



# Development roadmap discussion

---

*Martin Barisits* (CERN)

# Development roadmap discussion

---



Non-prioritized, cherry-picked, list of developments/ideas to discuss, some have been covered in previous talks; Goals of this session: Better understand community needs and priorities to derive development roadmap

- Tokens
  - Covered in yesterday's discussion
  - CILogon and other token providers, alternative token profiles
    - Needs to be addressed by community developments and cannot be covered by Dimitrios alone (Interface definition vs. specific implementations)
      - Anil and James already working on it
      - I believe we need to structure this work with clear objectives/expectations
  - Token timelines for non-WLCG communities
- DIRAC Interface
  - [DIRAC & Rucio mini workshop Jan 16&17 2025](#)
  - Implication for token workflows need to be understood / defined

# Development roadmap discussion

---



- **Data embargos**
  - We have a “concept of a plan”
    - Naturally this is tightly linked to token development and must be addressed in the permission layer of Rucio
  - More specific architecture, involving data access, rules, dids, metadata is a bit unclear though
- **Hierarchical data**
  - Major issue is that containers can only hold datasets but not files, making a representation of a directory tree difficult
  - One idea would be to entirely drop the concept of containers and only have files and datasets, with datasets being able to hold files and/or other datasets
    - Probably there needs to be settings (on the dataset) to enforce this (only\_files / only\_datasets)

# Development roadmap discussion

---



- **Metadata**
  - Big evolution coming (See Dimitris talk)
  - Metadata backend plugins will stay in Rucio!
  - Major plan is to deprecate the `did_columns` plugin, remove HEP specific columns from the schema and migrate ALL metadata to the JSON approach
  - Metadata schema enforcement
- **DC Inject tool**
  - Dedicated work on this coming from ATLAS
- **Tokens in Jupyterlab Extension**
- **Collection replicas**
  - We want to drop them
  - List-replicas forces `--deep` since a while, no noticeable performance impacts
  - No solution yet for `list_dataset_replicas_bulk` (should not be too hard)
  - No solution yet for `list_datasets_per_rse` (Unclear; can be derived from replica & contents, at a price)

# Development roadmap discussion

---



- **S3 credentials**
  - Should take token development opportunity to unify S3 (signed url) workflows
  - FTS should not need to store these credentials in parallel to Rucio but request signatures, on demand
- **Documentation**
  - E.g. Document staging RSEs (Staging without transfer)
  - DocuSprint in December!
- **Rucio-native, generic, in-situ file registration**
  - We have tried this before, but this always diverged into domain-specific needs
  - Maybe there is a path for a generic tool, but it is not clear
  - Documentation should be improved though
- **Protocols**
  - Migration away from gfal2 will take significant effort
  - gfal3?

# Development roadmap discussion

---



- Rules policies
  - Current idiosyncrasy that lots of data management knowledge needs to be packed into external systems which drive Rucio
    - Duplication of knowledge → Some of this could be moved into Rucio instead
  - Define policies (like workflows) on Rules e.g.
    - Think of it as Rule workflows based on pre-defined policies
    - After trigger action (time based, dataset close, external push, etc) remove/add a rule
  - Significant development

Other discussion points?

# 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/](http://rucio.cern.ch/doc../join_rucio_mattermost/)

Developer contact



[rucio-dev@cern.ch](mailto:rucio-dev@cern.ch)

Journal article



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

Twitter



<https://twitter.com/RucioData>

