



Rucio: State of the Donkey

Martin Barisits (CERN)

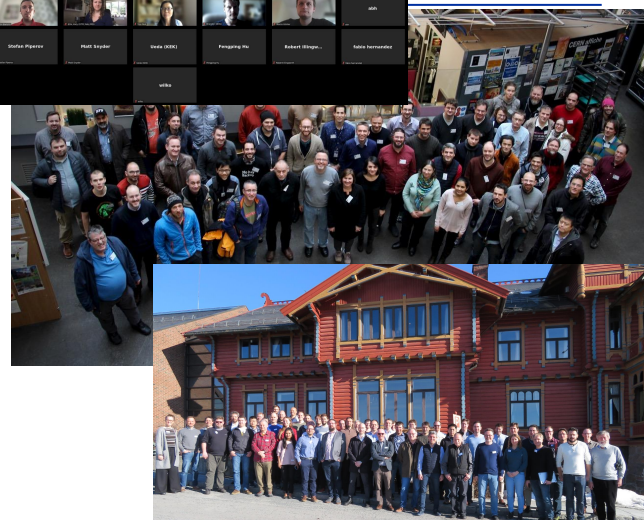
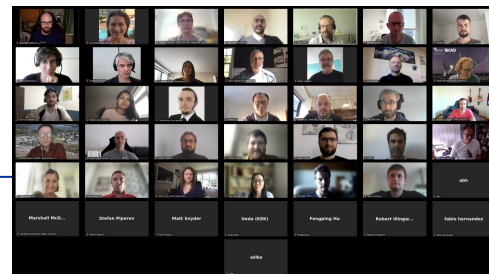
State of the Donkey

16 October 2023

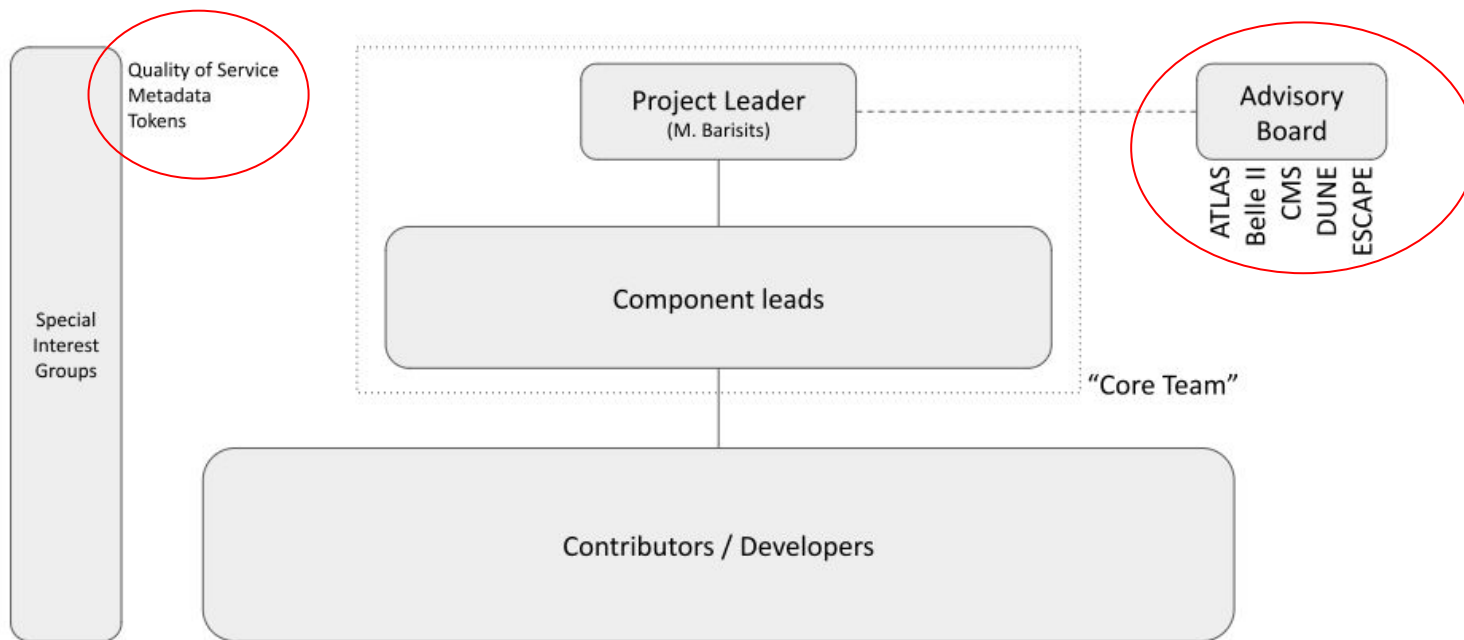


Welcome

- Welcome to the first DIRAC & Rucio workshop!
 - Not a DIRAC workshop AND a Rucio workshop but one joint workshop to bring both communities closer together!
 - Lots of interest to interface DIRAC & Rucio
 - Effort was initiated by Belle II several years ago
 - This workshop should initiate further integration of the two systems
- Thanks to the KEK team for the fantastic organisation!
- Thanks to the program committee!
 - Cedric, Federico, Andrei, Ueda, Eric



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
- **Two board meetings in 2023 (Minutes are internal)**
 - Initial meeting focused on state of the project, organisation, setup for future meetings
 - Second meeting focused on personpower situation and software evolution
- **Very successful and useful kick-off of the RAB**
 - Thanks a lot to the effort and motivation of the representatives!

Special Interest Groups



- One RAB recommendation was that all Rucio SIGs should have a well-defined structure
 - Worked with all SIGs to retroactively define
 - Well-defined objective
 - Expected end-date
 - Means to achieve the objective
 - SIG Convener
- Current SIGs
 - Metadata (Expected end Dec-2023, Convener: Rob Barnsley)
 - Objective: Reach out to community and create a report about metadata evolution in Rucio
 - Quality of Service (Expected end Dec-2023, Convener: Doug Benjamin)
 - Objective: Storage-managed tape QoS
 - Tokens (Expected end Mar-2026, Convener: Dimitrios Christidis)
 - Objective: Token integration in Rucio, more on this later

Personpower situation 1/2



- As with most projects, personpower situation is always fluctuating
 - More recently though, several senior experts took up additional responsibilities in their collaborations and will reduce Rucio development time
 - They will stay available in a mentoring/guiding capacity though
 - This is worrisome from a Rucio point-of-view; as senior experts cannot be simply replaced by new-hires
 - **Need to take extra care about long-term sustainability of project**
- [ALTO/Yale University](#) collaboration on transfer optimisations (FTS, Rucio)
- CERN participating in recently funded project DaFab “AI Factory for Copernicus Data at Scale”
 - Will fund metadata developments in Rucio
- [ESCAPE open Collaboration](#) agreement signed by CERN and many other organisations
 - Rucio is the data-core of the ESCAPE project
- Two Ukrainian students working on Rucio via [IRIS-HEP](#), no GSOC project in 2023

Personpower situation 2/2



- CERN Research & Computing sector technical committee
 - Initiated a [PSO](#) 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
 - Decision is on the RCS Steering Committee

Communication



- Discussions in weekly [Rucio](#) meetings
 - News, DevOps roundtable, hot topics, developers roundtable
 - Planning meeting (3-4 month plan) for each feature release
 - **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
 - Unfortunately none in the last years
 - Really need to bring those back!
 - DIRAC & Rucio hackathon!?

- Migration from Slack → [CERN Mattermost](#) last year
 - Mostly due to message retention reasons (Slack only kept ~1 month)
 - Data privacy
 - Being in better control of the service
 - 475 users on Slack on day of migration (February 2023)
 - 178 users on Mattermost today (October 2023)
- eMail lists
 - rucio-dev@cern.ch for developer contact
 - rucio-users@googlegroups.com for user discussion, basically dead
 - Plan is to deprecate and guide people to Mattermost
 - rucio-announce@cern.ch for Rucio announcements (a few times per year)
 - rucio-contact@cern.ch NEW - advertised as contact address on Rucio website

Rucio switch to semantic versioning 1/2



- Problem: Our versioning scheme does not follow [semantic versioning](#), the effective standard used in the python world, and thus is a source of confusion
 - X.Y.Z
 - Semver: X → MAJOR, Y → MINOR, Z → PATCH
 - Rucio (old): X → Constant, Y → MAJOR, Z → MINOR
 - **Thus our named/major releases look like semver-MINOR releases, which come with the assumption of general backwards-compatibility; This however is not true for Rucio;**
- Semantic versioning summary
 - MUST change MAJOR version when you make incompatible API changes
 - MUST change MINOR version when you add functionality in a backwards compatible manner
 - MUST change PATCH version when you make backwards compatible bug fixes

Rucio switching to semantic versioning 1/2



- Three named/major releases will become semver-MAJOR releases (32.0.0, 33.0.0, ...)
- Bi-Weekly minor/patch releases depend on their content
 - If they only include bugfixes they will become semver-PATCH releases (32.0.1, 32.0.2, 32.0.3, ...)
 - If they also include added functionality (in a backwards compatible manner) they become semver-MINOR releases (32.1.0, 32.2.0, 32.3.0, ...)
- This new versioning scheme will commence with the Rucio 1.32 “The Good, The Bad and the Donkey” LTS release line, which will become the Rucio 32 “The Good, The Bad and the Donkey”
 - **1.32.0 → 32.0.0**

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

- ~~○ 1.26 will be supported until **2023-07**~~
- 1.29 will be supported until **2024-07**
- 32 will be supported until **2025-07**
- 35 will be supported until **2026-07**

Python 3 / Containers



- From Rucio 32 LTS we require Python 3.9+
 - In the future we want to be more aggressive about Python requirements
- New legacy client compatibility policy
 - Clients backwards compatibility only guaranteed up to the oldest supported LTS release
 - Currently Rucio 1.29, 1.30, 1.31, 32
 - Older version MIGHT still work, but no guarantee given
- Switched all official Rucio containers from CC7 to Alma9

Some statistics



- Reminder: Commits and LOC not a good indicator to judge complexity of contributions
- Core repository ([rucio/rucio](#)) staying constant with ~50k LOC contributions a year
 - Constant ~30 contributors since 2018
- More and more work done in the secondary repositories though
 - [Jupyterlab-extension](#) 50k LOC added since 2020
 - [Documentation](#) 10k LOC added in 2023
 - [WebUI](#) 191k LOC (!!!) added in 2023
 - [Consistency-enforcement](#) 8k LOC added in 2023
 - And many more
- These secondary software products are fundamental part of the Rucio ecosystem

Legacy code and housekeeping



- 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
 - SQLAlchemy 2.0 upgrade
 - Temporary table workflows
 - etc.

Work done last year 1/2



- Some highlights presented in these slides
 - This does not reflect the entirety of work being done!
- Policies
 - More work has been done for a better integrated VO-policy handling
- Scalability
 - Re-factoring of large parts of the conveyor transfer workflows
 - New type of request dispatching within the conveyors
 - Introduce heartbeat handler to all daemons
- Consistency checking
 - CMS-style consistency checking added as new toolbox
- WebUI 2.0
 - More on this later this week
- Testing
 - Static type checking

Work done last year 2/2



- Housekeeping
 - Drop of function-based indices, deprecation of unused functionality, cleanup of tables, ...
- Tokens!
 - Start of SIG Tokens
 - See Dimitrios talk later this week
- Deprecation of hermes1 in favor of hermes2
 - Handles additional backends
- Transfers
 - Re-factored conveyor-throttler
- Documentation

Work program 2024 and beyond



- **Metadata**
 - Will play an increasingly important role in Rucio
- **Scalability**
 - Recent changes made to the conveyor should be brought to other daemons
- **Housekeeping**
- **Turn-key software / feature conciseness**
 - Parts of Rucio still heavily depend on semi-external pieces of code
 - Need to run an extra probe/cron job, or an oracle cron, have a dump available to guarantee certain types of core-system functionality
 - This should be done more concise
- **Documentation**
 - Setup tutorials, newcomer tutorials, etc. - Getting better, but still long way to go
- **Tokens!**
- **DIRAC integration**
 - Based on input from this workshop!

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>

