

# Metadata in Rucio

Which metadata should be in Rucio  
Lumiblock  
Rucio dev considerations

# What criteria for Rucio metadata

- ATLAS has a metadata catalogue – AMI
  - populated when Prodsys tasks finish
  - users can search here
- Rucio: subset of metadata needed promptly by upstream components
  - Prodsys needs metadata, perhaps before AMI harvests
    - #events/size for split & merge
    - lumiblock merge: AOD merge, AOD-to-DAOD(if lb not complete in merged AOD)
  - Rucio internal
    - accounting - AP/GP/T0/XP, group ownership, campaign!=scope/project(?)
      - campaign in dq2 but never used
    - data placement – transient(boolean)

# Lumiblock

- NUMBER(10)
  - Single value can be set for each data file
    - not set for most files: e.g.MC
    - not set for data files containing multiple LBs
      - i.e. no lists or ranges
      - could not think of Prodsys use-case for this
      - if all lumiblocks in a file are complete then prodsys does not need to know the numbers for merging

# Rucio dev considerations

- Extended metadata stored in separate table
  - size,checksum in main table. Guid is already in extra table
    - do we need the guid for anything? Back nav, picking? PFC building?
  - extra table not optimized for searching on metadata
- Prodsys pulls all metadata with getFileInDataset request
  - not a search. Once per task.  $O(100)$  prod per day.
  - most may not need extended metadata, e.g. user inputs?
    - 2-stage get info possible, but already get guid from extra table
- Any search for provenance, group?
  - Maybe for accounting, quota enforcement
    - give me all Higgs datasets at site X
    - get space used by campaign mc14
  - would be at dataset level,  $O(1000)$  fewer lines
    - ok when not frequent?

# Store metadata

- JEDI registers output files
  - currently guid,size,checksum
  - just add lb
  - Should be atomic operation
    - cannot have file without required metadata
- JEDI registers datasets
  - just add provenance, group, transient
- Rucio client changes for this ready
  - awaits Db & server changes
  - expect JEDI will be ready to put and use new metadata

# Miscellaneous

- Comment in doc
  - “Files inherit dataset metadata”
    - not sure what this means
    - knowing a file is owned by group, is not useful
- Timescale
  - good to test lb merge in repro
    - move a RAW dataset to Rucio?