NRC·CNRC

WD-CAOM-2.5

Patrick Dowler
Canadian Astronomy Data Centre

IVOA InterOp Nov 2024

Overview

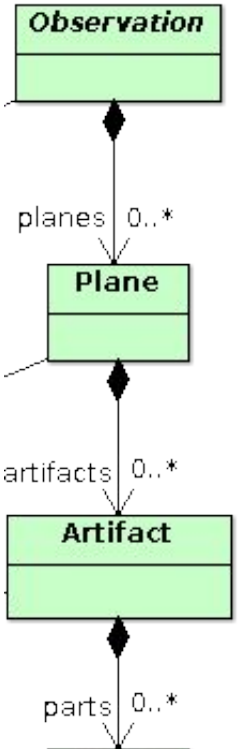
- What is CAOM?
- Current operational status
- CAOM vs current IVOA standards
- Updates (draft) since CAOM-2.4
- Identify key areas for discussion 

Common Archive Observation Model

- common metadata model to describe all science data
- design goals:
  - metadata only (no data formats)
  - static collections and data flowing from telescope(s)
  - public and proprietary metadata & data
  - support data discovery services
  - support data access services
  - support operational activities: metadata curation, sharing/mirroring

  - **describe the data we actually have**
  - **evolve to meet to new data challenges**
  - bring valuable old data forward (~free) as systems evolve

Common Archive Observation Model

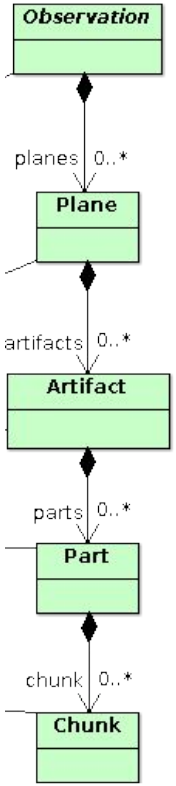


- Observation: result of a single experiment or process
  - SimpleObservation: ~single exposure by a telescope/instrument
  - DerivedObservation: an observation made from 1+ other observations, usually combined
- Plane: products of the observation
  - described by position / energy / time / polarization coverage
  - differ by processing applied: calibration level, provenance, sometimes data product type, etc.
- Artifact: stored resources of the plane
  - usually a file e.g. science data, previews, auxiliary data

Current Operational Status

- CAOM-2.4 released late 2019 and stable
  - some minor bug fixes and improvements
  - core java lib is 2.4.9
  - core python lib is 2.6.1
- CADC: supporting 40 collections (~30 static, ~10 live), ~60 million observations
  - ~5 with proprietary data, ~2 with proprietary metadata
  - pushing the limits of usefully describing radio data: ALMA, DRAO, LOTSS, POSSUM, VLASS, WALLABY, etc
- CADC and ESAC mirror HST, JWST, TESS metadata from MAST
- current CAOM community: CADC, ESAC, IRSA, IPAC, MAST, Rubin/LSST, SKA, TMT

CAOM and IVOA Standards



data discovery

TAP
SIA / DAP
Vocabularies

ObsCore (+radio)
Characterisation
Coords
Measurements
Provenance
DALI datatypes

reference to storage

DataLink

plane metadata generation
cutout computation

SODA

Cube

general processing & analysis

Draft CAOM changes – identifiers

- replace Observation.observationID (string) with Observation.uri (URI)
  - used directly in DerivedObservation.members
- replace.Plane.productID (string) with Plane.uri (URI)
  - used directly in Plane.provenance.inputs
- remove Plane.creatorID (URI) because it is now redundant
- Plane.publisherID exists as a concept but it not strictly part of the model 

Draft CAOM changes – vocabularies

- change Plane.dataProductType to explicitly refer to the ObsCore-inspired **product-type** vocabulary
  - add Plane.dataProductSubType vs narrower terms in vocabulary??
- change Artifact.productType to explicitly refer to the **DataLink-core** (semantics) vocabulary
- change Plane.observable.ucd to explicitly refer to the **UCD1+**, would prefer to treat it like a vocabulary
- add Plane.[observable,position,energy,time].calibration as a **calibration** vocabulary (non-existent: along the lines of ObsCore optional appendix material) 
- TODO: explore/use VO-DML semantic concept

Draft CAOM changes – radio

- reconcile with and support ObsCore Radio extension concepts
  - add Plane.position.minBounds (shape)
  - add Plane.position.maxAngularScale (interval)
  - add Plane.energy.resolution (real)
  - add Plane.energy.resolutionBounds (interval)
  - add Plane.uv.distance (interval)
  - add Plane.uv.distributionEccentricity (real)
  - add Plane.uv.distributionFill (real)
  - add Observation.telescope.trackingMode (vocabulary)
  - several of these have use/meaning outside radio… ObsCore-1.2?
- change Plane.polarization.states from enum to a (non-existent) vocabulary

Draft CAOM changes – data types

- removed SampledInterval in favour of separate Interval and Interval[ ] fields (energy,time,custom axes)

- remove MultiPolygon
- separate Shape and MultiShape fields in Plane.position

Draft CAOM changes – DataLink support

- add ArtifactDescription with ID and description
- add Artifact.descriptionID to reference it using the ID

- at scale, many millions of Artifact(s)...
- … but expect only thousands of ArtifactDescription(s)

- ArtifactDescription is an entity, so is independently curated
  - can be synced when sharing metadata between institutions

Draft CAOM changes – TODO

- analyze usage of all existing fields
  - identify rarely used and unused fields

- since CAOM-2.4, Plane.custom axis
  - use case: cubes with Rotation Measure axis
  - use case: cubes with Faraday Depth axis
  - pattern is unproven at this point, but need is there
  - axis type (ctype) as a vocabulary? 

If/when CAOM DM becomes an IVOA standard

- expect a separate CAOM-TAP relational model
- expect a separate CAOM web service (curation and sync API)

Let’s discuss!!