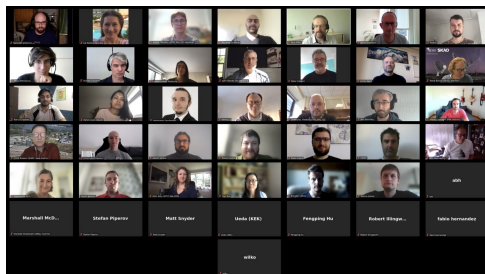# Rucio: State of the Donkey

Martin Barisits (CERN)

Welcome

- Welcome to our 5th community workshop!
    - First in-person Rucio workshop since 2020!
    - 46 in-person registrations + 40 remote
    - Thank you for attending!
- Thanks to the Lancaster team for the fantastic organisation!
- Thanks to the program committee!
    - Rob, Gareth, Kirby, Mario, Cedric, Eric

Organization 1/3

- Open and efficient development process
  - Software by the community for the community
  - Very motivated and capable development team - **Thank you!**
  - Contribution guide
- Component leads take formal responsibility of maintenance of their components
  - Plan patches and features, guide new contributors, review code
  - Component leads come from more than five different Rucio communities
- Project lead in charge of overall steering of the project
  - Overall technical guidance, final review of code contributions
  - Planning of deliverables based on **input of the community**
  - Progress tracking

Organization 2/3

- For each feature release (3 per year) - one release sprint lasts ~4 months
  - Release roadmap meeting, to plan the priority tasks for each component/developer
  - Weekly progress tracking
  - 1-2 checkpoints during sprint
    - Validate where we are vs. where we should be
    - Re-adjust roadmap and re-focus on highest priorities
- After each release we do a release retrospective
  - Somewhat similar to typical SCRUM retrospective, adapted to Rucio team peculiarities
  - Look at our processes in context of the last release
    - What went well, what could be improved, what will we do different?
 - Anonymous

Organization 3/3

Rucio 1.30 "The Donkeynator" development roadmap progress chart

Legend: Done, In Progress, Todo, Delayed, In Review

Rucio 1.29 retrospective, 2022-07-21

Aim of this retrospective is to look back at the 1.29 Rucio release and evaluate **what went well**, **what should be improved** and which **actions** we will take to do so. The focus is on the **processes**: roadmap planning (selection, effort estimation), weekly tracking and followup, differ from the goals, difficulties to identify the objectives, etc.

Specifically look on elements associated with the previous 1.29 release process, but this is also open to other project improvements.
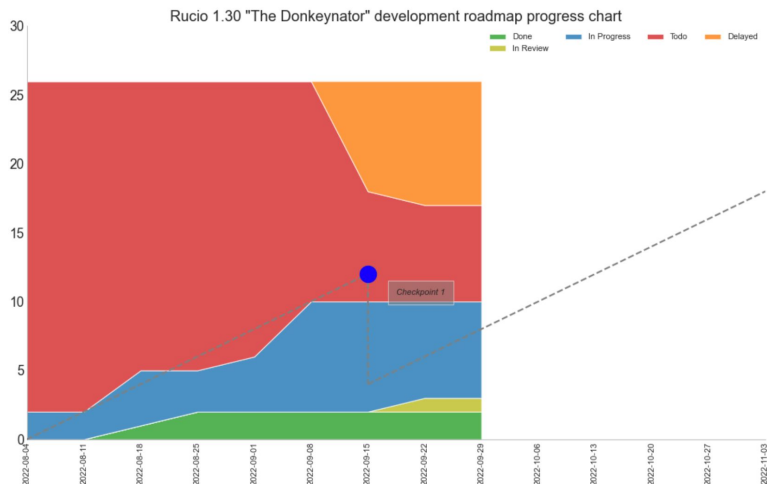
I like

- I didn't really work on the release. Sorry no comments.
- No hotfix releases needed in release line
- That we made some (a lot) improvements overall
- That we think more about clean code/ development
- Lots of progress in this release; many large features got done
- I like this mural idea as the format of this discussion
- Many ideas for improvements
- Good motivation

I wish

- More people would do reviews
- Output would be more in line with planning
- Having a hackathon again
- I wish the PR approval process was faster. Suggestion: consider making smaller releases combining groups of isolated changes and/or particular features ->
  1. prioritize the approval effort
  2. Shorten the development cycle
  3. Reduce the need to re-visit PRs due to other PRs getting merged ahead of time

We will

- Better clarify difference between patches and features in the contribution guide: Lots of contributions can be brought in as patches and do not need to wait for a major release
- Plan a hackathon: Next one will be a documentation hackathon
- Release planning: Increase awareness for realistic estimates and time it takes to implement issues.
- Release planning/followup: Have 1-2 checkpoints during a release sprint to decide which objectives can still be realistically achieved and potentially decide on re-focusing/dropping objectives
- Encourage people to do reviews

Rucio Advisory Board

- The way the Rucio project is organised is **key** to its success
  - «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»
  - Without much political and organizational overhead
- This process also has certain shortcomings
  - Cover long-term strategic perspective of involved communities
  - Address personpower and funding stream issues
  - Common submission/participation in funded projects
- Complement the existing organisational processes by creation of a **Rucio Advisory Board**

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

RAB - Mandate

- «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
- «The board provides non-binding strategic advice and is informal in nature.»

RAB - Composition

- The RAB is composed of a group of community representatives
  - «Goal is to complement the expertise of the development team, and thus representatives should have knowledge about the long-term strategy of their community and ideally have responsibility for their project programme and budget.»
- By invitation of Rucio project leader
  - Suggestions for expanding the board should ideally come from the RAB itself
  - 2-year, renewable terms
  - Invitations to join the board are made to the respective community
    - Nominates one, exceptionally two, representatives
    - Nominations discussed with the Rucio project leader
- Rucio project leader ex-officio member and chair of the RAB

RAB - Organisation

- RAB meets as needed, nominally twice a year
- Minutes are kept internal to the RAB
  - Members should feel free to openly discuss
- Next steps
  - Initial composition of the RAB will include representatives of the 4-5 most active Rucio communities
  - Invitations to these communities will go out in 2022
  - First meeting of the RAB planned sometime early 2023
- Probably no RAB composition extension in the first 1-2 years
  - Need to get settled in first

Communication 1/2

- 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 2020, 2021 & 2022 due to ongoing pandemic
    - Will be back in 2023!

Communication 2/2

- Slack
  - Open to everyone: Users, Operators, Developers
  - 350 users (2021) 474 users (Today)
- eMail lists
  - rucio-dev@cern.ch for developer contact
  - rucio-users@googlegroups.com for users, not massively used - most people head to Slack
- Special Interest Groups
  - Launched two SIGs in 2021
  - Forums to specifically discuss ongoing activities of these areas
  - Open to anyone (organized via public CERN e-groups, message me if you want to join without CERN account)
    - RUCIO-SIG-Metadata - Metadata evolution in Rucio
    - RUCIO-SIG-QoS - Quality of Service evolution in Rucio (and beyond)

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

- 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)
  - 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, …)

Work done the last year

- Huge program of work has been done the last year
  - Impossible to list all contributions
    - Apologies if I missed a specific contribution
  - Thanks to every single developer and organisation supporting Rucio
- Transfer system evolution 📅
- VO specific policies in Rucio
  - VO specific tests in the CI Framework
- Static type checking and CI framework evolution
- Cloud storage (Amazon, GCS, SEAL, …) 📅
- CMS-style consistency checking 📅
- WebUI 📅
- Documentation
- OIDC Tokens 📅

Work done the last year

- Hermes 2
- Database optimization (Switch workflows to temporary tables)
- Data rebalancing rewrite
- SIG Metadata and QoS work

SIG Status

- **Two Rucio Special Interest Groups**
  - Metadata
  - Quality of Service
- **Metadata**
  - Lots of progress
    - Inequality operators for metadata queries
    - New FilterEngine
    - New metadata plugins: MongoDB and external Postgres
  - More on this in the metadata session this afternoon
- **Quality of Service**
  - Support for self-managed QoS (Big disk buffers in front of tape) added
    - Storage decides when to transition QoS, but Rucio can ask for a minimum time to keep on disk
  - Currently production testing within ATLAS

Documentation

- Substantial documentation overhaul
  - https://rucio.cern.ch/documentation/
  - Markdown based
  - REST API based on OpenAPI spec
  - Full client API
- Old RTD documentation has been removed
- Very easy to contribute!

Getting started with Rucio

- Still a difficult process to get started with Rucio
- Lots of assumptions which are foreign for many new communities
  - GRID Storage and workflows
  - Authentication and Authorization infrastructure
  - FTS
  - Database
- Lack of documentation
- There has been progress in making this easier for people (k8s tutorial, …) but it is still a difficult process
  - Frustrating for newcomers
  - Time-draining for the Rucio team to guide newcomers
- We **need** to invest time here to make this experience better

DocAThon

- Host a DocAThon in early 2023 with the specific goal to improve the documentation & tutorials for newcomers
- 2-3 full days at CERN
- Step by step guide to get to know Rucio
  - Document about the environment Rucio "lives" in
    - GRID Storage, AAI, FTS
  - Different ways to deploy Rucio
    - Kubernetes
    - Non-k8s
  - Tutorial
  - …
- Please consider participating!

Python 3

- Fully dropped Python 2.7 with Rucio 1.29 (July 2022)
- Clients & Server now py 3.6, 3.7, 3.8, 3.9 compatible (In principle 3.10 + 3.11 as well, but we need to add it to CI)
- 3.6 already in EOL since 10 month
  - 3.7 will follow soon
  - Some packages already dropped 3.6 support
- Would a more aggressive Python requirement (Both Rucio clients & server) be an issue for any community?
  - Require Python 3.9 with Rucio 1.31 (March 2023)?
  - And then also push this more aggressively and not get into an Python-EOL situation
  - Many communities wrap/containerize their clients, but possibly not all?

Common containers

- Currently our common Rucio containers are based on CentOS7
  - Servers, Daemons, Clients
  - In principle supported until 2024-06
  - Ideally we prefer OS versions which natively package the software we need (Python 3.9, …)
- Move to CentOS stream 9 (CS9) with Rucio 1.31 (March 2023)?
  - Linux future committee meeting does not generally recommend it
    - Some issues encountered upstream
    - Concerns with release progress
  - Issues for deployments such as Rucio probably negligible though
  - Opinions?
- In principle we are not bound to CentOS as distribution
  - Should we consider other OS?
  - What is the impact for communities who have derived Rucio containers?

Google Summer of Code

- Rucio involved in GSOC since 2017
  - Very good experience with very bright students
  - https://hepsoftwarefoundation.org/activities/gsoc.html
- 2021
  - Rucio and CS3API to enable data management for the ScienceMesh cloud - Rahul Chauhan
  - New protocols for exascale data management with Rucio - Rakshita Varadarajan
- 2022
  - Rucio WebUI 2.0 - Nimish Bongale
    - More on this in the afternoon

Work program for 2023 and beyond

- Fine-grained OIDC token workflows
  - Makes Rucio the token-issuing component in the data management infrastructure
    - Enable communities to enforce data embargos
  - More on this in the afternoon
- Policies
  - Lots of work has been done to make Rucio flexible to VO defined policies
    - Server side, client side
  - Still some historic ambiguity: surl vs lfn2pfn methods; permissions vs policies
    - There is a risk with this ambiguity getting worse with fine-grained tokens → auth layer? Permission layer? Replica management layer?
- Scalability
  - Rucio is highly scalable
  - Workload scalability could be handled more dynamically
    - Workload aware activation/deactivation of agents (k8s pods) → Small, but positive impact for energy usage
  - Change of database managed partitioning to a job-dispatcher system

Work program for 2023 and beyond

- Consistency checking
  - Further integration of CMS-style consistency checking to Rucio
  - Work ongoing
- WebUI 2.0
  - State of the art web experience compatible and accessible on mobile devices
  - More on this in the afternoon
- SIG Metadata and SIG QoS
  - Further metadata evolution
    - Simplify metadata interaction (list_dids, list_dids_extended)
    - Include further external systems - IVOA?
  - QoS: One specific use-case implemented and being tested
    - Need to understand further needs and evolution

Work program for 2023 and beyond

- Documentation
  - See previous slides
  - DocAThon
- Hermes1 to 2 migration
  - Enable Rucio event submission to other services (Elastic, influx, …)
  - Hermes2 there, need to deprecate hermes1 and migrate
- Workflow system for rules
  - Right now rule workflows are mostly driven by external systems (Production systems)
    - This requires them to know more about data management than they ideally need to
  - Define workflows/policies for rules
    - E.g. Have a rule with 3 copies on disk for 2 month, then reduce to 2 copies for 2 month, then remove the rule and create a rule for tape for 3 years
    - Transition triggers do not need to be time-based, but could also be event-based
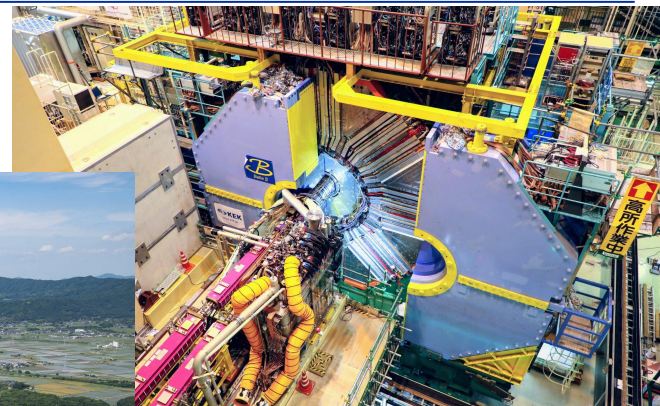
Work program for 2023 and beyond

- Housekeeping
  - Initial Rucio code now 10+ years old
  - Refactorization, maintenance becomes increasingly important
    - Review old code, remove unused parts → Remove unnecessary complexity
    - Review dependencies
      - Are they maintained and still well-used by the community
      - Security considerations
  - Well-maintained code makes the barrier of entry for newcomers easier
  - Examples
    - Migration to SQLALchemy2
      - Not complicated, but tedious

DIRAC & Rucio workshop 2023

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

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		https://rucio.slack.com/messages/#support/
Developer contact		rucio-dev@cern.ch
Journal article		https://doi.org/10.1007/s41781-019-0026-3
Twitter		https://twitter.com/RucioData