Rucio & ESCAPE

Long term collaboration

Martin Barisits
on behalf of the Rucio team

RUCIO
SCIENTIFIC DATA MANAGEMENT

Rucio in ESCAPE

- ESCAPE is a huge accelerator for Rucio
  - Thanks to ESCAPE Rucio is truly established at non-HEP sciences!
- Fantastic collaboration with ESCAPE communities
  - Lots of interesting and fruitful work for ESCAPE and the core Rucio team
  - ESCAPE and ESCAPE communities are a fundamental part of the Rucio community
    - First astronomy dedicated session at last Rucio workshop!
- Rucio strongly evolved thanks to ESCAPE contributions
  - Metadata
  - (Token) Authentication
  - Deployment
  - Documentation
  - Jupyter Rucio plugin
  - …

How is the Rucio project organised?

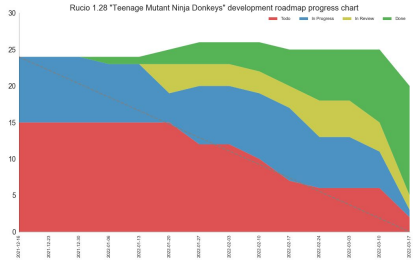
- Open-source community project
  - Contributions and participation is very welcome!
  - 2021: 46 distinct contributors to the codebase, only 13 coming from ATLAS
- Architecture is split into vertical/horizontal slices → Components (~30)
  - Each component is lead by a community developer responsible for the general maintenance of the component and go-to expert to coordinate contributions to it
  - Two Rucio components are lead by ESCAPE members (Rizart Dona & Rob Barnsley)
- Communication
  - Weekly Rucio meeting
    - News, DevOps & deployment discussions, developer roundtable
    - Feel free to join!
  - Yearly community workshop
    - State of the community
    - Long-term direction of the project

How is the Rucio project organised?

- Communication
  - E-mail list: rucio-users@googlegroups.com (Users), rucio-dev@cern.ch (Developers)
  - Slack!
- Special Interest Groups
  - Dedicated forum for a specific topic in Rucio, include non-Rucio experts as well
  - No strict “format”, some SIGs do regular meetings, others just on-demand
    - + Mailing lists and Slack channels
  - Currently
    - SIG-QoS: Quality of Service evolution in Rucio, topic right now is large disk buffers in front of tapes
    - SIG-Metadata: Metadata evolution in Rucio, Generic metadata, metadata operators, external metadata stores
  - Future
    - SIG-Tokens: Token evolution and data embargos in Rucio

How is the Rucio project organised?

- Release cycles
  - 1 patch release every 2 weeks (e.g. 1.27.X)
  - 3 feature releases per year (e.g. 1.28.0)
    - One of them being a Long-Term Support release, supported for 2 years
- Planning
  - Strategic direction discussed at Rucio workshop
  - Dedicated planning event for each feature release
    - Defines the 4 month priority plan which is monitored weekly
- Currently preparing the creation of an open steering board
  - Informal; Goal is to understand community priorities, identify common goals, collaboration and funding possibilities
  - Meets 2-3 times a year, details still in preparation
  - ESCAPE and ESCAPE communities should participate!



Future developments (of interest to ESCAPE)

- Non-exhaustive list
- Metadata
  - Support community needs in storing and searching generic metadata
- Quality of Service
  - See Paul's talk from yesterday
  - Complex topic involving the entire data architecture
  - Currently driven by specific use-cases (e.g. large disk buffers in front of tape)
- Tokens and data embargos
  - Some token functionality exists in Rucio, but it needs to be further evolved
  - Introduce fine-grained tokens into our data workflows
    - This will enable Rucio to truly enforce data embargos

Future developments (of interest to ESCAPE)

- Better documentation → New markdown based repository
- Make running and deploying the system simpler
  - Rucio is a very complex system able to manage advanced workflows at large scale
  - Lots of complexity
    - Partly coming from Rucio itself
    - Partly coming from the larger infrastructure: K8s, Databases, Certificates, Storage, …
  - It will probably never be trivial, but we still have a long way to go to make it “easy”

“Make the easy things easy, and the hard things possible”

ESCAPE long term collaboration

- Our goal is to actively include communities in the project
  - To sustain and evolve Rucio, we need community involvement
  - Not everyone can develop code, but there is also: documentation, support other communities, …
- Many possibilities for future collaboration, my thoughts:
- Having an ESCAPE representative (Riccardo & Rizart) in the weekly Rucio meeting and as a contact person in general was very beneficial
  - Summarize & handle ESCAPE issues
  - Shields Rucio team
    - Some issues are just handled within ESCAPE
    - Rucio team does not need to discuss the same issues with multiple people

ESCAPE long term collaboration

- ESCAPE as a driver for requirements/use-cases/priorities
  - Having ESCAPE collect & digest requirements/use-cases/priorities and communicate them to the project as a single voice
    - Many workflows and use-cases (e.g. astronomy) are similar among communities
    - Make sure all bases are covered
    - Relieves small communities from having to join Rucio meetings/discussions themselves
  - For developments like metadata, tokens, or data embargos it will be crucial to cover the use-cases and requirements from the beginning
    - Example: data embargos
      - This will be included in the design of the token evolution from the beginning
      - We need to truly understand the use-cases from communities who actually need these features
    - Later changes in the design might be difficult and costly

ESCAPE long term collaboration

- Deployment and infrastructure best practices
  - Many ESCAPE communities are on the verge of starting their own Rucio deployments
  - ESCAPE has lots of experience in setting up data infrastructure
    - Not only Rucio, also storage, authentication, data workflows, etc
  - This past and future experience should be captured in documentation & best practices in Rucio
  - ESCAPE should be a support/guidance forum for these communities in deploying Rucio as well
- ESCAPE Rucio instance
  - Was extremely useful for communities to get to know Rucio, try out actual workflows and experiment with the infrastructure
  - I would recommend to continue this for the foreseeable future

Thank you all for the inspiring years of collaboration

Community


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


LCLS


Belle II

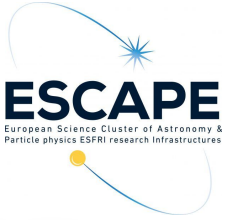

Science & Technology
Facilities Council


COMPASS


CERN

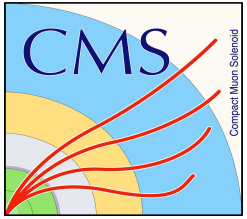

ARC


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

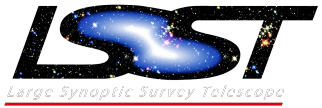

DUNE
DEEP UNDERGROUND
NEUTRINO EXPERIMENT


egi


3D
EISCAT


VIRGO


cta
cherenkov telescope array


LSST
Large Synoptic Survey Telescope


IceCube
SOUTH POLE NEUTRINO OBSERVATORY


FTS


eXtreme DataCloud


Fermilab


PAWSEY
supercomputing centre


SKA
SQUARE KILOMETRE ARRAY


LIGO


Open Science Grid


BROOKHAVEN
NATIONAL LABORATORY


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