

# DDM shortlist

# The big issue

- Problems with cross boundary queries
  - Example: Amount of AOD data for the last may reprocessing campaign for the group e-gamma ?
    - In BNL or dataset definitions ?
  - Information shared between DDM and AMI
    - AMI AOD datasets != DDM \*AOD\* datasets, e.g. hidden datasets

# Naïve solution

- Restructure DDM to be meta-data aware
  - Split dataset name into several indexed columns
  - Add more access patterns
  - Meta-data columns
    - Dataset level
    - Static
    - More metadata in the dataset names...

# Possible solution

- Query AMI API with the meta-data attributes
- Get a list of dataset identifiers
- Query DDM API with this list of identifiers to get a report
- 😊 Same list of datasets between AMI and DDM
- 😊 physics metadata to dataset identifiers resolution
- 😞 Scalability of the approach
- 😞 incremental report ?
- 😞 Not aggregation of physic file information, e.g. nb events
- 😞 It doesn't solve the problems of user querying DDM by meta-data in the dataset names

# Better integration

- Web service @cern which queries both databases
  - Query AMI db with the meta-data attributes and do a join against DDM tables
  - Integration with the DDM accounting framework
- New APIs can be part of the DDM clients ?
- Do we need all datasets in AMI? Users?
- Need to figure out the tech. details
- Needs to be a DDM+AMI cross project