


Commentary on Butler Data Model

These are some comments I have on what is missing from the current butler data model:

- ObsCore has the concept of observation identifier as a string. That is calculated by the translator but exposure here is defined as an integer. Should we have both OBSID and integer exposure ID?
- What exactly do we mean by visit now? Current model has that has an integer. Visits are disappearing from the acquisition system in that each exposure is given a unique OBSID string and can form a unique integer (YYYYMMDDNNNNN) but "visit" takes the form of a GROUPID "group identifier" which uniquely identifies the exposures that were taken together by the observing script (either unique per instance of the script, or in some special cases for calibration observations, unique for certain loops within the script). Are we required to translate GROUPID to an integer? It has also been noted that when processing, it may be necessary to define dynamic groupings. GROUPID is a concept that can therefore be applicable for calibrations, in fact we are planning to use it extensively for calibrations.

 DM-15536 - Jira project doesn't exist or you don't have permission to view it.

might be relevant.

- Are we trying to support in the model the concept of day YYYYMMDD and sequence number NNN? That's not a completely portable concept.
- We do declare that a detector has a "raft" although many instruments prefer not to use the term. Will people be upset if we retain raft (translator uses "detector_group").
- Subaru combines the detector name with the detector group but there are HSC use cases where raft is queried (the new YAMLCamera explicitly defines raft). Do we continue with defining a raft for HSC in gen3 (we currently do not) and a name that includes the raft? Or do we split them up? DECam has grouping of "N" and "S" but combines them as "N12" rather than "N_12" so how do we handle that? The translator recently introduced detector_name and detector_unique_name so you can decide which to use for ingest.
- We need to put observing mode somewhere (OBSTYPE aka IMGTYPE). Raw files can be bias, flat etc.
- What do we do with test stand TESTTYPE concept?
- Each field observed by the scheduler will have a name so we should use it. Normally "OBJECT" header and called "target" in ObsCore.
- snap is in the model but should be removed.
- We are missing the concept of observing program so how would we tell the difference between a deep drilling field or a fast/wide/deep field? Would we be parsing target names and guessing? This is sometimes called "run" in gen 2. How can we tell that some observations were taken in engineering time?
- What do we do about data quality? What happens if an exposure is good enough to be processed to a PVI but is not good enough to be combined into a coadd? Where do we keep that information? If it's a separate data quality table somewhere in the data backbone does the butler get to see it? (at my previous telescope we had ratings of good, bad, questionable, junk where questionable was meant to be a transient state before deciding whether it was good or bad, and junk meant there was something fundamentally wrong with it that would cause it to be unusable for anything).