MAST Users Group Meeting

Nov 26 2012

Common Archive Observation Model

What is CAOM

and why should MAST use it?

Brian McLean

MAST Users Group Meeting

Nov 26 2012

Roadmap for the Space Telescope Science Institute Archive and Data Center

Strategic Goals

- Increase the science research productivity of the astronomical community through the products and services we provide.
- Enable exemplary science through the development of science products and services.
- Maximize the long-term archival benefit of the data holdings within the STScI Archive and Data Center.
- Actively collaborate with other archive centers to increase the overall productivity and science return to the research community.
- Promote and engage the user community on the value and use of the STScI data holdings.

MAST Users Group Meeting

Nov 26 2012

Public Science Multi-Mission Archive Advantages

- Primary point of entry for access to browse and access SCIENCE data for ALL missions housed at STScI.
- Browser-based search and visualization tools.
  - All data is available on disk for immediate access
  - proprietary rights can be enforced
- Identify related datasets from different missions to enable multi-wavelength science analysis.
  - Including community contributed or mission generated high level science products.
- Virtual Observatory compliant data access.
  - Users can easily discover data, retrieve and analyze using VO tools.
  - Inter-archive access

Improves scientific return and impact for all missions

Common Archive Observation Model

- Unified model for metadata for all archives
  - Enable cross-mission searches and enhance scientific value
  - Simplify access with common set of software
  - Model is extensible for mission specific metadata if needed
- Layered software
  - Use VO protocols for all services (internal & external)
  - Write access services once, use for all missions
  - Lower maintenance costs
- CADC developed and implemented CAOM v1
  - Collaborate with CADC to implement/extend model
  - Use CADC lessons learned and implement v2

MAST Users Group Meeting

Nov 26 2012

Common Archive Observation Model v2.0 Summary of main concepts

- OBSERVATION
  - Metadata describing the characteristics of an observation
    - *simpleObservation : describes a single set of observed photons*
    - *compositeObservation : describes the combination of 2 or more simpleObservations*
  - Proposal, Telescope, Instrument, Target, Proposal, Combination type, etc
- PLANE (1:n planes per observation)
  - Metadata describing a distinct construct within the observation
    - e.g. multiple detectors or filters would each be a plane
  - Position (Footprint), Energy, Time, Polarization, Metrics, Provenance
- ARTIFACT (0:n artifacts per plane)
  - Metadata describing a distinct product file within each plane
    - e.g, image/spectrum , errors, weights, previews
  - Includes pointer (URI) to actual data files
- PART/CHUNK(0:n per artifact)
  - Metadata describing each unique product within file (e.g. multi-extension FITS)
  - SpatialWCS, SpectralWCS, TemporalWCS, PolarizationWCS

- Queries can access any of these layers
  - Can create views for VO defined data models e.g ObsCore and protocols e.g. ConeSearch, SIAP, SSAP

Current Status

- Leave mission databases unchanged and only create combined CAOM database for science queries
- Created CAOM database to support Portal development
- Created spatial functions, STC/S conversion functions
- Database loading
  A) Scripts to copy data from mission db to CAOM db
  B) IDL tool to harvest data from headers and/or mission db and load CAOM db

- Missions done : GALEX GR6, HLA images, IUE low dispersion spectra, FUSE
- In progress : IUE high dispersion, HLA spectra, HST, SWIFT-UVOT, KEPLER FFI

Current Plans

- Load Additional Missions
  - High Priority : KEPLER, GALEX GR7, HST High Level Science Products
  - Medium Priority : SAGE, XMM-OM, HST PR images
  - Low Priority : EUVE, HUT, UIT, WUPPE, HPOL, BEFS, TUES, IMAPS, COPERNICUS, FIRST
- Implement updated CAOM schema developed with CADC
  - Includes additional metadata about data provenance, processing, access control, statistics
- Build more sophisticated queries including temporal and spectral coverage