

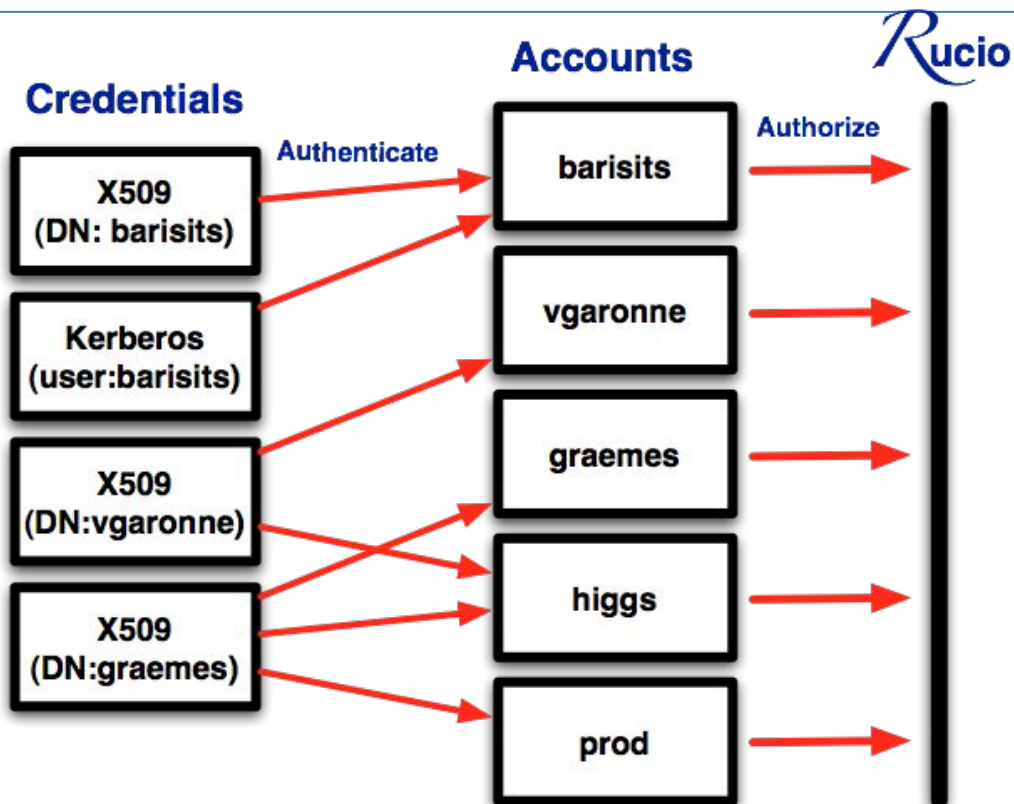


Chapter 1: Namespace

Rucio Account

- Entry point and unit of assigning privileges
 - Flexible management between credentials and group, activities, ownership, permission, accounting, quota, etc.
- A Rucio account can represent
 - Individual user, e.g., user account
 - Group of users, e.g., group account
 - Organized activity, e.g., service account
- Users are authorized to log to an account using their credentials
 - E.g.. Userpass, X509, GSS, etc.
- The same user credential can be used to connect to several accounts

Example Account



Rucio Namespace

Data is federated in a single namespace providing a logical view and transparent access of data across multiple locations:

Logical
view:

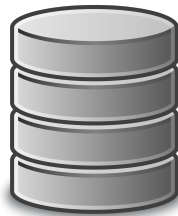
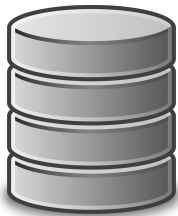
`MyScope:MyFile`

`MyScope:Mydataset/MyScope:MyFile`

Storage X at location A

Storage Y at location B

Storage Z at location C



Physical
View:

`/pnfs/site.fr/rucio/MyFile`

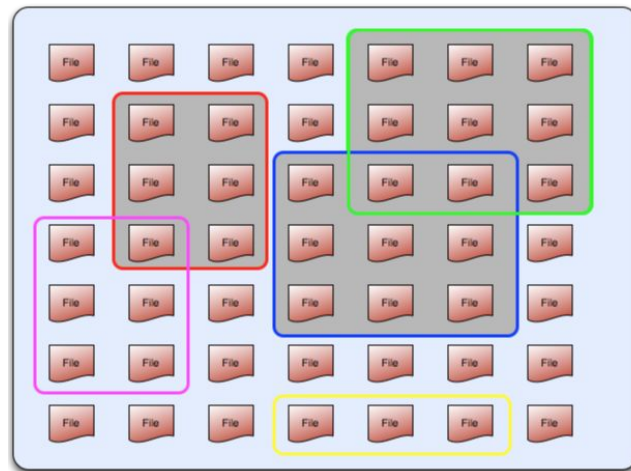
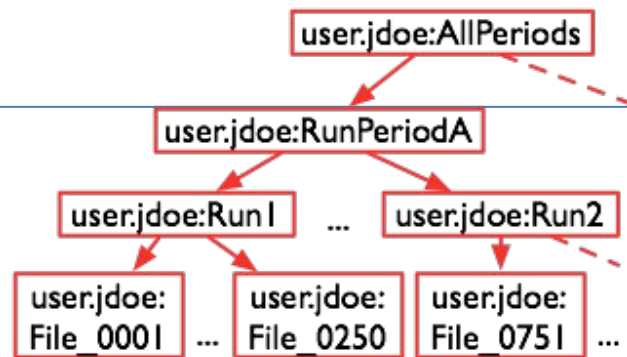
`/dpm/site.uk/rucio/MyFile`

`/rucio/MyFile`

Data Hierarchy

- At the heart of everything is a file*
- Files are grouped into datasets
- Datasets are grouped into containers
 - Datasets only hold files
- Containers are grouped into containers
 - containers only hold datasets or containers
- Collections can be organised freely
 - Files can be in multiple datasets
 - datasets can be in multiple containers
 - containers can be in multiple containers

* sub-file support being explored



Data Identifier (DID)

- Data Identifier (DID) is the primary addressable unit of data
 - DIDs can be either files, datasets or containers
- Files, datasets and containers (DIDs) are identified by **{scope} : {name}**
 - A data identifier is uniquely identified over all time and cannot be reused
 - Files, datasets and containers follow the same identical naming scheme
 - All Rucio commands work with DID independently from its type
 - E.g., `$> rucio download {scope}:{name}`
 - Where **{scope} : {name}** can be a file, a dataset or a container

Scope

- The scope is part of the data identifier (DID) `<scope : name>`
- The data identifier scope isolates the namespace in several sub-spaces
- The data identifier name is unique within the scope
- A scope is mapped to an account
- An account can have access to multiples scopes
- The scope helps users to immediately associate a data identifier to a specific group of data, such as a project

Examples of {scope}:{name}

User data

Scope: User name, e.g., user.jdoe:this.is.my.test.file001
user.jdoe:this.is.my.dataset1

Group data

Scope: Group name, e.g., group.thebest:this.is.our.test.file001
group.thebest:this.is.my.dataset1

Detector data

Scope: data + <year>, e.g., data18:16TeV.00199926.calibration.daq.RAW

Reprocessed data

Scope: Campaign +<year> + real data scope, e.g., Repro18Data18:00169783.AOD.r2059

Metadata

- Metadata are custom attributes on data identifiers
 - Support for arbitrary metadata being explored
- Rucio supports different kinds of metadata
 - System-defined, e.g., size, checksum, did_type, is_open, created_at
 - Physics, e.g., number of events, GUID
 - Workflow management system, e.g., which task or job produced the file
 - Data management, necessary for the organisation of data
- Metadata provides another namespace
 - Datasets are searchable by name and metadata

Pluggable Policy & Permission

- Rucio offers the possibility to enforce the semantic and type constraint for some attributes
 - A Data Identifier name can have a naming convention enforcement
 - A metadata value can have a type constraint and semantic enforcement (e.g., enum)
- Rucio supports configurable permissions to reflect collaboration policies
 - Permissions are defined for an account and an operation on resources
 - E.g., Which accounts can write in a scope

Questions?

If you have a question but don't get the chance to ask it directly during the session, you can do it here: <https://goo.gl/BdSGoC>