

ALMA Data: Access and Processing

This is an overview of how the AAT/PPI handles finding and using ALMA data.

Data Hunting:

Rawdata Discovery:

AAT/PPI selects the set of data with PASS and SEMIPASS status for QA0, using a combination of data from the tables: AQUA_STATUS_HISTORY, AQUA_V_EXECBLOCK, PIV_SCHED_BLOCK. The current query is a modification of the one used by the NAASC MAGMA project to provide information about raw data to PIs when requested.

The amygdala system performs a periodic check (set via CAPO, currently hourly), and checks for all updates to the status of EBs for the previous number of hours (currently 12, configurable via CAPO). These updates allow for a broader selection of raw data for ALMA than the initial methodology implemented.

Additional Backfill:

alma_butler does a broader comparison of known EBs for ALMA and the AAT/PPI (with out a timescale attached) with the appropriate status, and adds the missing data sets to the list of what to ingest. This tool is run daily via cron. This utility still finds some gaps in the data, even with the recent set of updates. This indicates that the variability of QA0 status values is not yet well understood.

Calibration Discovery:

The AAT/PPI selects MOUSes which had products recently ingested for delivery. AAT/PPI gathers date information from ASA_DELIVERY_STATUS (similar to the NAASC LAVA project), and then the ingestion tool gathers information from the ASA_PRODUCT_FILES, ASA_SCIENCE, and ASA_ENERGY tables for use in restores or AUDI purposes. This reliance upon ASA_* tables constitutes a particular risk, as will be discussed below.

The amygdala system performs a periodic check (set via CAPO, currently daily) for new calibration information. Like the EB search, we look for updates over the last number of days (currently 2, configurable via CAPO).

Limitations Upon Calibrations:

Once potential MOUSes are identified from the ASA_DELIVERY_STATUS, the contents of the products are considered:

- MOUSes with an ARI-L project readme file is rejected.
- MOUSes with additional scripts to facilitate calibration (*scriptFor*Cal*) indicate that the calibration isn't suitable for the PPI and rejected.
 - Note: The conversion to Workspaces might be able to alleviate this restriction
- MOUSes without any images classified as for science are delayed until the next cycle and checked again.
- MOUSes without any rawdata recorded for them are delayed until the next cycle and checked again.
 - When EBs are ingested, the OUS structure is looked up and persisted to the AAT metadata database

Processing Considerations:

Session Mapping:

In addition to the basic information of EBs and calibration files, there's an additional layer of organizational information required for processing ALMA data via the CASA Pipeline. The organization of EBs into 'sessions' is required within the ProcessingIntents section of the PPR.

The best method discovered to date for re-creating the session structure within the MOUS involves collecting data from the tables: AQUA_EXECBLOCK, AQUA_SESSION, SHIFTLOG_ENTRIES, and BMMV_SCHEDBLOCK to generated a time-ordered list of EBs, with session number and status information. This is the data that allows for the creation of a reasonable set of ProcessingIntents entries (including creating blank sessions in the correct locations).

Runtime Data Gathering:

In order to mitigate the the effect of changes to the ALMA metadata database after it has been ingested by the AAT/PPI, the constituent components to facilitate a restore (raw data, calibration files, and session information) are all queried as part of the setup for performing the processing.

Caveats:

ASA_* tables are known to be of questionable reliability. Those tables are reportedly dropped and rebuilt on an irregular basis.

Unknown timescale, unknown changes, etc

Also intermittent failures of the processing infrastructure and occasional ingestion issues.

- [ALMA Optimized Imaging for the AAT/PPI](#)