

ALMA Archive/VO meeting 9th  
November 2010 at ADASS XX,  
Boston, MA

## Present:

Mark Lacy, NRAO

Brian Kent, NRAO

Bob Hanisch, VAO, STScI

Doug Tody, NRAO

Masatoshi Ohishi, NAOJ

Severin Gaudet, NRC

Patrick Dower, NRC

Stephane Leon, ALMA (JAO)

Felix Stoehr, ESO

Juande Santander (ESO, via skype)

## Meeting summary:

(Immediate action items in bold)

Mark gave background to the current situation. At the previous ADASS meeting, Doug Tody, Brian Glendenning, Alberto Michol and Andreas Wicenic met and outlined a high level description of how the ALMA archive would interface with the VO, which Doug wrote up. In the meantime, ALMA has begun taking data, but the current archive design, based on an XML database store, although adequate for running the telescope, is not suitable for use as a science archive for external users to query.

Thus, the ALMA archive development team has decided that the ALMA Science Archive (ASA) will use a relational database, populated by harvesting key metadata from both the XML database (which will continue to be used for telescope operations as the "ALMA front-end archive" (AFA)) and the ASDM files that store the raw visibility data. Views will be constructed onto the ASA database that will support the VO TAP and SIAPv2 protocols.

**Felix will lead the task of defining the metadata for the ASA archive to be extracted using the harvester into the ASA data model, assisted by Mark and Stephane.** We will liaise with the EVLA archive through John Benson, who are performing a similar metadata harvesting from the ASDM for EVLA data. We will then discuss in detail with Doug and other interested parties in the VAO and wider VO community whether the metadata are sufficient for forming the VO-specific views. The ALMA archive scientists can help to determine the SIAPv2 requirements through interaction with the

VO organizations, and through commenting on the SAIPv2 white paper which is currently being drafted, and which should be available for comments early next year. Supporting ALMA is a high priority for the VO.

The pipeline plays a key role in generating the metadata for the processed products, we will liaise with the pipeline team to ensure appropriate metadata is generated by the pipeline. The pipeline also needs to generate a processing history that is stored with, or as part of (e.g. in the FITS header), the pipeline product to ensure data provenance.

Reprocessing at the ARCs can leverage VO infrastructure for providing web interfaces to pipelines to supply parameters and queue pipeline jobs etc.

VO access is required to both raw and pipeline processed data. For raw data, we need to define a footprint. The FOV parameter in ObsCore should be able to do this, perhaps supplemented by a rectangle or polygon for mosaicked data within a single ASDM. The Common Archive Object Model, developed by the CADC, will include such an option.

Data access: the ALMA archive allows for anonymous querying via a VO user, but anonymous data downloads are currently not allowed. This needs a policy change. As lead of the archive policy WG, **ML will write this into the archive policy document** for consideration by the directors. Support from VAO through Bob Hanisch and through NAOJ by Masatoshi Ohishi (former IVOA chair) would be forthcoming if obtaining agreement to this from the ALMA project is problematic. For proprietary data, the VO supports authentication, allowing access to be restricted during the proprietary period. Metadata should be open access, if only for ease of implementation.