

Ad-hoc data model working group 2018-09-20 meeting notes

Date

20 Sep 2018

Attendees

- [Gregory Dubois-Felsmann](#)
- [Colin Slater](#)
- [Kian-Tat Lim](#)
- [Jim Bosch](#)
- [Tim Jenness](#)
- [Fritz Mueller](#)

Goals

- Advance the plans for the LSST image metadata model and its alignment with CAOM2 and ObsCore

Discussion items

Time	Item	Who	Notes
	Scope clarification	Colin Slater , Gregory Dubois-Felsmann	<ul style="list-style-type: none">• Clarify the two senses of "metadata": per-observation ancillary data about images ("image metadata") and static (e.g., per-data-release) data about the data model ("schema metadata" or "table and column metadata"). Both need discussion but...• This initial meeting is about the former.• Tim Jenness: Can we talk about actually <i>using</i> ObsCore and CAOM2 ourselves and not just about mapping our data model to these standards on output?<ul style="list-style-type: none">◦ Gregory Dubois-Felsmann has been working on mapping to the relatively weak ObsCore data model. It may not be worth attempting to use it directly, though we should be sure that we can satisfy it on output. But CAOM2 has a lot of structure we could probably just use.
	Tour of CAOM2 and rough mapping to LSST	Gregory Dubois-Felsmann as tour guide, input from all	<ul style="list-style-type: none">• CAOM2 UML diagrams and detailed description: http://www.opencadc.org/caom2/• Reviewed the whole model superficially. Note that it has a lot of complexity from having to handle collections of data from a wide variety of observatories. Some of this can be dummied out for us. This can be done by actually instantiating the dummy objects or by faking them when we serve data. Should be a topic of discussion <i>after</i> we've understood the model together.• Discussed the "Observation" and "Plane" layers of the CAOM2 object model in some detail.• Results appended below (for easier formatting). <p>Next steps</p> <ul style="list-style-type: none">• Jim Bosch felt that he had enough information from the "tour" to be able to start to look more concretely at how to map these concepts onto the Gen3 Butler database concepts, and which of the CAOM2 elements could be used directly, rather than extracted via a translator.

Initial conclusions

Observation layer

- We propose using "SimpleObservation" only for raw images.
- We propose using "CompositeObservation" both for standard visits (comprising two raw images) and for alternate standard visits (containing only a single raw image). This retains the run-time flexibility that current requirements mandate, as well as maintaining a single data type for visits. It also better supports the notion of the ability to (re)define visits after the fact based on the archive of raw images taken, potentially even in an N:M way.
- We propose that an "Observation" from the main Camera represent the whole of a time-synchronized data acquisition from the Camera. In simpler terms, a main Camera Observation covers the entire focal plane. Substructure at the CCD level will be represented by lower layers of the CAOM2 model.
- CompositeObservation will be used for coadds as well. We only briefly touched on this, but it seems to make sense to define CompositeObservation at the "tract" level, with the "patch" structure at a lower level.
- CompositeObservation will also be used for calibration data such as synthetic dome flats, assembled from a set of individual SimpleObservations for each input flat. We contemplated, but decided against, making a full set of, e.g., 10 raw flats a SimpleObservation with multiple artifacts.

Observation metadata

- Very little metadata is absolutely mandatory at the Observation layer, in fact, just `Observation.collection` (string) and `Observation.observationID` (string).
 - "observationID" will correspond, for raw images, to the image ID field (not including CCD) discussed under [LCR-1424](#). We did not discuss its value for visits, nor is there a fully satisfactory definition of visit IDs to rest on at this time (i.e., this is a wider problem).
 - We did not discuss the other attributes of "Observation" at this time. "metaRelease", the metadata-release timestamp, should be discussed to determine whether it should match the release of alerts (c.f. OTT1) or the remainder of the Prompt data (c.f. L1PublicT).
- Apparently there is also a mandatory "Algorithm" object with a mandatory "name" string value, though its semantics are weakly defined. *Note added after the meeting:* It appears that it is expected to have the value "exposure" for SimpleObservation objects, and a value describing the composition process for CompositeObservations. I propose the values "StandardVisit", "AlternateStandardVisit", and "SyntheticCalibration" for the cases we described; other values may be needed for Special Programs observing.
- We did not discuss the "Telescope" metadata object, but for those reading along, it is worth pointing out that although the *explicitly* specified attributes look like they are survey constants (the name and geographic coordinates of the observatory) the documentation states that the additional "keywords" may "describe the telescope or telescope configuration at the time of observation".
- Similarly, the "Instrument" metadata object "keywords" may "describe the instrument or instrument configuration at the time of observation".
- We briefly discussed "Target" and "TargetPosition". We agreed that a "TargetPosition" should be supplied for science observations, with "coordinates" describing the nominal pointing of the telescope on the central boresight axis. An additional rotation-angle attribute should be added. A "Target" may be supplied if the Scheduler operates in terms of a grid of fixed fields plus dithers, and could define the fixed field. (Note added during minutes transcription: it may be reasonable to use the "Target.name" to refer to the dome screen and/or CBP, as appropriate, for calibration observations.)
- We did not discuss the other Observation metadata objects in any detail yet.
- We find the Plane metadata object "Energy" may also be useful at the Observation layer for the main camera, as the filter and its bandpass are known independently of the image data reduction process. We would like to make it available directly from the Observation in addition to its normal association with a Plane. See the discussion under "Plane metadata" below.

Plane layer

- The Plane layer should be used to refer to the state of data reduction of an image.
 - "dataProductType" should be "IMAGE" for all image data, raw, calibrated, synthetic calibration, coadds, etc.
 - CAOM2 (just as ObsCore) allows for reduced data to be treated as associated with Observations as well. So although in the final released data we do not plan to explicitly release per-Observation artifacts such as FITS files of Sources detected on a particular image, this is a very relevant concept internally and well-defined in the Butler, so we should support this. In this case CAOM2 allows for the use either of the ObsCore standard "dataProductType" = "MEASUREMENTS" or an additional non-ObsCore value "CATALOG".
 - "calibrationLevel" is an integer value taken from ObsCore; we propose the use of the value 1 (RAW_STANDARD) for raw data (since our raw data is available in a community-standard format), 2 (CALIBRATED) for calibrated single-visit images, 3 (PRODUCT) for coadded images, and 4 (ANALYSIS_PRODUCT) for Observation-linked catalog data artifacts such as FITS tables. See [ObsCore](#) section 3.3.2 for more background.
 - (Note added in transcription of minutes) [Gregory Dubois-Felsmann](#) proposes the use of ObsCore's "dataprodukt_subtype" to represent Butler dataset types, with an "lsst:" prefix, e.g., "lsst:calexp". This attribute could also be included here at Plane level, or at a lower level.
- Note that a single Plane for single-epoch data still represents the entire focal plane.
- A standard (two-snap) visit would then be modeled in this way:
 - The visit is a "CompositeObservation" with two "members" pointers to the "SimpleObservation" objects for the snaps.
 - The snaps would have only a Plane with calibrationLevel=1, representing the raw image data.
 - The visit would not have a calibrationLevel=1 plane, but rather a calibrationLevel=2 plane for the PVI data from the visit.
- The "TargetPosition" and "Energy" objects for the visit and the snaps should all be the same, and, *pace* the use of the ownership-implying "black diamond" UML relation in the CAOM2 model, should probably share a single actual physical object.
- We did not discuss how to represent the multiple PVIs that would emerge, over time, from a single visit: the Prompt (Level 1) PVI, and the per-data-release PVIs. One possible representation of this would be to make new CompositeObservations for each one, with the Prompt-DR1-DRn axis represented in the "collection" attribute of the Observation. This may be the most natural way and the way which most clearly expresses the idea that at least in principle visits are defined at processing time, not once-for-all-time (though in practice these definitions are likely to be very stable). Another would be to create multiple Planes of the same visit's CompositeObservation. In this case they would have to be distinguished on the "productID". (In the presumably rare cases where certain visit definitions *did* change over time, this would be accommodated with new CompositeObservations.) [Gregory Dubois-Felsmann](#): I don't think that representing the reprocessing of visits *below* the Plane layer is compatible with the intent of CAOM2, as Provenance is at Plane level.

Plane metadata

- The only *mandatory* Plane-level metadata is the "productID", "collection- and observationID-specific identifier for this product", i.e., something defining the role of a particular Plane within the set of data associated with an Observation. (The productID can then be used to create the "creatorID", "typically made up of the Observation.collection, Observation.observationID, and Plane.productID and in the form of an IVOA dataset identifier".) We did not discuss this at the initial meeting.
- "calibrationLevel" and "dataProductType" were discussed above.
- We did not discuss the "metaRelease" and "dataRelease" time stamps.
- We did not discuss the DataQuality or Metrics metadata objects, nor did we discuss Provenance beyond noting its existence.
- The Polarization metadata object will not appear in the LSST data model.
- The Energy metadata object will be derived from largely static filter information. It must represent both the nominal filter bandpass name and the physical serial number of the filter. The "bounds" Interval will represent the bandpass of the filter at some nominal transmission level; we did not discuss whether it would be an entirely static, measured-once mapping, or updated from time to time with information from the detailed filter calibration.
- The Time metadata object will be derived from shutter metadata. Note that this Time object represents a visit, and for standard visits cannot be identical to the Time objects of the associated raw data Observations' Planes. It is unresolved whether for the alternate standard visit the Time object might be refined from the raw Observation/Plane to the visit/PVI Observation/Plane (e.g., using the detailed shutter transit metadata as opposed to only the basic start and stop OCS events).
- The Position metadata object was not discussed in enough detail to determine what to do with the "bounds".

Artifact/Part/Chunk

We did not discuss this level of the model in any detail beyond noting that only the Chunk level appears to have detailed WCS metadata objects. [Gregory Dubois-Felsmann](#) promised to provide, after reminding himself, an explanation of the intended use of these three levels, which for many projects are not all meaningfully distinct. It was clear that somewhere in these layers, though, the breaking-apart of a focal-plane notion of a visit into CCD-sized pieces would occur, and similarly for coadd tracts and patches.

Action items

- ☐ [Gregory Dubois-Felsmann](#) Send around an explanation of the intent of the Artifact, Part, and Chunk layers of the CAOM2 model. **27 Sep 2018**
- ☒ Jim Bosch Use what was learned in the initial CAOM2 meeting to come up with an initial, more detailed, proposal for a mapping between CAOM2 and the Gen3 Butler database, and for where CAOM2 types could be used directly in that data model. 03 Oct 2018