



# ESCAPE

European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures

## Thoughts on future of Storage Orchestration and Quality of Service

Paul Millar (DESY)



# Introduction

“Trying to predict the future is a mug’s game. But increasingly it’s a game we all have to play because the world is changing so fast and we need to have some sort of idea of what the future’s actually going to be like because we are going to have to live there, probably next week.” Douglas Adams

So, here are some thoughts on “the future”:

Not going into ESFRI-community specifics.

I did ask some others to share their thoughts: my thanks go to Agustin, Fabio, Martin, Mieke and Yan.

Just to make it abundantly clear: none of this should be even remotely considered a roadmap for any software or service.



# Storage technology, etc.

## Magnetic disks:

Pressure to increase aerial density will continue: LMR → PMR → SMR → HAMR?. This may involve trade-offs (→ QoS classes)

## SSD:

Increasing in popularity, but likely to remain a niche for scientific storage.

## Tape:

Future is uncertain, but true for last ~20 years. Consolidate has left tape vulnerable (see LTO-8 vs LTO-M8 debacle). Sites exploring alternatives.

## Exotic new technology:

Continue to see articles extolling tech. X. Almost all will never be products.



# Rucio will support more transfer services

Rucio **supports FTS**, which is a kind of Swiss-army knife for transferring data. However, it requires someone to provide the service and sites to use supported storage solutions.

**Globus support** is being added to Rucio, mostly driven by HEP communities in USA needing access to HPC facilities.

Rucio is a **very attractive** storage orchestration solution, even for communities that do not want to run their own FTS service.

Could “buy” an FTS service (from EGI or EOSC) but there are other possibilities: commercial (e.g., IBM Aspera) and bespoke/in-house.

Somewhere between “possible” and “likely” that someone will **add support** for another transfer service Rucio.





# Better integration with workflow/compute engines

Basic tenet: **move compute to storage**. Compute needs to know where data is located, possibly moving it around.

Users don't want to manage data, they want to “do science”.

Pressure to hide data management from users

DLaaS is an **excellent first step**, but I suspect the user experience will improve over time.

We may see sites offering compute resources more at the **OpenStack/K8 level**. Compute resources provisioned dynamically:

User wants to do an analysis: the system selects the best site, moves any missing data, spins up a new JupyterHub instance.



# QoS model unlikely to evolve once storage provisioned

Provisioning hardware involves **agreements** (MoUs or contracts).  
Adjusting QoS model may involve renegotiating those agreements.  
May be hard to justify against the benefit of an additional QoS class.

- Agreements are organised by different people from data management.  
Further complicates the communication.
- There is also the “don’t change a working system” maxim.

Likely a **gradual shift** (setting concrete) rather than a sharp cut-off.

Innovation is **still possible**; e.g., in which QoS classes files are stored,  
dynamic changing QoS class, incrementally tweaking the definition of  
QoS classes.



# Current “ESCAPE” QoS model is good enough

RSE has a **single QoS class**. Data managers reference this QoS class **directly**. QoS transitions (currently) always **trigger a transfer**.

Scientific communities are likely to see a fairly static set of QoS classes:

- Adding new QoS class is (relatively) hard.
- New work-flows’ storage requirements taken from existing QoS classes.

Overhead from transferring files is likely **negligible**.

Rucio could identify **transfers that could be QoS transitions**, allowing storage to optimise this.



# QoS will evolve, but slowly

Example is MAS: a very large disk buffer in front of tape. This is forcing both **compute and storage orchestration** to evolve.

QoS is (and will likely remain) a low priority for many communities.

We're likely to see small, focused, incremental changes to Rucio that are trying to solve **specific "pain points"**. Progress will likely depend on those communities providing the necessary resources.

The **challenge** is for Rucio to maintain a simple, consistent model, avoiding ad-hoc and community-specific solutions.

Supporting some kind of plug-ins is an attractive solution, but risk fragmentation.





# Better integration between DataLake and HPC

Predict a **shift** from Rucio supporting (almost) exclusively HTC workflows to supporting some HPC work-flows.

Supporting HPC facilities in DataLake becomes increasingly important.

Two approaches: deploying caches and including HPC storage as RSEs.

EU funded project “interTwin” (starting 2022-09) may help.

In USA, the answer is Globus\*. Elsewhere (e.g., here in Europe) it’s more complicated. The **FENIX solution** (Swift edge storage) is one approach, but has some drawbacks.

May see an **solution expansion** (with several competing approaches) before there is consolidation, as sites wish to reduce their operational load.



# Conclusions and final remarks

There are a number of predictions here.

If (as planned) ESCAPE continues to exist (in some form) beyond the official end of project, we could use that forum to provide input on these points and help develop a common approach.

Who knows, in five years time, we may be able to meet up in person for a retrospective talk on which predictions (if any) turned out to be right.



# Thanks for listening !

