Science Mining the ALMA Archive

Peter Teuben, Marc Pound, Lee Mundy
University of Maryland, College Park, MD

What we will be doing:

In this just started ALMA Cycle 7 Development Study we plan to build a prototype of the ALMA Science Archive, and show via cloud computing how users can query this archive.

Based on results from **ADMIT** (ALMA Data Mining Toolkit) we compute rich science meta-data that users can query via remote Jupyter Notebooks.

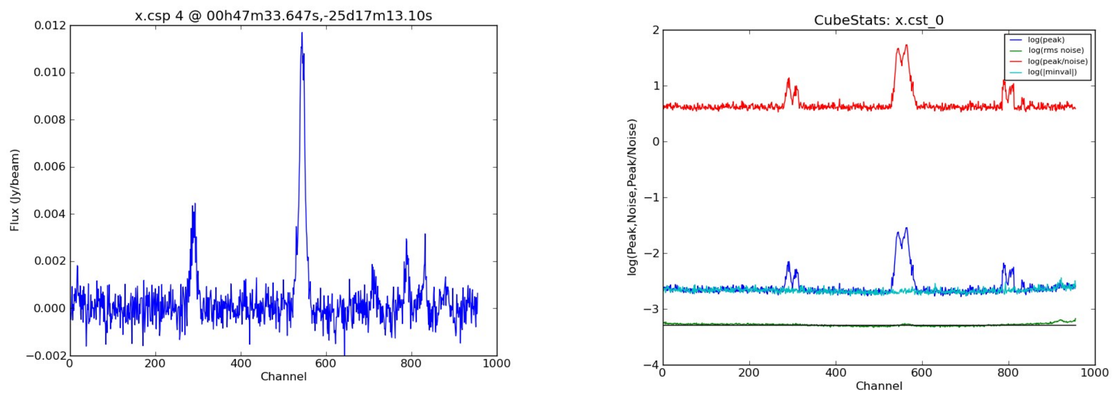


Figure 1: Example "spectra" used in the LineID task in ADMIT.
On the left a CubeSpectrum, on the right a CubeStats spectrum.

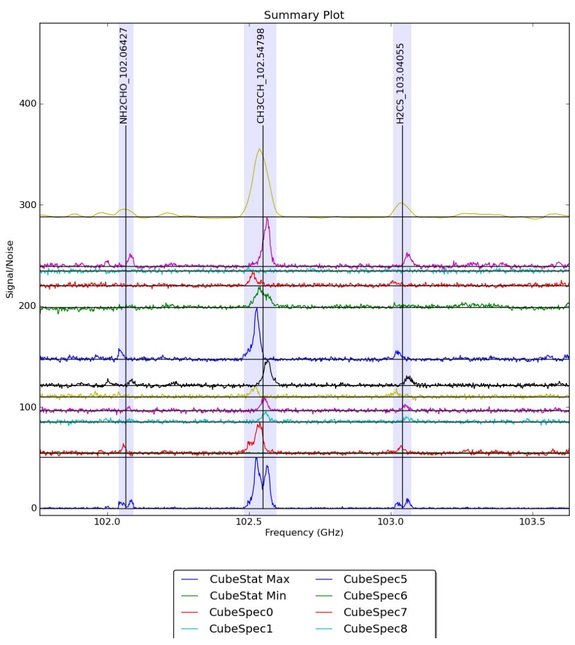


Figure 2: A final LineID is based on comparing a number of "spectra". Here 3 lines are identified, and in blue we show the channel range Where the line is identified from which LineCube's are extracted.

AQLite query example

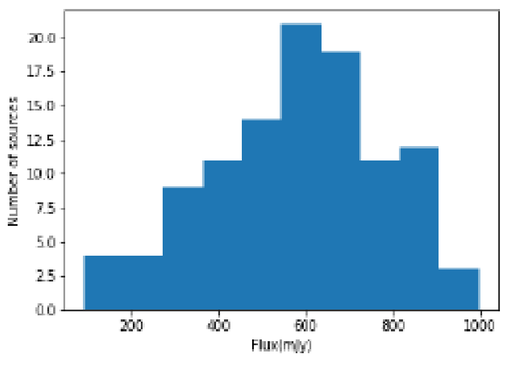
```
# define a typical astroquery payload
Payload = {
  "RA"          : "04 41 45.9",
  "Dec"         : "+25 41 27",
  "radius"      : "1 degree",
  "Abstract" : ["YSO", "young stellar object", "protostar"],
  "Lines"       : ["CO(2-1)"],
  "S/N"         : [">4"]
}

# get a panda's DataFrame
df = database.query(payload)
```

```
%matplotlib inline
import matplotlib.pyplot as plt

# create a new DataFrame
# by selecting on flux AND fwhm
df2 = df[(df["flux"]>100) & (df["fwhm"]>2)]

# plot the histogram
plt.hist(df2["flux" ])
plt.xlabel("Flux (mJy)")
plt.ylabel("Number of sources")
plt.show()
```



How will we do this?

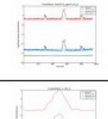
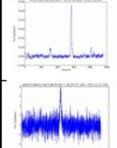
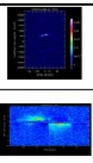
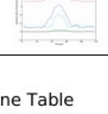
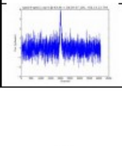
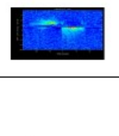
We will (on Amazon Web Services compute nodes):

- Grab a few dozen public data from Cycle 5
- Re-image (if needed) the calibrated visibilities
- Get the basic ALMA archive meta-data into our database
- Run ADMIT to obtain:
  - Signal and Noise characteristics
  - Spectral Line Identification
  - Line Cubes
  - Moment 0,1,2 for all Line Cube's
- Add ADMIT meta-data in our database
- Build AQLite (astroquery light) to return Pandas dataframes
- Provide a remote Jupyter Notebook for users

Combined ALMA and ADMIT Results Table

	ALMA Science Archive Results					ADMIT Results		
index	Project Code	Source Name	RA	Dec	...	SPW	Tasks	...
0	2017.1.005.S	CenA	12:34:56	01:02:03	...	23,25,27,29	6	...
1	2018.3.040.S	NGC 1234	16:17:18	-19:20:21	...	1,2,3,4,5,6	8	...

Spectral Window Table

index	ALMA index	SPW	# lines	# sources	CubeStats	CubeSpectrum	PV Slice	...
10	0	23	4	10				...
11	0	25	1	1				...

Line Table

index	Spectral Win. index	Frequency	Formula	Transition	velocity	startchan	endchan	...
9	10	110.20137	13CO	1-0	238.238	786	812	...
8	11	102.54798	CH3CCHv=0	6(0)-5(0)	179.740	529	586	...

Source Table

index	Spectral Win. index	Line index	Ra	Dec	Peak	Flux	S/N	...
41	10	9	12:34:56	01:02:03	12.3	20.1	30	...
42	10	5	12:34:49	01:03:00	3.6	7.5	5	...

Figure 3: Examples of the metadata tables we create from ALMA and ADMIT products for sophisticated science queries. The tables are linked via their indices (color highlights), allowing combined queries like "Sources with 13CO detected at S/N >10." Note we can also insert visual cues for the user such as thumbnail images of ADMIT products [Spectral Window Table].

Interested in playing with this archive?

Contact us here (Peter and Marc are at ADASS) or send us email. We hope to have a portal available Spring 2020 for testing.