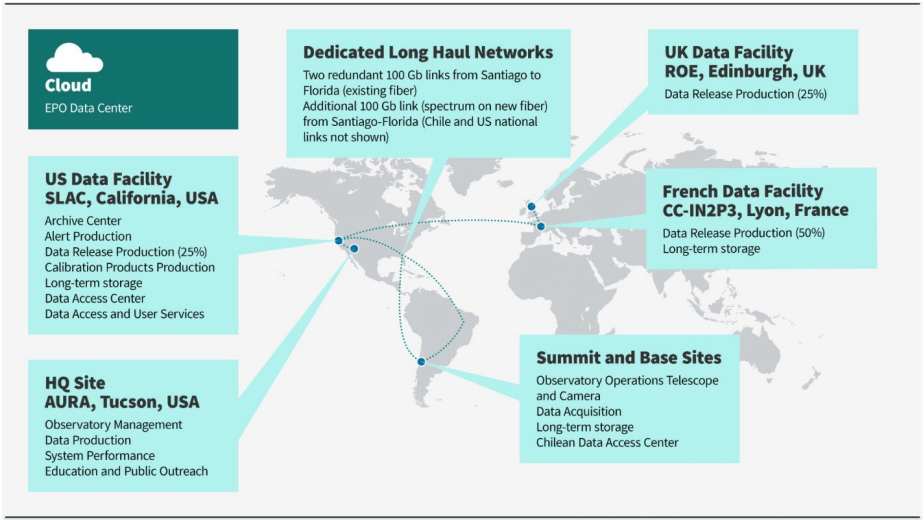
Summary

- interTwin is designing and "road testing" a blueprint for a **next-gen AI/ML-driven federated computing infrastructure**.
  This infrastructure should be widely deploy at **facilities**, and being available through **EOSC**.
- ESCAPE DataLake (and, therefore, Rucio) is a **core component** of federated data management within interTwin.
- The work will involve some familiar faces, but also some **new communities**.
- New use-cases will likely lead to **new requirements** on Rucio.
- We anticipate working with the Rucio team to have these enhancements **included with Rucio**, for everyone's benefit.

29

On the intended use of Rucio for LSST (Wei Yang)





Fermilab Rucio Operations (Brandon White)

DUNE Rucio Deployments

- DUNE has used a Rucio deployment on OKD3 (frozen at 1.26) for production operations since late 2019
  - Uses a custom image build/deployment framework to allow for customization such as policy packages
  - Versus K8s: are Routes vs. Ingresses, custom ServiceAccount permissions, monitoring config
  - Deterministic RSEs. No custom data placement
- Running a development deployment of 1.29
  - Deployed on Fermilab's development OKD4 cluster
  - Uses the DUNE policy package for custom data placement at tape sites
  - Production OKD4 cluster is newly online!

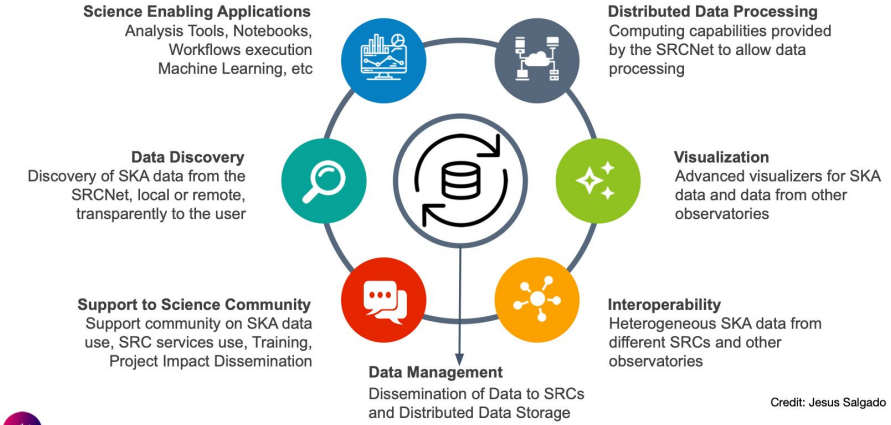
ICARUS Rucio Deployments

- Experiment at Fermilab searching for sterile neutrinos
- Fermilab supports a 1.26 deployment for ICARUS
  - Will be upgraded to 1.29 once the FNAL OKD4 cluster comes online, scheduled for after the DUNE upgrade
  - Limited issues with this deployment, has run fairly stably on OKD3 since 2020
  - Stability is usually limited by usual service management issues (e.g. certificate renewal)
- The ICARUS use case is fairly simple
  - Deterministic RSEs, no custom data placement
  - Primarily used for data movement from FNAL->INFN
  - For the moment replica cataloging functionality for ICARUS is provided by the legacy Sequential Access via Metadata service

SKA Regional Centre Data Management Update (Rohini Yoshi)

SKAO data product delivery

SKA Regional Centre Capabilities



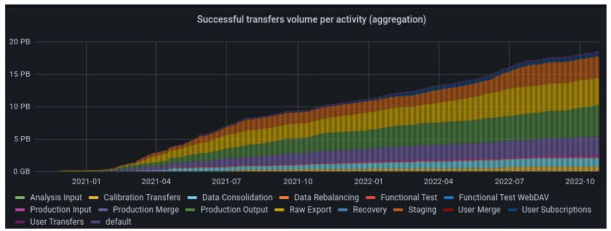
Credit: Jesus Salgado

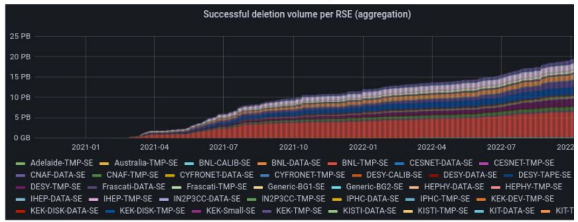
Slide /

Belle II (Cedric Serfon)

Introduction

- Belle II now uses Rucio in production for 1.5 years
- No major issues. It fits for Belle II purpose

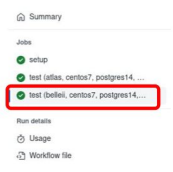


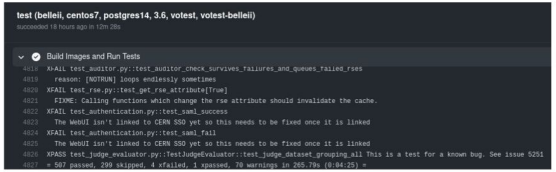


Brookhaven National Laboratory

Belle II Rucio unit-tests

- In the past, we suffered from some regression due to a change not taking into account Belle II convention
- In addition some methods specific for Belle II were introduce (e.g. dirac add_file)
- For these reasons, votest test suite was introduced
  - It includes specific Belle II tests
  - In addition, most of the other tests that only worked with ATLAS convention also work with B2 convention
  - We also spotted some methods that can only work for ATLAS convention (see later)



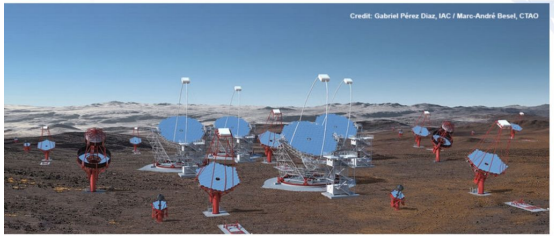


Brookhaven National Laboratory

5

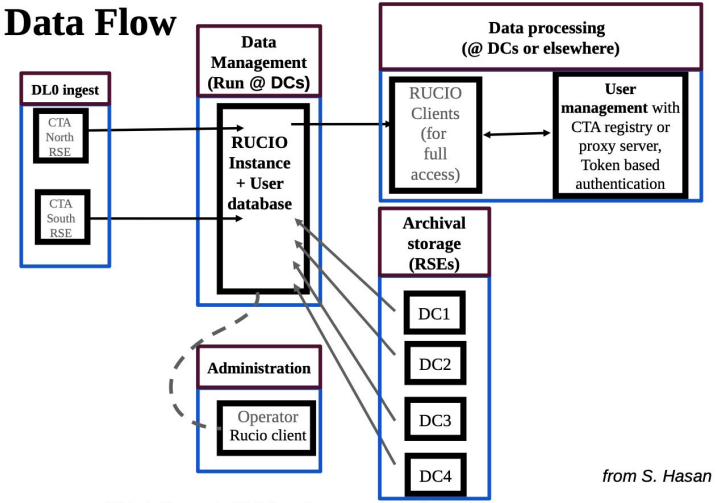
Rucio framework in … the CTA Archive (Georgios Zacharis)

The Cherenkov Telescope Array Project



- Two sites (*CTA North* and *CTA South*) taking data with four data centers in Europe
- O(10TB) of raw reduced data required to be transferred 'daily'
- O(50PB)/year from both sites to be archived
- Data must to be duplicated at the European DCs and removed at origin
- Processed up to science data
- Stored long-term on tape and periodically reprocessed

CTA Bulk Data Flow



from S. Hasan

Rucio & dCache in the SDC@KIS (Peter Caligari)

Background

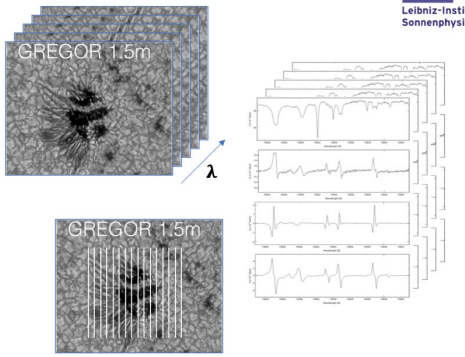
Solar Observational Data

Versatility: bone and bane of ground-based solar physics.

- Highly inhomogeneous data
- Existing data is not standardized
- Tailored post-processing needed

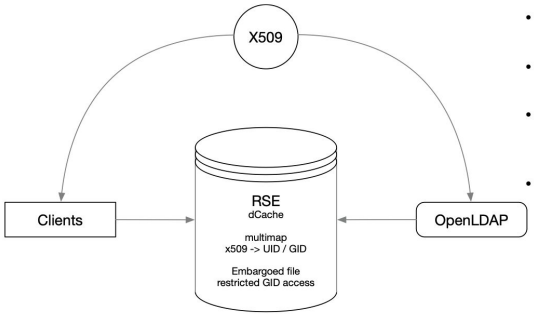
⇒ 2D/3D maps of physical parameters ($\vec{B}$, $\vec{v}$, P, T, …) as a function of time

- 3 – 70 TB/d



Schlichenmaier et al. (2016)

Embargoes



- X509 based (certificates only valid for 1 year!); no VOs
- Several asynchronous daemons per RSE involved
- Is this concept still feasible after switching to token-based AAI (IAM)?
  - Should we wait?
- File-based, not scope-based! Would require:

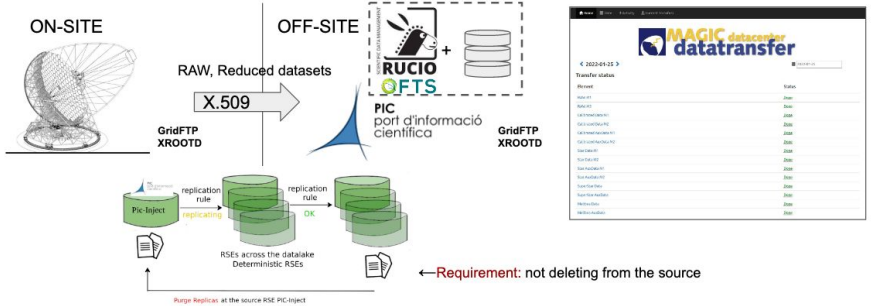
```
rucio list-dids*:instrument*
```

Rucio for Gamma-ray astronomy (Jordi Delgado)

CTA-ESCAPE WP2 USE CASE (I)

Long haul ingestion and replication - **MAGIC Configuration**

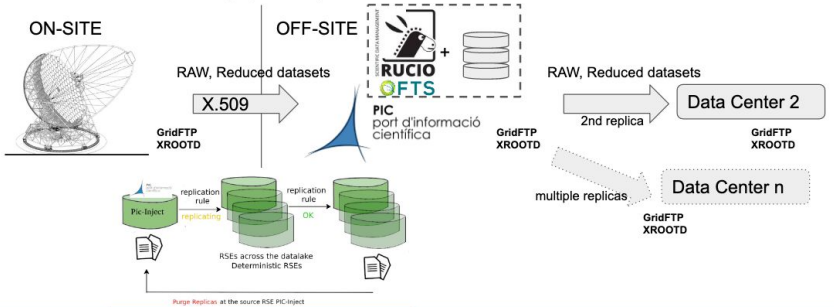
Description/Goal: Automatic detection and transfers of data from a remote site (RSE at 'on-site') produced from observation at ORM-la Palma, transfer and replication in off-site RSEs (PIC).



CTA-ESCAPE WP2 USE CASE (I)

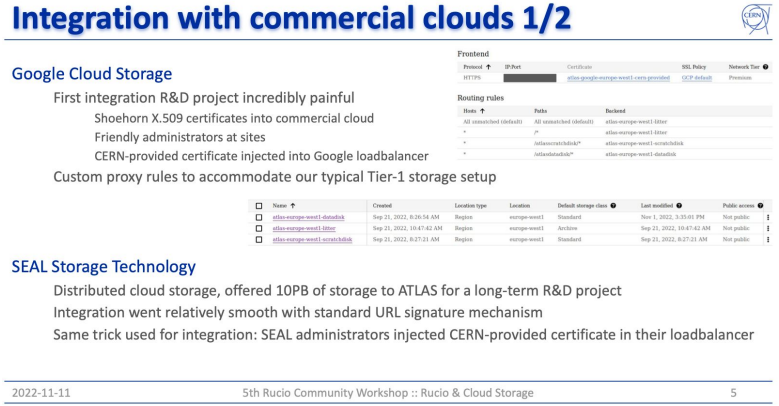
Long haul ingestion and replication - **CTA Configuration**

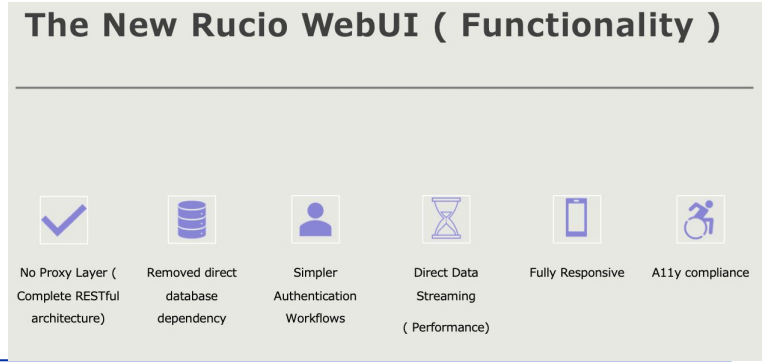
Description/Goal: Automatic detection and transfers of data from a remote site (RSE at 'on-site') produced from observation at ORM-la Palma, transfer and replication in off-site RSEs (PIC) and after replication is successfully done, RUCIO triggers the deletion of the files from the origin (La Palma).



Tech session: Tech stack 1/2

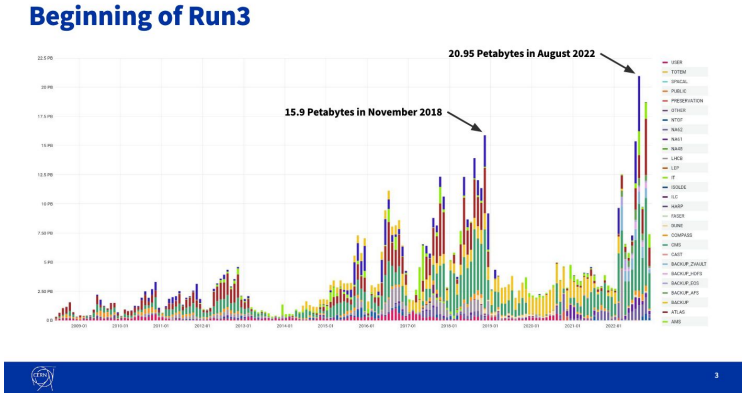
- Commercial clouds (Mario Lassnig)
  - Lots of interest
  - Google, Amazon, SEAL
  - Giant steps made the last years
- Database Overview (Cedric Serfon)
  - Only ATLAS and CMS on Oracle, others on Postgres
  - ATLAS by far largest database
    - Belle II and CMS second largest
- New Rucio WebUI (Mayank Sharma)
  - New Rucio UI coming soon
  - Fully responsive, digital accessible, new tech

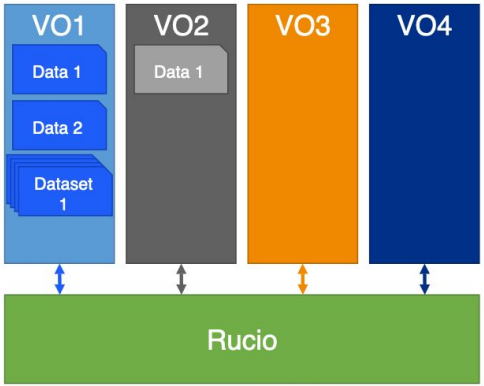




Tech session: Tech stack 2/2

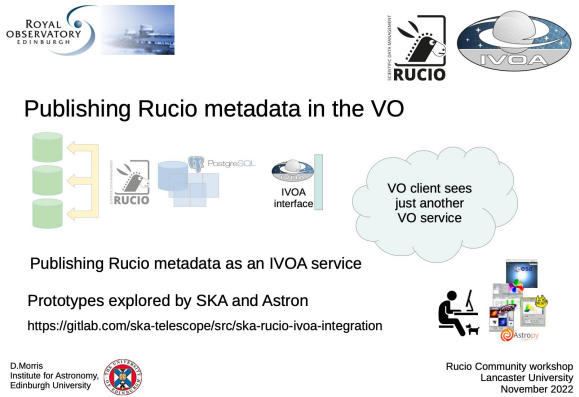
- CERN Tape Archive
  - Impressive performance (21 PB in Aug-22!)
  - Plans: Archive metadata, tape collocation
- Multi-VO Rucio
  - Many improvements for Multi-VO mode implemented
    - Multi-certificate handling for transfers
    - Deployment simplification
  - Monitoring

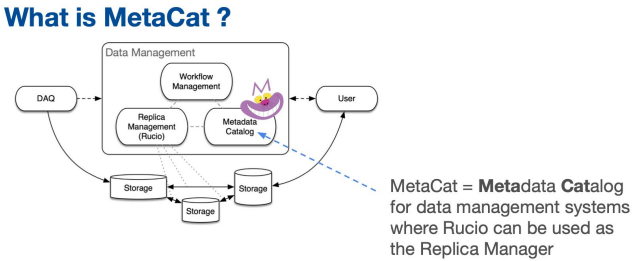




Tech session: SIG Metadata

- Lots of work done in the last year
  - New plugins: MongoDB, Postgres
  - Inequality operators
- IVOA and beyond (Dave Morris)
  - IVOA-Rucio prototype explored by SKA and Astron
  - But can we go further?
    - ⟶ Virtual Laboratories
- MetaCat - Metadata catalog (Igor Mandrichenko)
  - External metadata catalog for Rucio - Used by DUNE
- Rucio Metadata tests for Belle II (Cedric Serfon)
  - Import of Metadata + metadata queries
  - Rucio delivers the required performance for Belle II



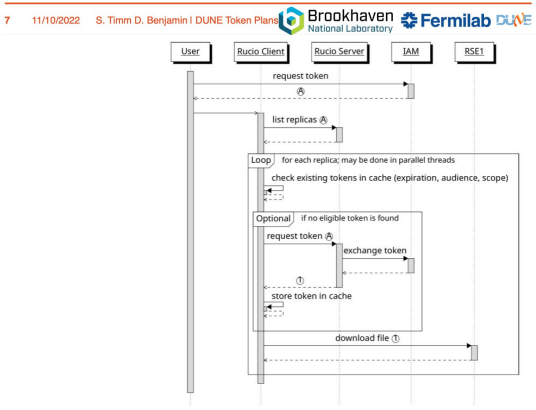


Tech session: Tokens

- Evolution from X509 certificates to OIDC/oAuth2 tokens
  - Lots of interest from the community
  - Not only WLCG, but hugely important for astronomers
    - → Data embargos
- WLCG AAI summary (Maarten Litmaath)
  - Key points from WLCG
- Token requirements from DUNE (Steve Timm)
- Rucio token plans (Martin Barisits, Dimitrios Christidis)
  - Scope limitations
    - Making it configurable for each community
  - IAM vs CILogon, WLCG token profile
  - Download / Upload workflow changes

Permission Layout: Legacy vs. Tokens

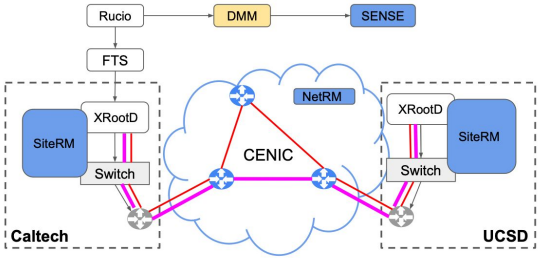
- Legacy situation (X.509 proxy / NFS v4.1):
  - Fermilab dCache has each individual user mapped to individual user id.
  - Unix permissions govern which users can write in which directories.
  - "dunepro" production user owns all the raw data and production outputs, (managed by Rucio).
  - Other sites—DUNE VO maps to one or two users. "dune" and "dunepro"
  - All data at external sites managed by Rucio (except small interactive space @ CERN).
- Tokens on dCache:
  - By default all tokens issued by the dune issuer map to dune:dune
  - Can override this with dCache inheritance—then files will inherit user:group of the directory
  - This allows us to have group quotas
- Corner cases exist:
  - Files can and do wind up with different owners
  - X.509 proxies can write directories tokens can't, and vice versa.
- GOAL: want to keep everything manageable by the production user in the era of tokens.



Tech session: Transfers

- Network optimisation is key
  - SENSE (Diego Davila Foyo)
  - ALTO (Richard Yang)
  - Packet Marking (Shawn Mc Kee)
- Transfers: the new bearings of the Conveyor (Radu)
  - Evolution of the Rucio transfer machinery
  - Globus Online anyone?
- FTS 2023: Plans and direction (Mihai Patrascoiu)
  - FTS is the primary transfer service of Rucio
  - Rucio is the primary data management system for FTS
    - 80% of all FTS transfers come from Rucio
  - Plans
    - HTTP error reporting improvements
    - Token support

How it works? For a priority Rucio request



For a priority Rucio request (pink) DMM picks a pair of free IPv6s and requests a bandwidth allocation on them to SENSE

DMM return the selected pair of IPv6s to Rucio

SENSE instructs SiteRM to implement specific routing and QoS on the given IPv6s at the site level

SENSE instructs NetworkRM to implement specific routing and apply QoS in CENIC nodes in between the 2 IPv6 endpoints

When the transfer is finished Rucio signals DMM which request the deallocation of the priority services 8

Recent transfer machinery evolution

- **Rucio 1.26**
  - Complete re-write of the Submitter. Many major bug-fixes and some functional changes.
- **Rucio 1.27**
  - Transfertool-agnostic code in the core. Improvements in the Receiver.
- **Rucio 1.28**
  - Introduction of a new database table for multi-hop transfers. Fixing long-standing issues with multi-hop transfers and enabling multi-hop between transfertools.
- **Rucio 1.29**
  - Preparer becomes more intelligent. If used, performs path computation and source selection. Improves multi-transfertool and throttler workflows.
- **Rucio 1.30 (upcoming release)**
  - Throttler should become more usable.

11/11/2022 Radu Carpa 7

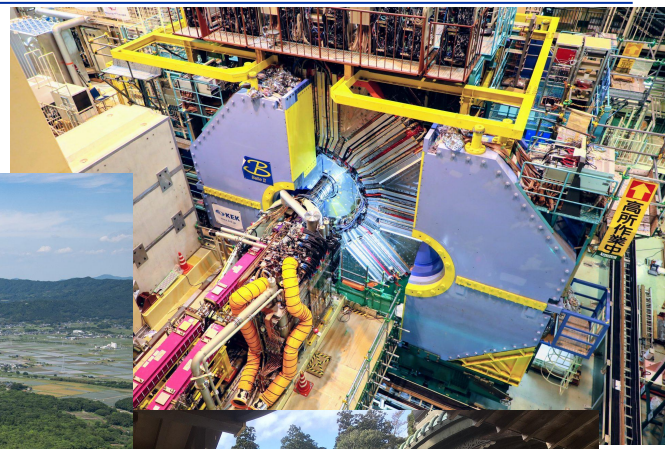
DIRAC & Rucio workshop 2023

- October 16-20 2023
- KEK, Tsukuba, Japan
- Stay tuned for more…

KEK

DIRAC
THE INTERWARE

RUCIO
SCIENTIFIC DATA MANAGEMENT



Additional information

http://rucio.cern.ch

https://rucio.cern.ch/documentation

https://github.com/rucio/

https://hub.docker.com/r/rucio/

https://rucio.slack.com/messages/#support/

rucio-dev@cern.ch

https://doi.org/10.1007/s41781-019-0026-3

https://twitter.com/RucioData

CERN