Cycle 7 NRAO ALMA Development Study Proposal – ALMA Archive Science Mining

PI: P. Teuben with M. Pound

Institutions: U. Md.

ABSTRACT

We propose a study to create a prototype Science Query Database that enables broad science-driven queries of ALMA projects. Our goal is to enable science discovery with ALMA archival data by enhancing users' ability to identify, access, and examine relevant data sets through database access to scientific and observational metadata. This study proposes a design and prototype implementation as a pathfinder for a full ALMA implementation.

We will design and construct a Science Query Database on an Amazon Web Services (AWS) testbed using selected public Cycle 5 data. We will image, as necessary, and run the ALMA Data Mining Toolkit (ADMIT) on full projects to create a standard set of science products, and ingest the ADMIT science metadata (e.g., line identifications, line characteristics, source intensities, image statistics, source coordinates) into the Science Query Database. We will merge these metadata with metadata harvested from the ALMA Science Archive system and the u,v and image data files. Combining these with the existing archive interface capability of searching project abstracts and science keywords will allow investigators to make queries that dig through the data rich archive to facilitate new science and explore new ideas.

We will create a new AstroQueryLite Python package to showcase how this implementation can be integrated into many user environments. We will use remote Jupyter notebooks for our study, which are familiar to many astronomers. The outcomes of this study will be: the design framework for including science metadata in future ALMA archive upgrades, a prototype implementation of a Science Query Database and associated access tools, and a test of the viability of AWS as an archival database server for public use.