Getting the most out of your ALMA DATA with ADMIT: The ALMA DATA Mining Toolkit
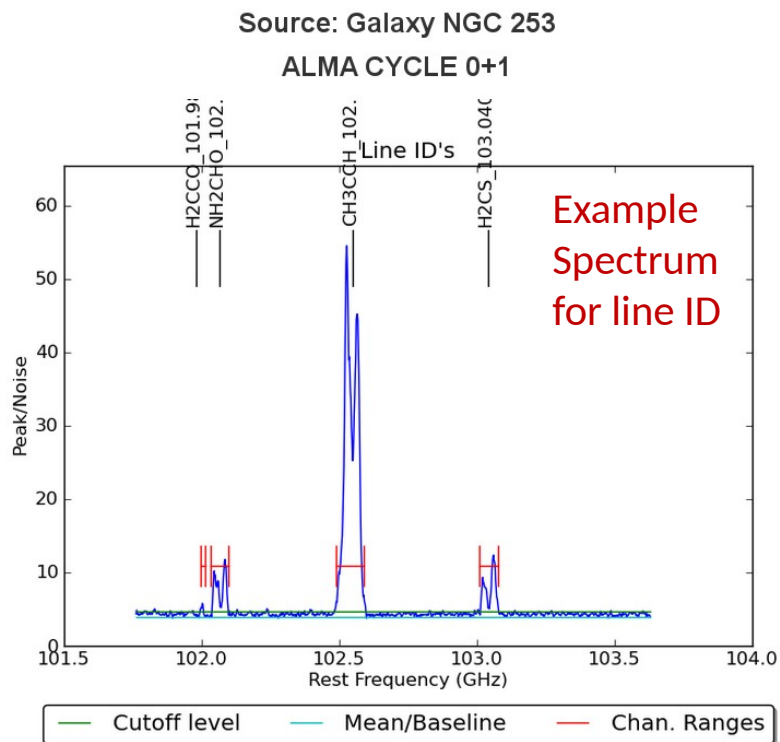
Lee Mundy, Leslie Looney, Peter Teuben, Doug Friedel, Marc Pound, Lisa Xu, Kevin Rauch, Robert Harris, Jeff Kern, & ADMIT team

ADMIT: ALMA Data Mining Toolkit

◆ Developed by University of Maryland, University of Illinois, and NRAO (PI: L. Mundy)
➢ **Goal #1:** First-view science data products into archive: spectra, line ID, moment maps, etc
➢ **Goal #2:** Python Toolkit allows user to generate their own science products from cubes.
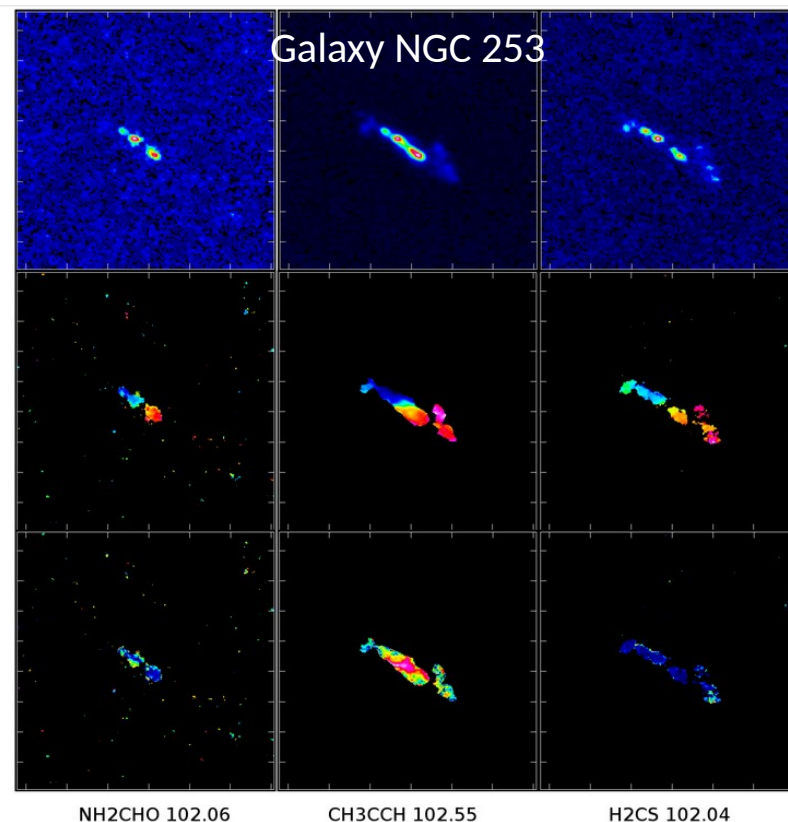
ADMIT Products shown from automated flow

Source: Galaxy NGC 253
ALMA CYCLE 0+1

Example Spectrum for line ID

Galaxy NGC 253

Moment 0

Moment 1

Moment 2

NH2CHO 102.06   CH3CCH 102.55   H2CS 102.04

Identification of common lines

Moment maps for lines

Identified lines

| frequency	formula	transition	velocity	fwhm	startchan	endchan
101.98143	H2CCO	5(1,4)-4(1,3)	2.938E+02	2.413E+01	828	836
102.06427	NH2CHO	5(1,5)-4(1,4)	2.382E+02	2.354E+02	783	816
102.54798	CH3CCHv=0	6(0)-5(0)	2.300E+02	2.343E+02	532	583
103.04055	H2CS	3(0,3)-2(0,2)	2.359E+02	2.332E+02	283	318

ADMIT

- ❖ Operates on data cubes only; cubes can be FITS or casa image format
- ❖ Compatible with CASA environment and utilizes CASA routines where possible
- ❖ Products are self-documenting with XML; compatible with future ingestion by a database

Each data cube get a full set of products
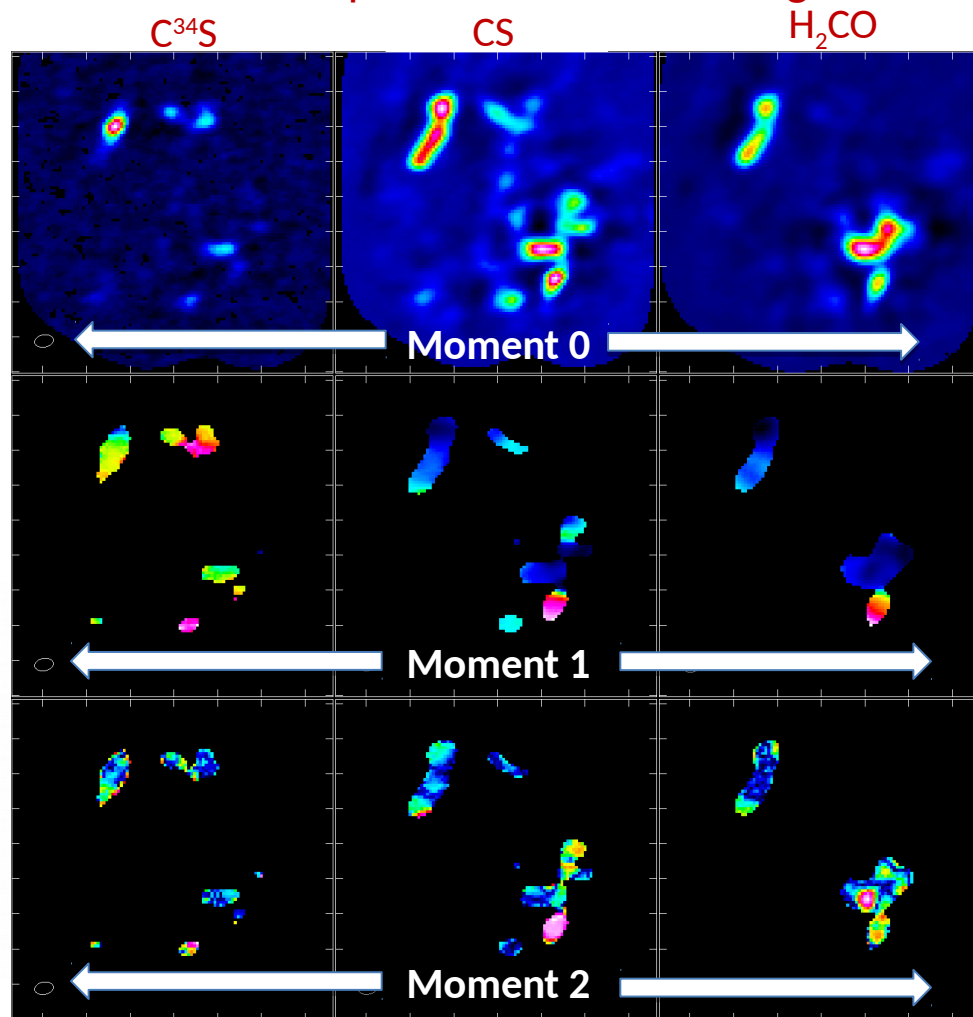See: admit.astro.umd.edu/admit-M4 and click on an xxx.admit directory
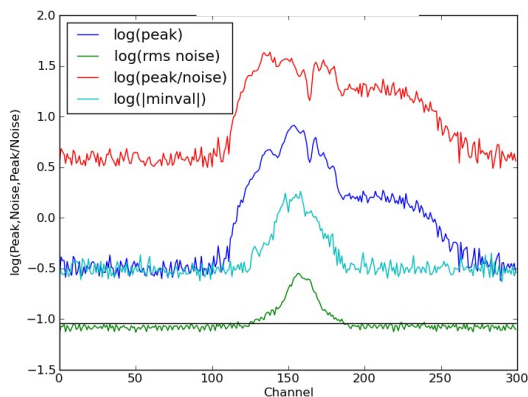
Spectra based on peak flux and noise in each channel

Blue and green spectra highlight impact of missing flux
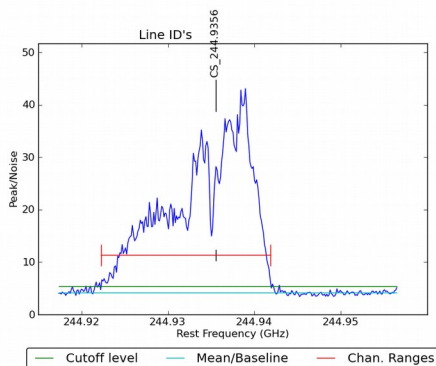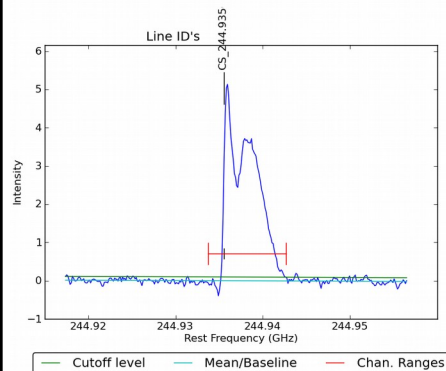
Two spectra used in line ID of CS J=5-4

Serpens Main Mosaic Image

C^{34}S   CS   H$_2$CO

Moment 0

Moment 1

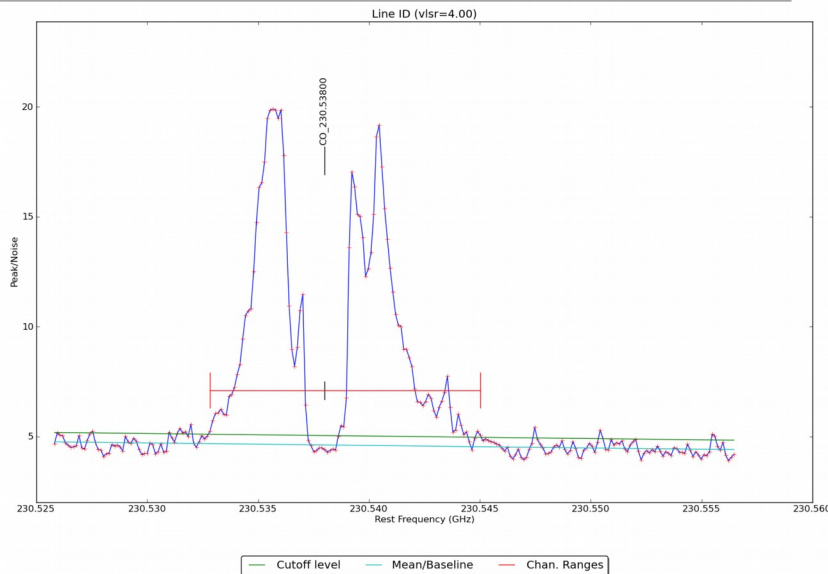Moment 2

ADMIT

Two modes of Operation:

- **On-line   (pipeline mode producing standard set of products):**
  - ADMIT runs after QA2 and before archive ingestion (ideally as a pre-step to the archive ingest process) – details being worked with ALMA Project
  - ALMA archive user can select to download ADMIT tarball (20-40MB)
    - XML, PNG, and HTML files; limited FITS files – details to be decided with ALMA Project
  - Browser-based view allows user to inspect products once downloaded

- **Off-line (user created data products):**
  - The ADMIT Toolkit "add-on" available from the CASA download page
  - Flow-model for creating and re-creating products - viewable in the browser
  - Environment for expanded exploration of data sets:
    - Principle component analysis of emission
    - Overlap integrals
    - Comparisons across multiple windows and multiple sources
   - New tools for examining large data cubes
    - Fine tune line ID

ADMIT

- ❖ Automated line ID allows line-based operations: moment maps, PV slices, etc
- ❖ Flow produces set products for users, which can be determine by ALMA
- ❖ Users can create their own custom products locally, which can be applied across sources

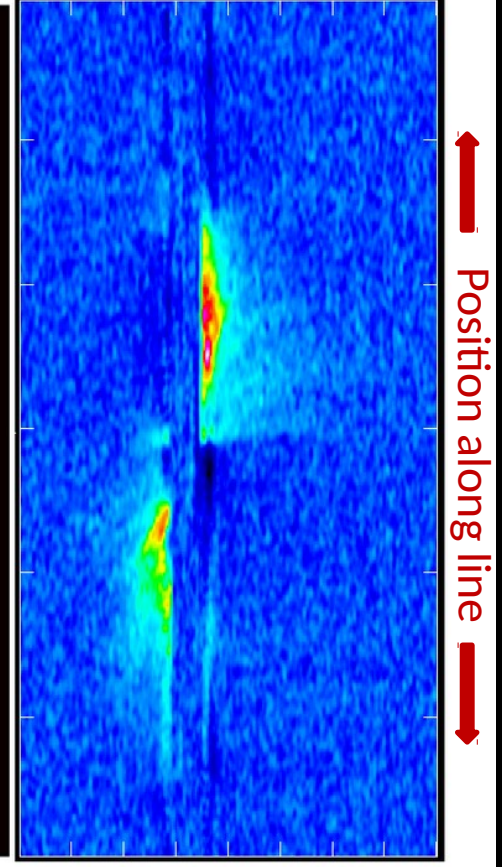Conservative automated line ID:
do no bad ID

Spectrum: peak emission over noise
per channel

Moment 0
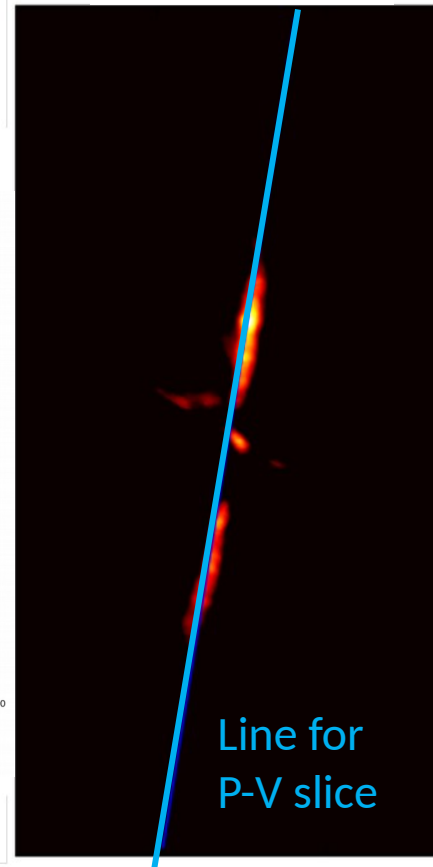
Position-Velocity Slice

Line for
P-V slice

Position along line

Velocity

Identified lines

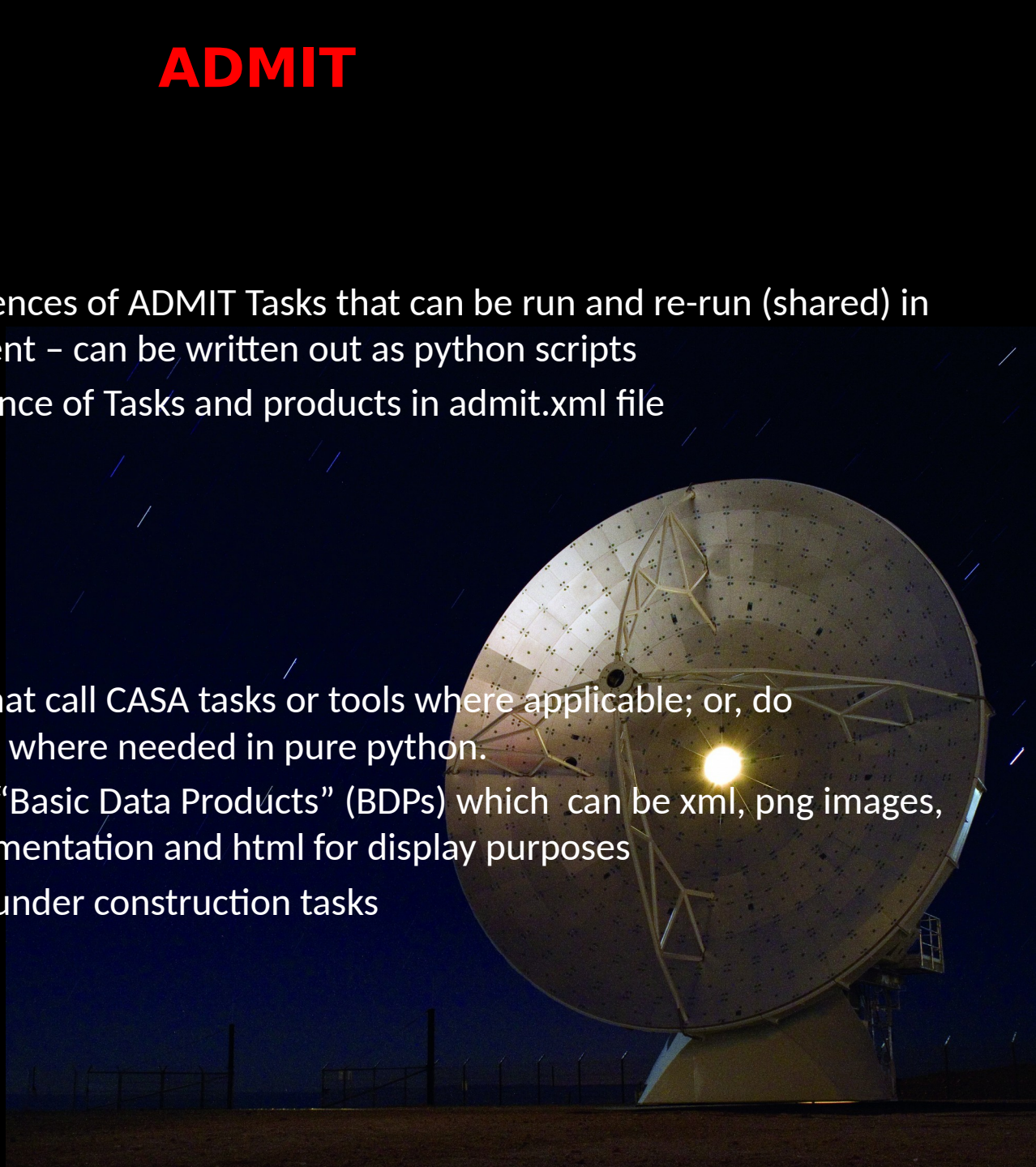| frequency	formula	transition	velocity	fwhm	startchan	endchan
230.53800	CO	2-1	1.200E+01	3.503E+00	100	155

ADMIT

Flow Manager

- Allows creation of sequences of ADMIT Tasks that can be run and re-run (shared) in CASA python environment – can be written out as python scripts

- Keeps a record of sequence of Tasks and products in admit.xml file

ADMIT Tasks (ATs):

- ATs are python scripts that call CASA tasks or tools where applicable; or, do appropriate calculations where needed in pure python.

- The output of tasks are "Basic Data Products" (BDPs) which can be xml, png images, and FITS files with documentation and html for display purposes

- 12 currently existing or under construction tasks

AT Flow

Ingest

CubeStats

CubeSpectrum

LineID

LineCube

Moment

Ingest

SFind2D

CubeSpectrum
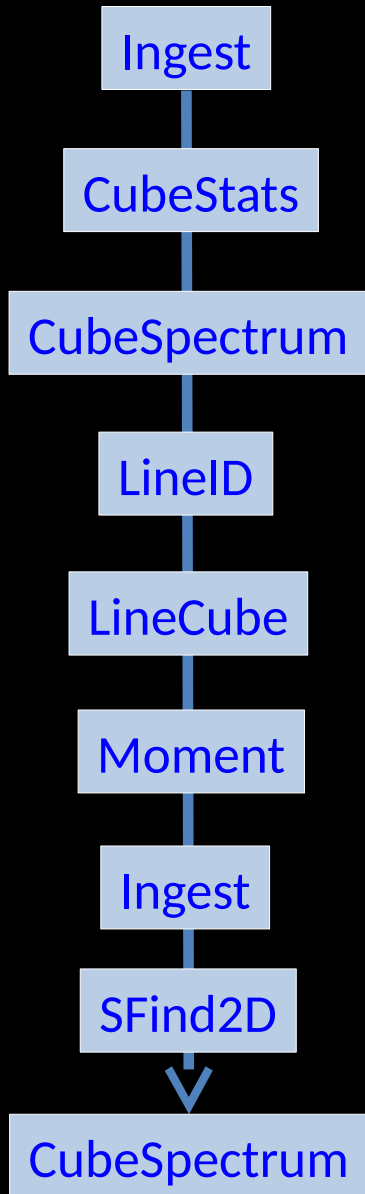
ADMIT

Read in full window data cube: from FITS or CASA image
        Output: CASA image and xml information

Calculates statistics of data cube: RMS, Min, Max per channel, etc
        Output: xml table, png's

Makes Spectra which characterize the emission –used for LineID
        Output: xml table, png's

Identify lines present in data cube: where, and which transitions
        Output: xml table and png's

Creates separate cubes for each line found with transition or freq labeling
        Output: "N" CASA images with xml information

Creates clipped moment maps for each line (0, 1, 2... as requested)
        Output: CASA images, png's, xml information

Read in continuum map
        Output: CASA image and xml information

Find continuum sources to some selected depth
        Output: xml table, png's

Make Spectra at each continuum position
        Output: xml table, png's

ADMIT

Data Product Viewer

- Browser-based with style files similar to the ALMA calibration pipeline.

- The browser view is self-generated as ATs are run.

- The browser view is started by pointing the browser to the admit directory.

ADMIT

Timeline for Science users:

- Jan 2016: We are discussing with ALMA a plan for release of the ADMIT software for users. We are open to supporting a limited number of beta testers over the next few months

- TBD: Deployment of ADMIT pipeline within ALMA to create products for archive ingestion.
  - Date and final design to be decided in discussion with the ALMA project

- May 1, 2016: Delivery of completed software system
  - End of funded ALMA Development Project is April 30, 2016
  - Contract requires delivery of all software and documentation
  - Options for continued support will be explored with ALMA/NRAO