

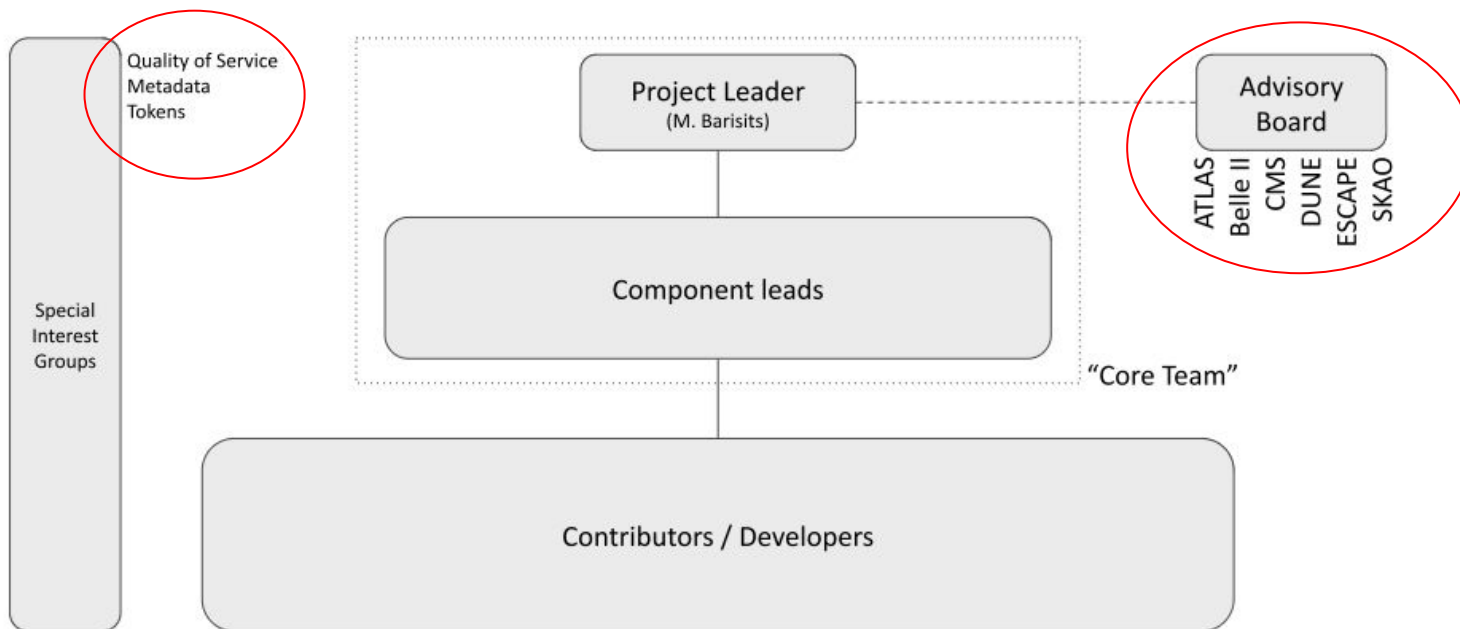


Rucio: State of the Donkey

Martin Barisits (CERN)

-
- **Igor V. Mandrichenko passed away in December 2023**
 - He was a big part of our community, working on metadata and consistency checking
 - Igor worked on Rucio contributions for DUNE and CMS
 - Since 2022 he was co-leading the Rucio consistency checking component
 - **He is dearly missed**

Organisation



Rucio Advisory Board



- «The primary function of the Rucio advisory board (RAB) is to provide expertise from representatives of Rucio communities and to advise the Rucio project leader.»
 - Long-term priorities and plans of communities
 - Advise on alignment of Rucio project objectives and plans
 - Identification of common objectives to form common development efforts
 - Advise on resource and person-power situation within the Rucio project
 - Advice on collaboration on funded projects
- Discussed at the workshop in Lancaster 2022
- Established in February 2023

Rucio Advisory Board



- Current composition

- ATLAS David South, DESY 2023-Apr → 2025-Mar
- Belle II Cedric Serfon, BNL 2023-Jun → 2025-May
- CMS Katy Ellis, STFC 2023-Feb → 2025-Jan
- DUNE Mike Kirby, BNL 2023-Feb → 2025-Jan
- ESCAPE Xavier Espinal, CERN 2023-Feb → 2025-Jan
- SKAO Rosie Bolton, SKAO 2024-Jun → 2026-May

- Three board meetings so far, another one in November (Two per year)

- Very successful to address difficulties (e.g. Personpower situation) and to shape the long-term evolution of the project

- Updated the [RAB mandate document](#), specifically describing eligibility criteria

Special Interest Groups



- **Quality of Service (Ended Dec-2023, Convener: Doug Benjamin)**
 - Objective: Storage-managed tape QoS
 - Lead to a [storage-managed QoS feature](#) now available in Rucio
 - SIG concluded
- **Metadata (Expected end Dec-2024, Convener: Rob Barnsley)**
 - Objective
 - Reach out to community and create a report about metadata evolution in Rucio (DONE)
 - Evolve metadata functionality based on communities need
 - Will very likely be extended!
- **Tokens (Expected end Mar-2026, Convener: Dimitrios Christidis)**
 - Objective: Token integration in Rucio

Personpower situation 1/2



- As with most projects, personpower situation is always fluctuating
 - Over the last year(s) several senior developers took up responsibilities in their collaborations
 - Somewhat available in mentoring/guiding capacity, but not active development
 - We were able to sustain this by introducing new motivated developers to the project
 - Alex, Dimitris, Enrique, Francesc, Hugo, Maggie, Riccardo
 - Big thank you to them and their collaborations/institutes!
 - Biggest difficulty
 - Sustainability and stability
 - Introducing new developers to the project is great! (More will come, see funded projects slides)
 - BUT: Project requires a small group of long-term/senior experts and cannot live on short-term contributors alone
 - Constant ~30 contributors per year (Core team + others)
- Component leadership team in a good state
 - 19 component leads from 10 Rucio communities/organisations!

Personpower situation 2/2



- CERN Research & Computing sector technical committee
 - Initiated an [Ininitiative](#) on Rucio
 - Two activities
 - CERN IT to participate in DevOps support for ATLAS & CMS
 - Establishing a Rucio reference data management service at CERN (For SMEs and other sciences)
 - Reference data management service should lay the groundwork for a possible reference compute service based on Rucio & DIRAC
 - Was approved by the RCS Steering Committee
- Hugo Gonzalez Labrador leading these activities for CERN IT
 - Took over Component leadership in Rucio for several components
 - Rucio instances for CERN SMEs
 - More in Hugos slides later this week

Communication



- Discussions in weekly [Rucio](#) meetings
 - News, DevOps roundtable, hot topics, developers roundtable
 - Planning meeting (3-4 month plan) for each major release
 - Release retrospective meetings well received!
 - **Everyone** is welcome to join!
- Yearly Rucio [community workshops](#)
- Yearly Rucio [coding camps](#) (hackathons)
 - Introduce new developers to the project; Spend some focused days on discussing ideas and implement them
 - No in-person hackathon this year
 - But we will have a Docu-Sprint in December

Communication



- CERN Mattermost
 - 300 users on Mattermost (Had 475 users on Slack pre migration 2023)
- eMail lists
 - rucio-dev@cern.ch for developer contact
 - rucio-users@googlegroups.com **DEPRECATED**
 - rucio-news@cern.ch for Rucio announcements (a few times per year)
 - rucio-contact@cern.ch Contact address on Rucio website

Release plan



- Three major releases in 2023

- 1.31 "Donkeys of the Caribbean" March 2023
- 32 LTS "The Good, The Bad and the Donkey" July 2023
- 33 "Eternal Sunshine of the Donkey's Mind" November 2023

- Three major releases in 2024

- 34 "Donkey Potter and the Data Cache" March 2024
- 35 LTS "Donkey and the Data Factory" July 2024
- 36 "Donkey Unchained" November 2024

- LTS support period

- 1.29 will be supported until **2024-07**
- 32 will be supported until **2025-07**
- 35 will be supported until **2026-07**

Funded projects 1/2



- Rucio community very successful in securing funding for the evolution of Rucio
- [DaFab](#) HORIZON-EUSPA-2022-SPACE, 3M EUR (Coordinator: DDN)
 - AI Workflows and data management for the Earth Observation domain
 - Started in March 2024, 3 years
 - Significant metadata developments for Rucio, more in Dimitris talk
- [OSCARS](#) (via HORIZON-INFRA-2023-EOSC-01-01)
 - 1st OSCARS open call
 - 13M EUR (max 200k EUR per project), 264 proposals, 58 selected
 - 3 Rucio proposals selected for funding!
 - Open Data, more on this by Hugo
 - ETAP, multi-rucio clients, more on this by Lia
 - MADDEN, rucio-fs etc., more on this by Lia as well

Funded projects 2/2

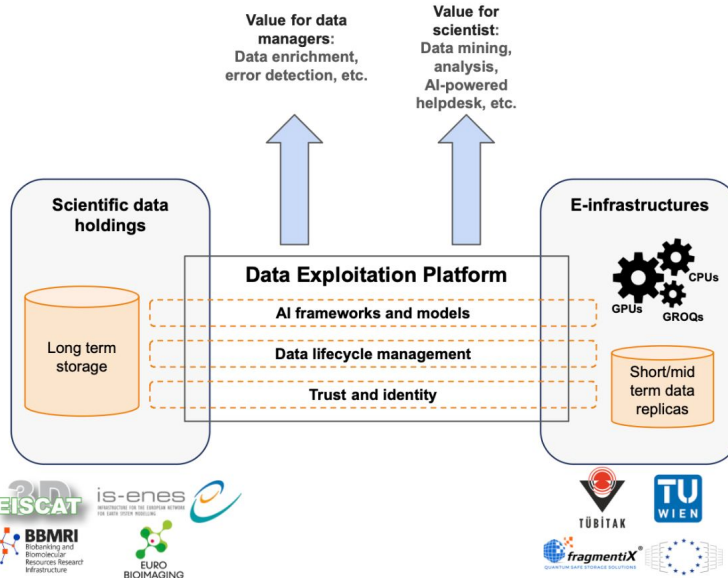


- Globus transfers (via [HEP-CCE](#))
 - Expand current Rucio Globus transfer/deletion capabilities and bring them to a production-ready state
- RI-SCALE HORIZON-INFRA-2024-TECH-01-01, 10M EUR (Coordinator: EGI)
 - Partnership between RIs and e-infrastructures
 - Shared goal: scalable data exploitation
- Remarkable:
 - Funding secured by different organisations for different scientific communities
 - All development will flow back to upstream Rucio!

RI-SCALE 1/2



RI-SCALE: technology innovation for scalable RI data exploitation



- **New EGI-coordinated project to start in March 2025**
 - Horizon-infra-2024-tech-01-01
 - 10 Million EC funding
 - 29 partners
 - 3 years
- **Partnership between RIs and e-infrastructures**
 - Shared goal: scalable data exploitation
- **Flexible policies for data replication**
 - Popularity, Sensitivity, Size, etc.
 - RUCIO and FTS
- **AI frameworks for data valorisation**
 - Itwinai, AI4HPC, BioEngine, BiolmageModelZoo, Hypha
- **Credit based usage accounting**
 - ARGO system
- **Interoperability validation with data spaces**
 - Destination Earth
 - Copernicus Data Space
 - EUCAIM cancer image federation

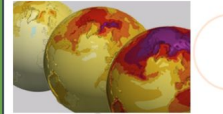
20

egi RI-SCALE use cases



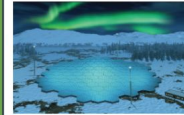
Climate modelling *CMIP6 data repository at DKRZ, 20PB in total*

- High-resolution **downscaling** of climate scenarios and **risk trend analysis** in agriculture
- Smart **detection of anomalies** in climate data usage



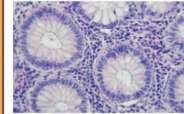
Near-Earth geospace studies *40 years of EISCAT and new EISCAT3D data, few PBs/year*

- **Intelligent Scheduling** of Radar Observations and Experiments
- **Space Debris and Anomaly** Detection



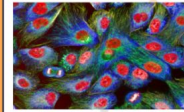
Cancer treatment *Cancer & digital pathology data at 3 BBMRI sites, 3 PB in total*

- Colorectal **cancer prediction** with explainable AI
- **Synthetic data** for computational pathology



Imaging in life sciences *Biologie Archive and EMPIAR DBs at EMBL-EBI, 4.5 PB in total*

- **Foundational models** for heterogeneous biological image data
- **Generative AI-Powered assistant** for data discovery and analysis



Legacy code and housekeeping



- Reminder from 2023 State of the Donkey:
- Rucio is a 10+ year old Software system
- While the principal design is rather future-proof, dependencies, certain features, at the time chosen best-practices, etc. are not
- We have to spend an increased amount of time on housekeeping
 - This is normal, but cannot be ignored
- Examples:
 - Web.py → flask migration **DONE**
 - Sqlalchemy 2.0 upgrade **DONE**
 - Temporary table workflows **ONGOING**
 - Dependency management **DONE**
- Need to fight dormant rod of unused features and remove them
 - Follower mode, rse limits vs. attributes, collection replicas vs. --deep, etc.

Work done last year 1/2



- Some highlights presented in these slides
 - This does not reflect the entirety of work being done!
- Clients
 - Complete Restructuring of client commands
 - Usability features (rich)
- Policies
 - Work on this continues
 - Long-term goal is to remove all community policies from upstream Rucio
 - They should only be used as policy packages
- WebUI 2.0
 - More on this later this week
- Jupyterlab extension
 - Migration to JupyterLab 4
 - New search features

Work done last year 2/2



- **Testing**
 - Static type checking
 - Code coverage
 - Better separation of unit tests, intake and integration
- **Tokens**
 - See Dimitrios talk later this week
- **Transfers**
 - Metadata for tape co-location
 - Bittorrent
- **Documentation**
- **Housekeeping**
 - Lots of quality of life improvements for developers

Work program 2025 and beyond



- **Metadata**
 - Will play an increasingly important role in Rucio
- **Housekeeping**
 - Fight dormant code rot
 - Deprecate unused features
- **Turn-key software / feature conciseness**
 - Parts of Rucio still heavily depend on semi-external pieces of code
- **Documentation**
 - Significant improvements over the last year, but still more to come
- **Tokens**
 - Third party copy in a good state
 - Client workflows next
 - Other Identity providers (CILogon) and non-WLCG profiles
 - Data embargos!

Work program 2025 and beyond



- **Protocols**
 - Migrate away from gfal
 - Use native protocol implementations instead
- **Cloud cost control**
 - More and more communities move their data to commercial clouds
 - Introduce features to safeguard/control for costs
- **Globus transfertool evolution**
 - Bring it to a production-ready state

- Ongoing collaboration between Rucio and DIRAC
- Several communities (Belle II, CTAO, JUNO, ...) are using/planning to use both DIRAC and Rucio
- Ongoing objective is to improve the interfacing between the two systems
 - [DIRAC & Rucio workshop](#) @ KEK last year
 - Very beneficial discussions between developers and administrators
- Next event: [Dirac & Rucio mini-workshop and hackathon](#)
 - Jan 16-17, 2025 @ CERN
 - Will have a hybrid part, but f2f participation is encouraged
 - Handling of tokens will likely be a big part of this discussion

PSA: Community page on our website



- Ambitious plan: List all communities who use Rucio in production
 - Not trivial, even I don't have an exhaustive list
- For now, all RAB-communities are listed on the website
 - Plan: Complete this list with all Rucio communities!
 - Please contact me to add your community, or submit a PR

<https://rucio.cern.ch/community.html>

Additional 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



http://rucio.cern.ch/doc../join_rucio_mattermost/

Developer contact



rucio-dev@cern.ch

Journal article



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

Twitter



<https://twitter.com/RucioData>

