

Some issues for followup: I

Transform and metadata transport issues

- Transform architecture and its support for metadata
- Architecture for writing/reading metadata at the job (and Athena) level
- Metadata packaging and transport
 - Also improved consistency between T0 and distributed production
- Would be nice to be able to configure tasks/jobs from a dataset-level metadata source rather than by peeking

General

- Metadata validation as part of validation process for production
- Standard and optional metadata content



Some issues for followup: II

AMI-related

- Improvements to DDM/production/AMI agreement, understanding, and representation of dataset status
 - (Is the dataset ready, and what does that mean?)
- Detection and recording in metadata of unexpected event count differences and other production anomalies
- AMI provenance improvement
 - Not always clear or complete; particularly problematic at the file level
 - Which lumi blocks are in which file?
 - Beyond input/output provenance, most other provenance questions are hard to answer
 - Though configuration tags are a big help
- File naming differences between Tier 0 and distributed production?
 - Metadata in the file names is different (maybe this matters only to those who try to use DDM system to find data without AMI?)



Some issues for followup: III

Simulation metadata (defer to Borut here)

- Input of simulation metadata seems fragile, and reliant upon human input that may or may not happen
- Recurring issues related to use of run number in simulation
 - Affects metadata of various kinds adversely

Offline “quality” and similar metadata

- Stream-dependent offline “quality” flags?
 - More generally, a means to flag files that are problematic or interesting or anomalous in some way

Robustness improvements to in-file metadata handling

Metadata about the content of standard data products

- Can we answer questions like, does the AOD for period E2 contain AntiKt6H1TopoJets?
 - What is the EDM content of a given dataset?
 - Some kinds of queries/provenance questions are difficult without reading (running?) all the nested job options that went into a job’s definition

