Archive status and plans

Mark Lacy

ANASAC 16-17 Sept 2013

Current status

- Cycle 0 archive complete, software working.
  - Query tool allows search on observational metadata (position, frequency, resolution, date, etc), and project metadata (project code, PI etc).
- Cycle 0 data products are in tar files containing:
  - Reference images (FITS)
  - Calibrated measurement set
  - Measurement set with only Tsys and WVR corrections applied
  - Calibration tables
  - Scripts, logs, QA2 information
- Cycle 1 archive products will be different:
  - Tar files will not contain measurement sets.
 - Instead, raw ASDMs will be available through the request handler interface.

Future plans for the ALMA archive

- The archive user experience will change significantly for Cycle 2 and beyond, once the pipeline is commissioned
  - Pipeline products (images and ancillary data [PPRs, PSFs, Flux images, weblogs, caltables, previews]) will be ingested, and made available on a per-product basis (no more giant tar files).
  - One primary product FITS image/cube will be delivered per source per spectral window [large files, or single dish observations, may be split], with other files listed (and downloadable) as ancillary files.
  - Searches will be conducted on the ObsUnitSet level (the unit the pipeline uses for processing) by default (though observation-level ["raw"] and project-based views will remain available).
  - VO capabilities will be enabled for the product files via a database view, so ALMA data will be available via VO interfaces.

Advertisement of the archive

- A news item was placed on the ALMA Science Portal to announce the availability of the archive query tool in February (archive data was available prior to this on an temporary webpage).

- Items have appeared regularly in the NRAO E-news

- Would be interested in ANASAC ideas for further advertisement, particularly once "full" archive is available in Cycle-2+

- Archive usage statistics show a total of 4TB of anonymous downloads since February, indicating community awareness of the public data.
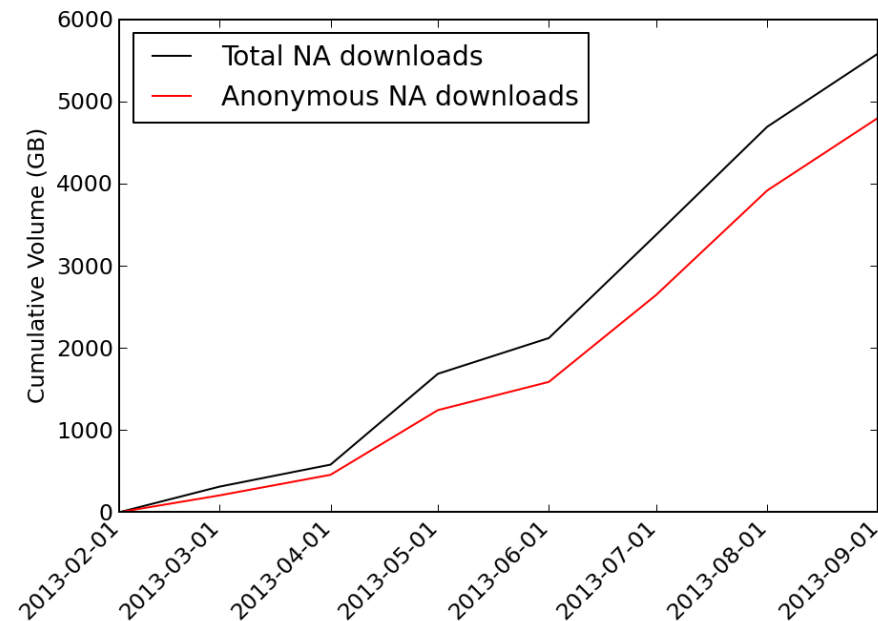
Enhanced processing capabilities

- NRAO is planning to build a pipeline processing interface which can be used to reprocess ALMA (or VLA, GBT) data on demand using a current version of the pipeline to produce new images (with, e.g. different uv weighting), or a calibrated measurement set (these capabilities will be extended to include tuning of further pipeline parameters later).

  – The software infrastructure we are building to support this will be the basis for server-side visualization and analysis of ALMA data products in the future.

  – The other ALMA ARCs are also interested in pushing forward along these lines, with collaboration to ensure similar user experiences for each ARC.

Proposed Stage 1 PPI options page for ALMA

Proposed Stage 1 PPI Cart page

ALMA development proposals

- Several ALMA development project proposals have been submitted for advanced analysis and visualization capabilities (two from NA), plus an NRAO-led one for building a portal to the NCSA facilities for data processing. If any of these are funded, they will allow us to accelerate development of support for data analysis:

  - E. Rosolowsky (Alberta): Improvements to visualization tools, including server-side visualization of large data cubes.
  - L. Mundy (UMD): ASTUTE – improved archive metadata from sophisticated image and spectral analysis tools.
  - J. Robnett (NRAO): – External on-demand computing for ALMA data via a science gateway to the XSEDE supercomputing environment.

Summary

- The ALMA archive will undergo significant development over the next 1-2 years.

- Key change to the ALMA archive will be the ingestion and availability of individual products from the pipeline (no more huge tar files!)

- Integrated into this will be the development of software infrastructure to support remote access to NRAO computing facilities (focusing initially on the Pipeline Processing Interface, but expandable to include other computing facilities [NCSA], analysis and visualization).