



# Rucio project

Robert Illingworth

SCD Projects

3 October 2019

# Rucio “project” overview

- This is a bit complex, because there are at least four interacting projects here
  - The CERN Rucio project
  - Rucio for CMS
  - Rucio for DUNE
  - FNAL Rucio (including IF)
- There’s a lot of overlap, but no single person is in charge of everything

# Rucio for DUNE + IF

- Mostly overlapping people and tasks between DUNE and IF
- DUNE does not yet have a formal Rucio DM project (primary focus has been ProtoDUNE)
  
- Short term objectives
  - Complete deploying Kubernetes instance at FNAL
  - Replace SAM file catalogue with Rucio
- Medium term
  - Replace SAM metadata catalogue with something Rucio integrated
  - Fully integrate Rucio into DUNE workflows (POMS? DIRAC?)
- Longer term
  - Push this out to all ongoing IF experiments
    - Exactly what “ongoing” means needs to be decided

# DUNE + IF Effort

FNAL	FTE	
Robert Illingworth	0.25	Technical lead
(Steve Timm)		DUNE DM convener; some ops
Igor Mandrichenko	0.1	Metadata design
Brandon White	0.75	Everything else
DUNE non-FNAL		
Edinburgh	0.6	Monitoring; Rucio configurability; QoS?
Other UK	?	Site DB (CRIC?) evaluation

This is barely sufficient for the short term tasks; insufficient for more

# CMS Rucio

- Had hoped to complete a transition to Rucio by early CY 2020. Will not make this.
  - Very difficult to replace a running system of the complexity of PhEDEx + Dynamo
- Targeting summer 2020. Run 3 begins in spring 2021
  - Lack of disk space overhead is a concern
- Data consistency is necessary for the transition. Popularity-based replication should be there shortly thereafter
- Recent losses, especially at Fermilab are a concern if the effort is not replaced soon (typically a one-month ramp-up period for new people)

# CMS Effort

FNAL	FTE	
Eric Vaandering	0.75	Project Management; development; Kubernetes integration
Natalia Ratnikova	0.5 → 0	Development, database integration
(Nick Smith - PPD)	(0.25)	Development, operations
(Pradeep Jasal - Intern)	(0.1)	Operations
Yuyi Guo	0 → 0.75	Replace Natalia
Igor Mandrichenko?	0 → 0.6	?
Non-FNAL CMS		
Stefan Piperov (Purdue)	0.3	Exploring data consistency (new effort)
UK	0.25	on tape testing, integration — could use help
UK	0.25	popularity generated replicas (new effort, will need help)
IT	0.25	on user data management

# Some R&D proposals

- Improve Rucio dataset scalability
  - Rucio dataset rule management is designed for ATLAS workflows; individual datasets can't get too big or rule evaluation gets very slow
- Advanced data stores
  - Object stores; etc. How does Rucio interact with non-traditional file storage?
- Storage Quality of Service and Data Lakes
  - Provide Rucio with a view of tiered performance in a storage system; how and when to transition between tiers
  - Rucio core project is already starting to think about this; there's scope for us to take on part of it
- Clearly the current effort is not sufficient to do these on top of everything else