ALMA, GBT, VLA, VLBA

IVOA Meeting, Heidelberg Germany

NRAO Data Management and Software

Brian Glendenning

Associated Universities, Inc.

Atacama Large Millimeter/submillimeter Array
Karl G. Jansky Very Large Array
Robert C. Byrd Green Bank Telescope
Very Long Baseline Array

NRAO

NRAO in 2013

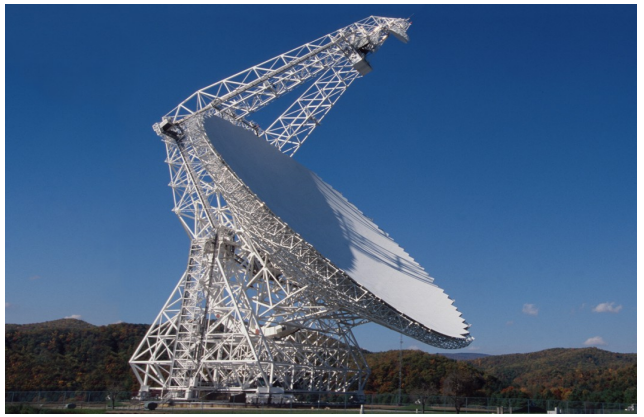
DMS

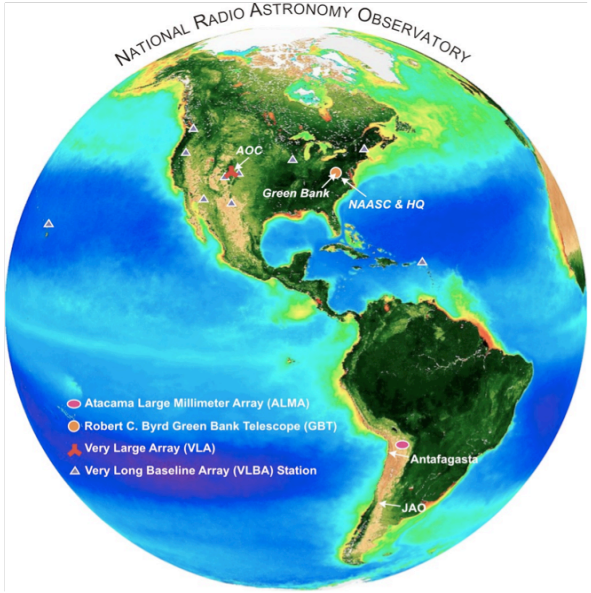
Complementary suite of facilities under Open Skies policy

[λ: m ➔ submm; δθ: arc-min ➔ milli-arc-sec]

Jansky VLA

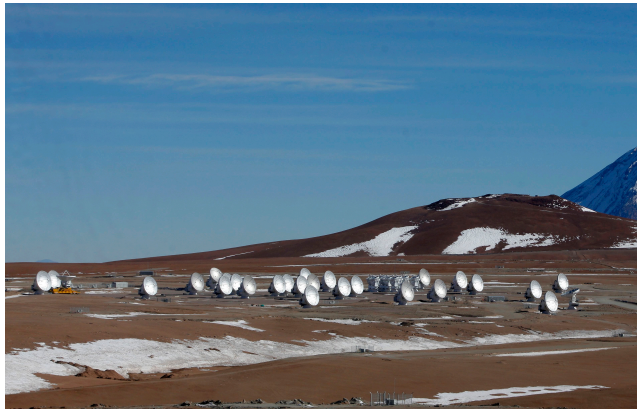
GBT





VLBA

ALMA



GBT and VLBA under threat ("divestiture" – NSF Portfolio Review)

Sample Science GBT: Surveys for water mega-masers

DMS

- 100m GBT
  - Most sensitive telescope to search for H_2O mega-masers at λ = 1.3 cm
  - 10 minutes for each galaxy
- Power collected by GBT from a typical mega-maser at 50 mJy

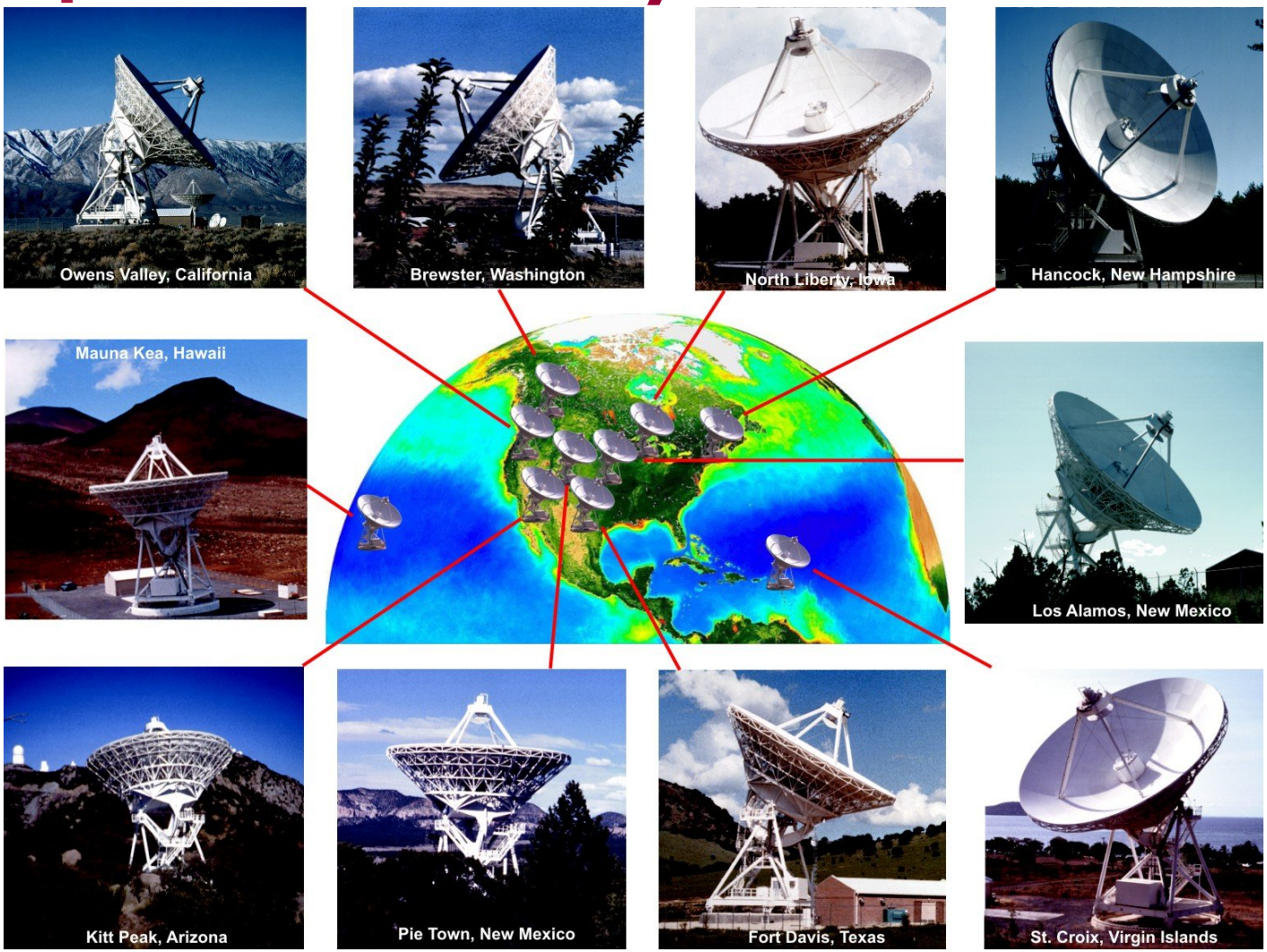
  ~ 10^{-17} watt

NRAO

VLBA: 10-μsec Astrometry

10
25m Antennas

λ = 1.3 cm
D = 5000 miles
λ/D ~ 0.33 mas



Karl G. Jansky Very Large Array (2012)

Plain of San Agustin, New Mexico

DMS

NRAO

Top Level Jansky VLA Performance Goals

DMS

- Providing orders of magnitude improvements

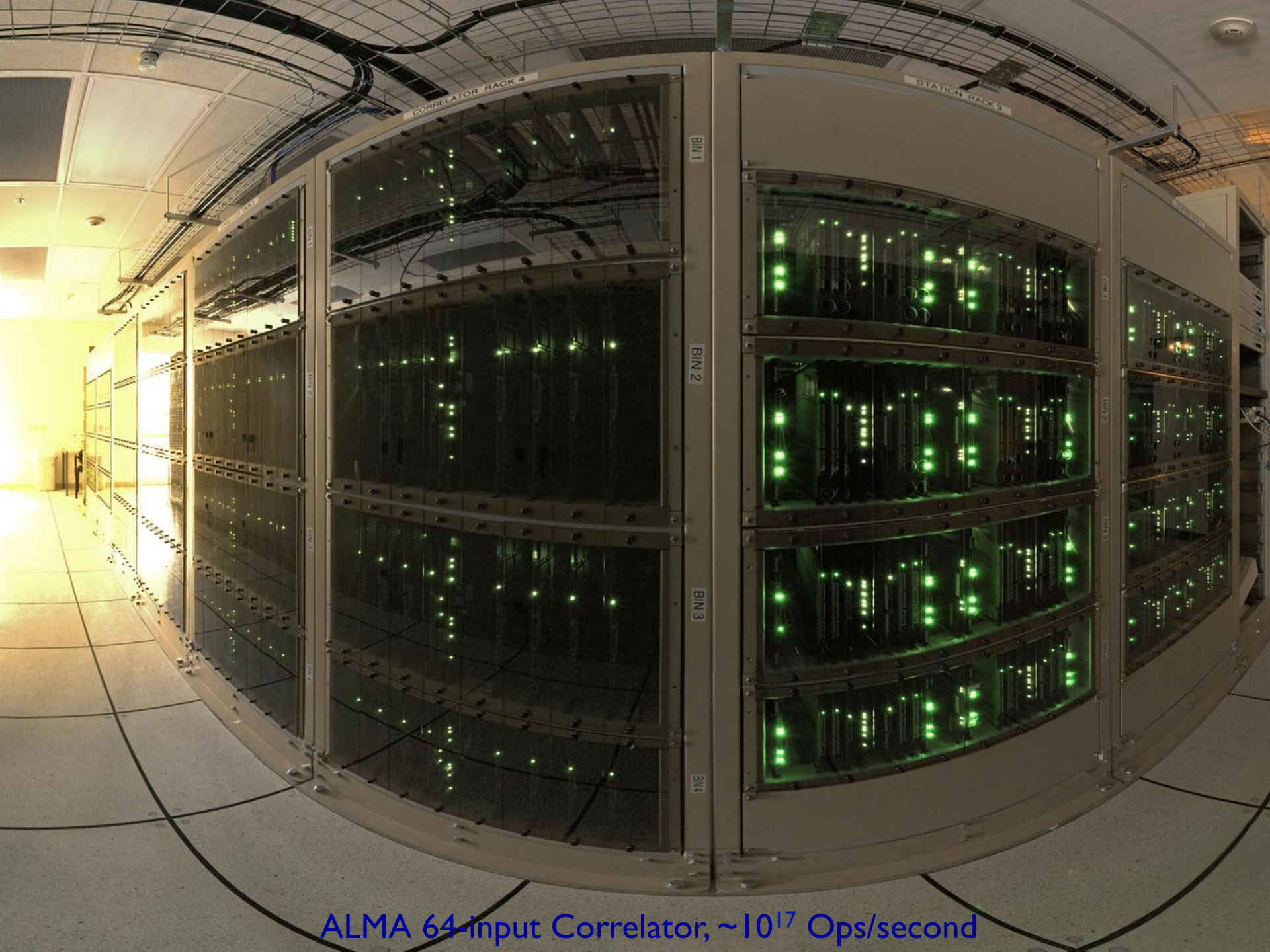
Parameter	VLA	Jansky VLA	Factor
Frequency Coverage (1 – 50 GHz)	22%	100%	5
Continuum Sensitivity (1-σ, 1 hr.)	30 μJy	3 μJy	10
Maximum BW in each polarization	0.1 GHz	8 GHz	80
# of full-polarization spectral windows	2	64	32
# of frequency channels at max. BW	16	16,384	1024
Maximum number of freq. channels	512	4,194,304	8192
Finest frequency resolution	381 Hz	0.12 Hz	3180

ALMA in Sept. 2012, ~50% antenna completion

ALMA March 2013, ~75% antenna completion

Raw Data

- ALMA & VLA use the (shared) Science Data Model/Binary Data Format
  - Complex (~60 table types)
  - XML with binary MIME attachments for bulk data
- GBT, VLBA: FITS BINTABLE based
- Data rates to Archive
  - ALMA 6.4 MB/s average, 64 MB/s peak (16-bit values)
  - VLA 7.5 MB/s - 75 MB/s (32 bit values)
  - GBT < 1 – 50 MB/s (pulsar data 200 MB/s, not archived)
  - VLBA 1 – 10 – 100 MB/s (rare)
- Potential data rates
  - ALMA = 3 GB/s (correlator clusters already handle this data rate, limitation is Archive) (32 bit values)
  - VLA = 16 GB/s
  - GBT = 1 GB/s (new spectrometer being commissioned)

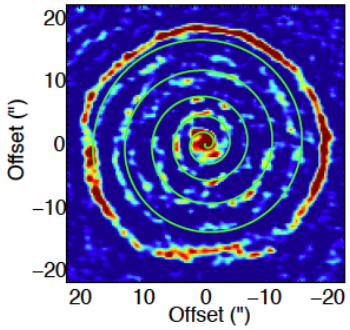


ALMA 64-input Correlator, ~10^{17} Ops/second

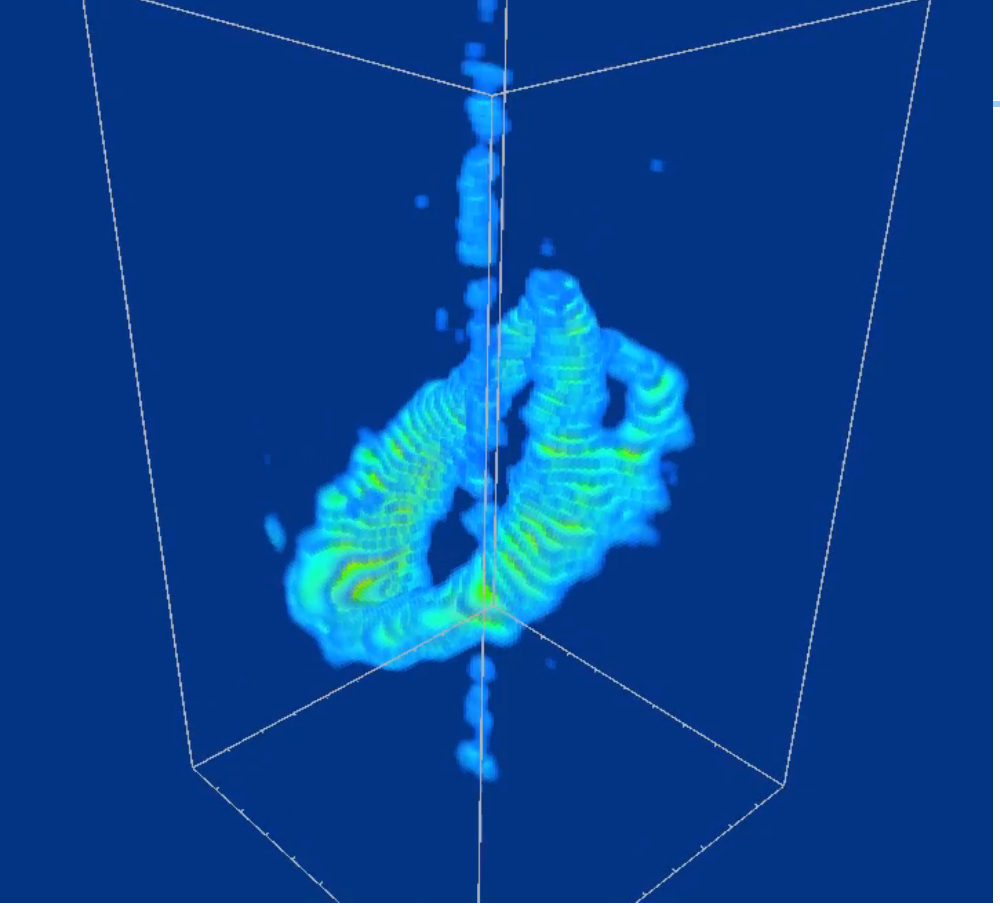
Image Data

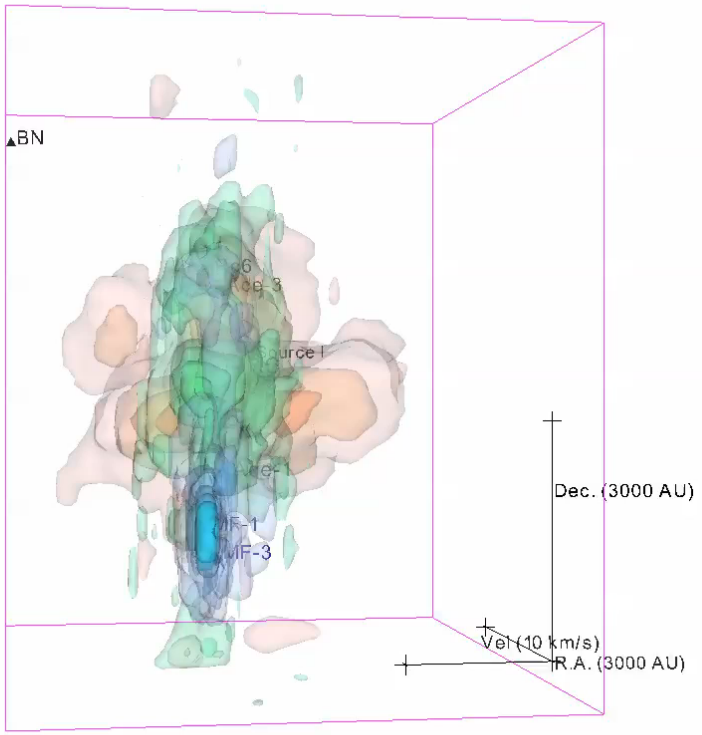
- Images typically produced as end result of observations and data reduction
  - But also: pulsar data, spectra, model-fits to raw data, …
- ALMA estimate: 5% of raw data volume (=10 TB/year at current specs)
- Common axes: RA, Dec, frequency/velocity, polarization
  - Rare: time
  - Starting to commonly have multiple "spectral windows" with varying #chan, resolution
    - Typically represented as multiple image cubes (often have different intents for the various spectral windows)
- Typical: 1000^3 (Gpix), Possible: > $10{,}000^3$ (Tpix), (x1-4 polarizations)
  - Do not have good visualization solution for very large images
  - Computation infrastructure should be OK for large images (e.g., tiled), but little tested

ALMA Measures Stellar Feedback



- ALMA's high sensitivity high resolution image measures the mass (0.003 M_{sun} and timescale (200 years) of feedback to the interstellar medium from the AGB star R Scl and reveals the star to be an unrecognized binary

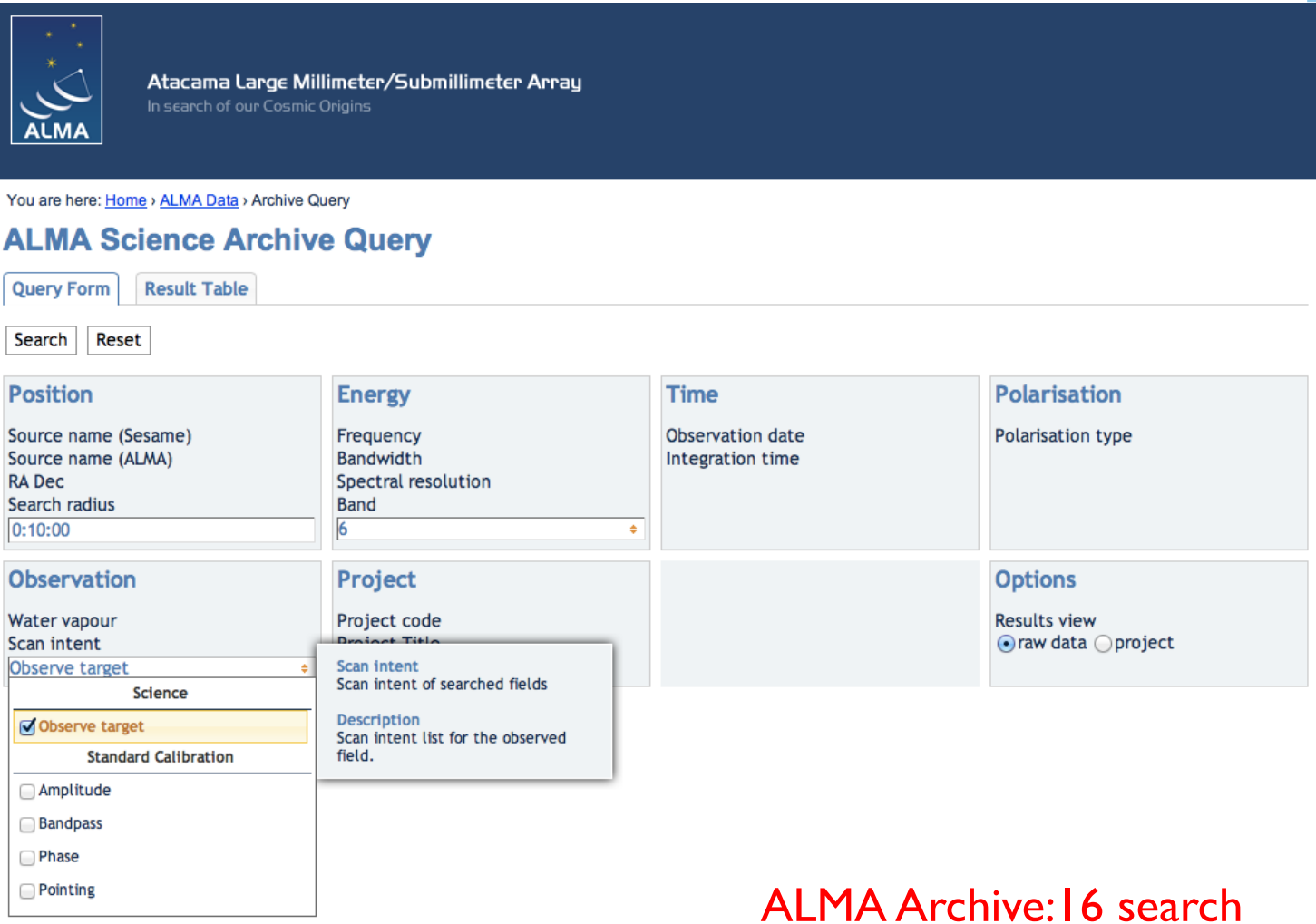




NRAO

Archives

- Two separate codebases: ALMA Archive (ESO, CADC), NRAO (non-ALMA) Archive
- ALMA Archive
  - Started serving data in December when first data left proprietary period, full software suite deployed early February
  - Multi-site (4), distributed; master repository in Santiago
  - Based on ESO NGAS, Oracle, Oracle products (replication etc)
- NRAO Archive
  - Live for 10+ years
  - Two sites (only Socorro visible)
  - Based on ESO NGAS, Oracle
  - Significant rework planned
- Neither visible from VO (ALMA does use significant parts of VO software stack)



ALMA Archive: 16 search parameters.

NRAO Archive, advanced search form: 23 search parameters (plus some output specification parameters)

General Search Parameters :

Telescopes ☑ All ☐ Jansky VLA ☐ Historical VLA ☐ VLBA ☐ GBT

Project Code
GBT: AGBT12A_055
JVLA: 12A-256

Project Session

Dates From

Observer Name

Archive File ID (partial strings allowed)

To (2010-06-21 14:20:30)

Position Search :

Target Name

Search Type: SIMBAD or NED

Min. Exposure (secs)

RA or Longitude (04h33m11.1s or 68.29d)

DEC or Latitude (05d21'15.5" or 5.352d)

Equinox: J2000

Search Radius: 1.0' (1d00'00" or 0.2d)

- OR - ☐ Check for automatic VLA field-of-view, freq. dependent.??

Observing Configurations Search :

Telescope Config ☑ All ☐ A ☐ AB ☐ BnA ☐ B ☐ BC ☐ CnB ☐ C ☐ CD ☐ DnC ☐ D ☐ DA

Observing Bands ☑ All ☐ 4 ☐ P ☐ L ☐ S ☐ C ☐ X ☐ U ☐ K ☐ Ka ☐ Q ☐ W

Sub_array ☑ All ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Frequency Range (In MHz : 1665.401 - 1720.500)

Polarization: ALL

Receiver ID: ALL

Data Type: ALL

Backend ID: ALL ((GBT only - select GBT in Telescopes list))

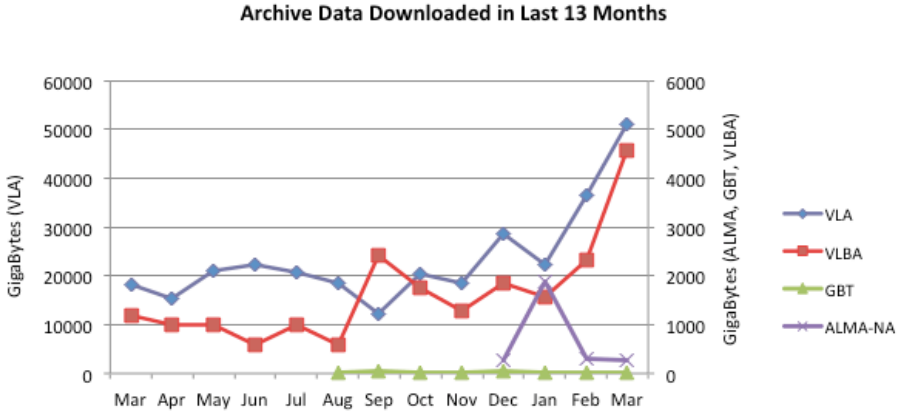
Data Retrieval from Archives

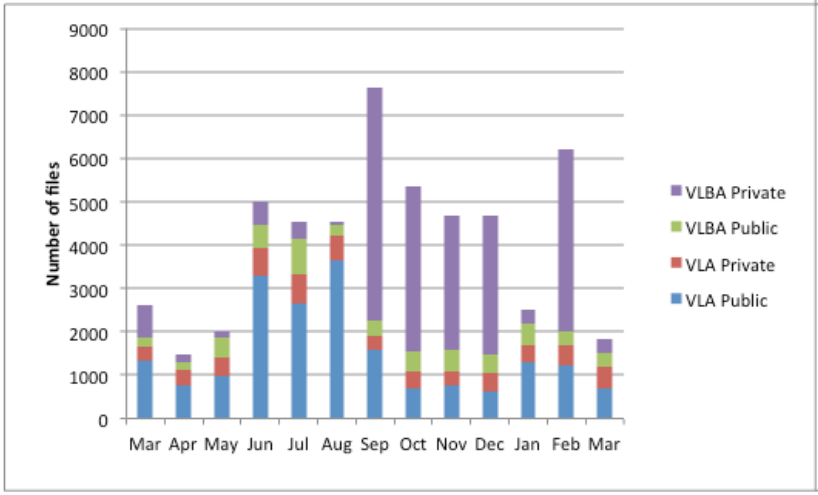
DMS

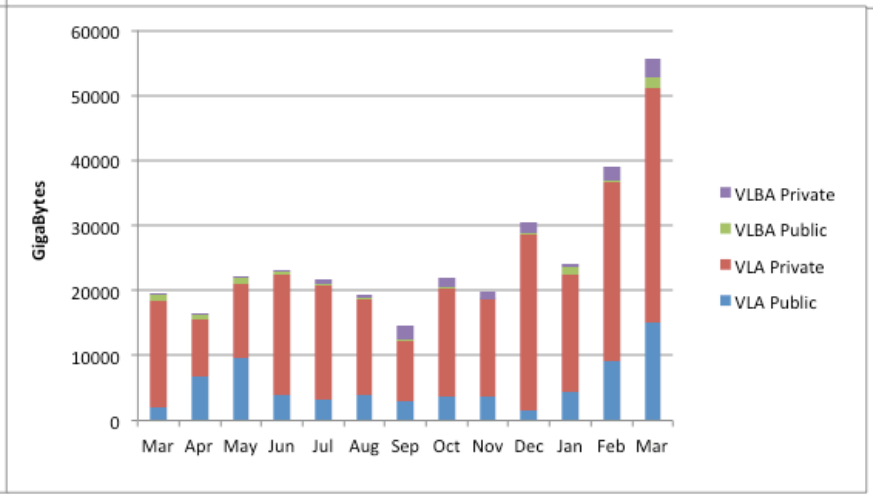
Dominated by VLA. Most Only ~1 month of ALMA public data. GBT data retrieved from filesystem

FY2013 Q2 Totals

	# of Files	Data Volume (GB)
ALMA-NA		
Proprietary	0	0
Public	2,607	2,469
Total	**2,607**	**2,469**
GBT		
Proprietary	24	3
Public	77	34
Total	**101**	**37**
VLA		
Proprietary	1,338	81,432
Public	3,242	28,669
Total	**4,580**	**110,101**
VLBA		
Proprietary	4,841	5,322
Public	1,135	3,145
Total	**5,976**	**8,467**
Pipeline Images		
Downloaded	**137**	







Data Processing - CASA

- Main package used for ALMA and VLA
- Unofficially used for other radio telescopes, developments (25% of helpdesk tickets)
- Long, complex (=difficult) development history (was AIPS++)
- C++ libraries, applications; Python (ipython) user interface, scripting
  - Very powerful facilities available to Python developers (much of the C++ library is exposed), not pure Python
- Significant recent performance optimization (parallelization, I/O) – small cluster
- Powerful image facilities in library
  - Separation of data model from file format, e.g. the same interface can use CASA, FITS, AIPS, Miriad, HDF5, Gipsy, Miriad format files
    - Not sure all have been exercised recently
  - Tiled (chunked) image storage for native images – can traverse data quickly either spectrum by spectrum or plane by plane
  - Good WCS support (wcslib for RA/DEC, good support for spectral/polarization axes)
  - Blanking & Regions
  - Virtual/OTF image calculations (e.g., image calculations with no additional storage)

Images reduced in CASA

ALMA: R Sculptoris

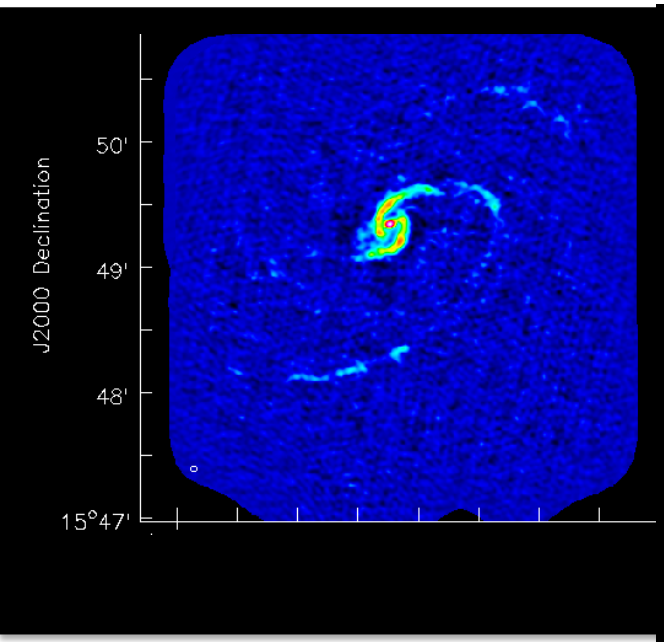
Maerker et al.

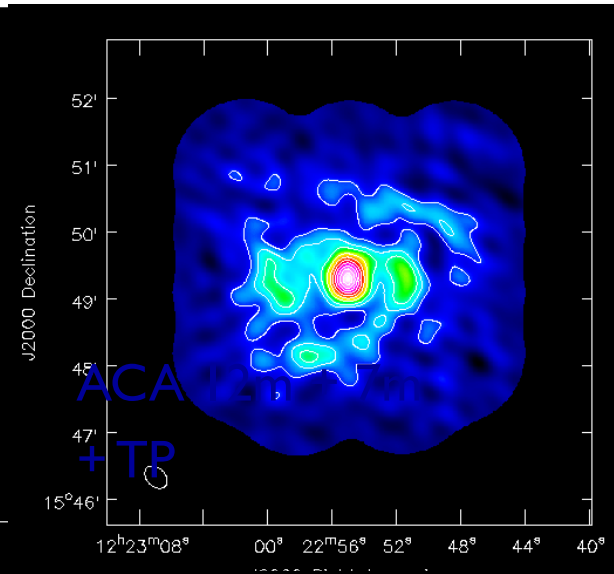
VLA: Manatee Nebula (W50)

Golap et al.

Images reduced in CASA

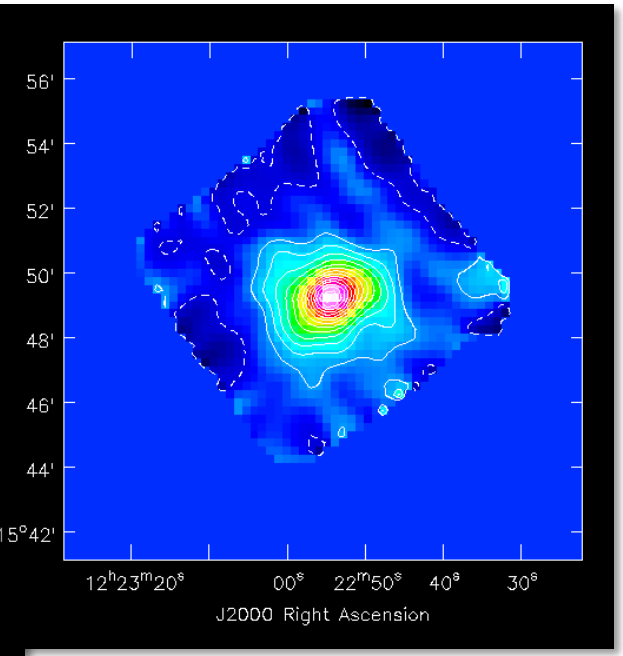
ACA 7m





ACA 12m

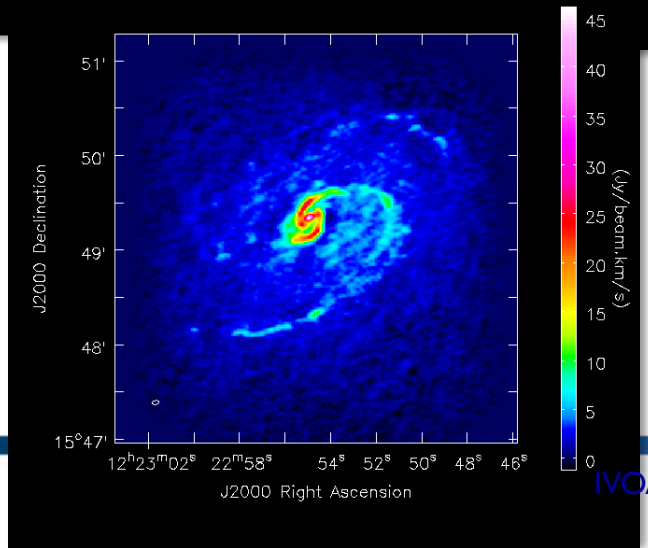
+TP



ALMA 12m

ACA TP

All combined



(from D. Iono)

Images reduced in CASA

VLA: GLOSTAR

C-band Galactic Plane Survey

PI: A. Brunthaler

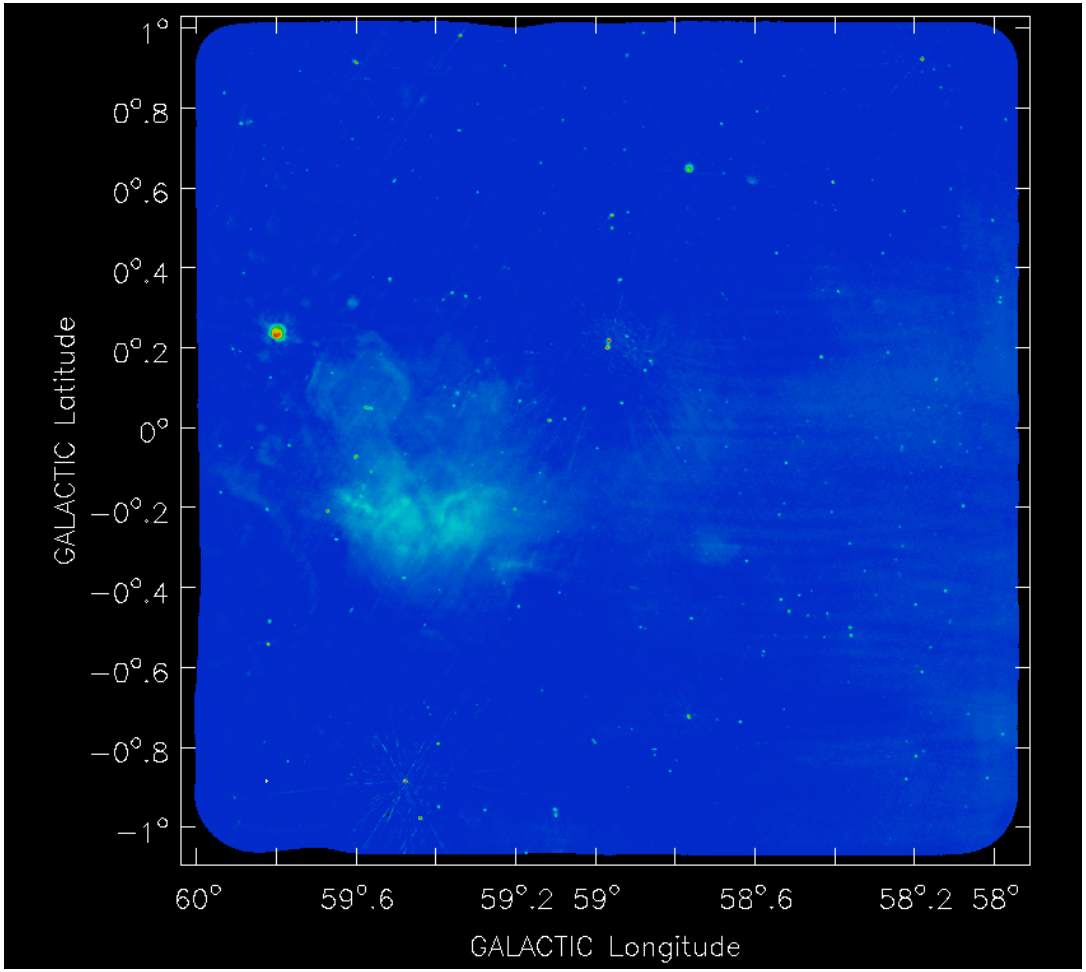
Pilot image:

G059+0.0

1276 pointings, 2x2deg field

Effelsberg added as a starting model in clean

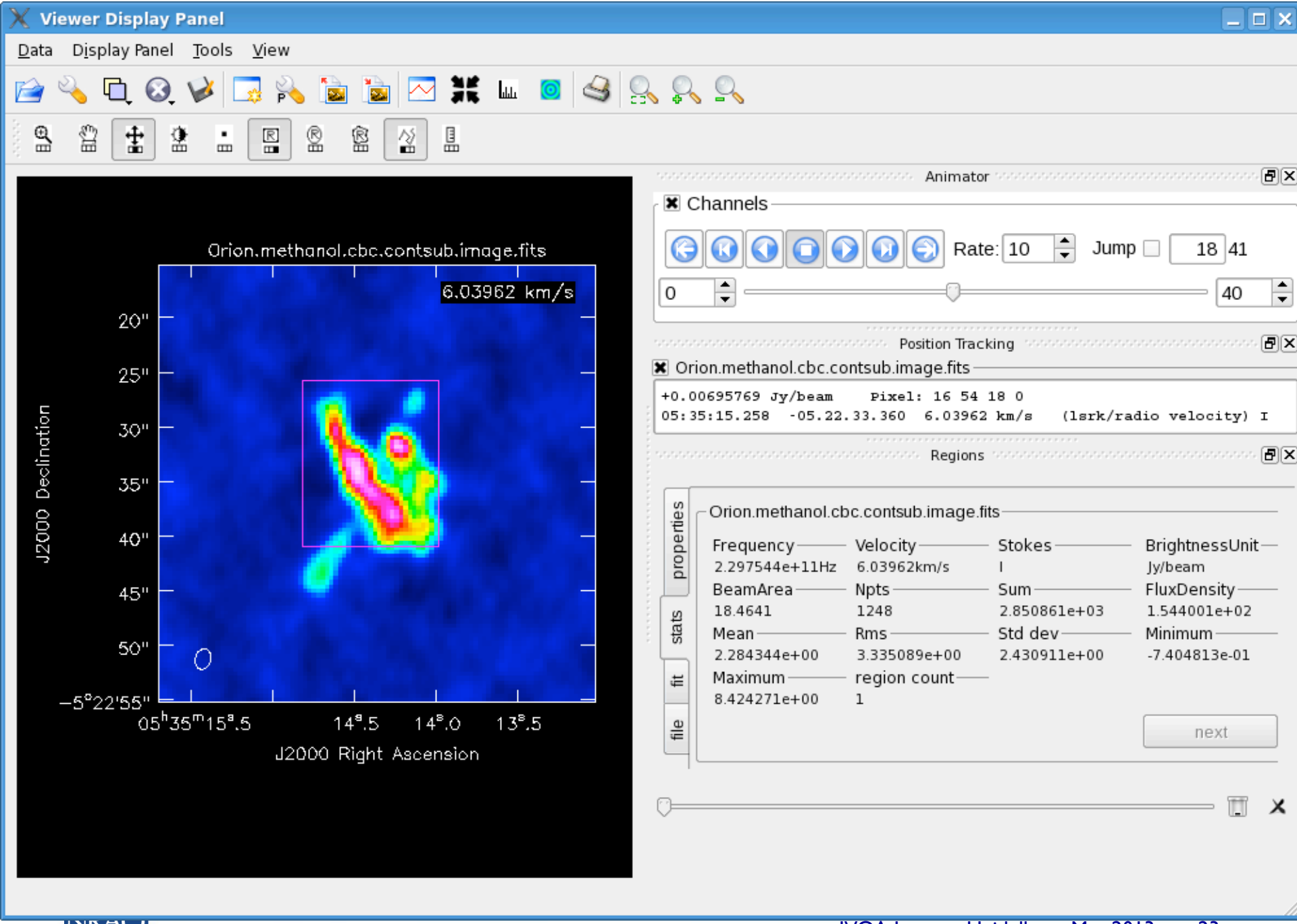
Carasco-Gonzales et al. (in prep)

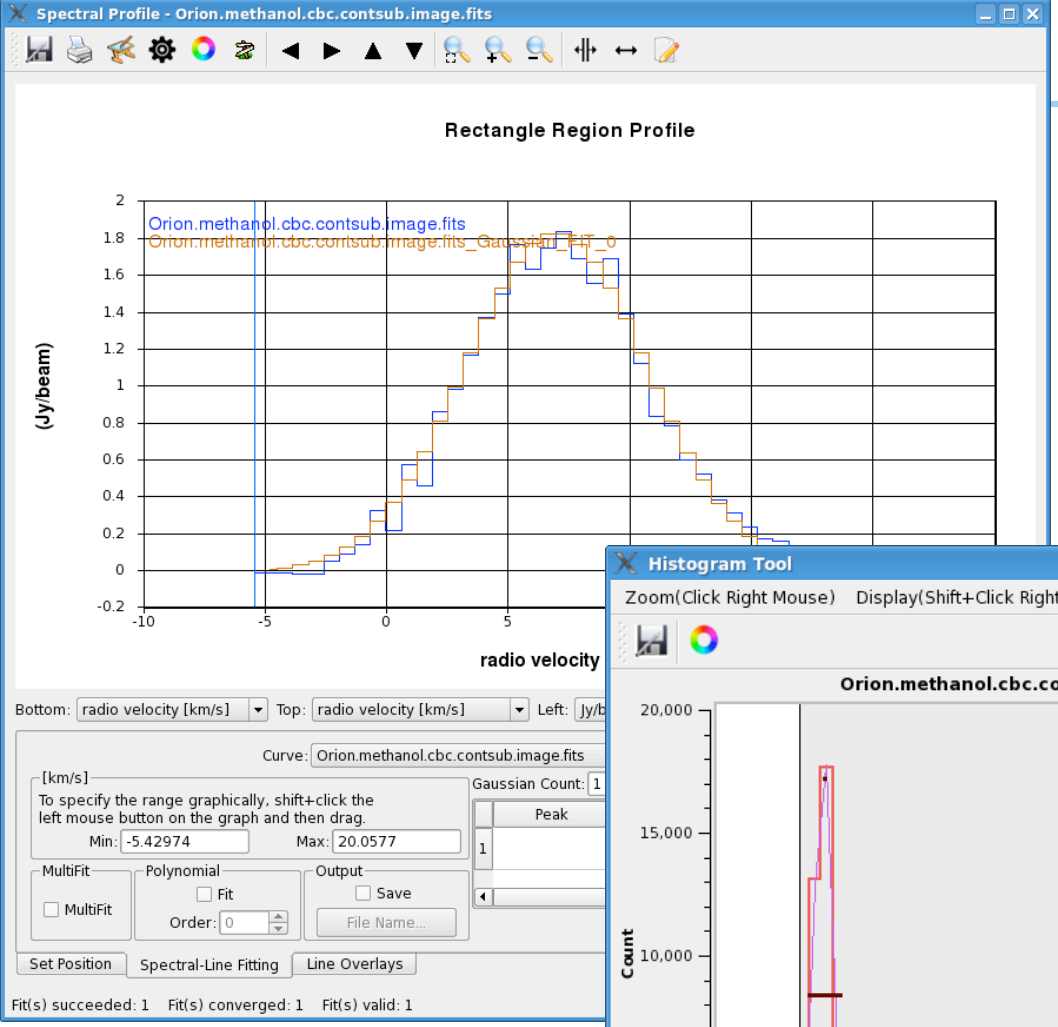


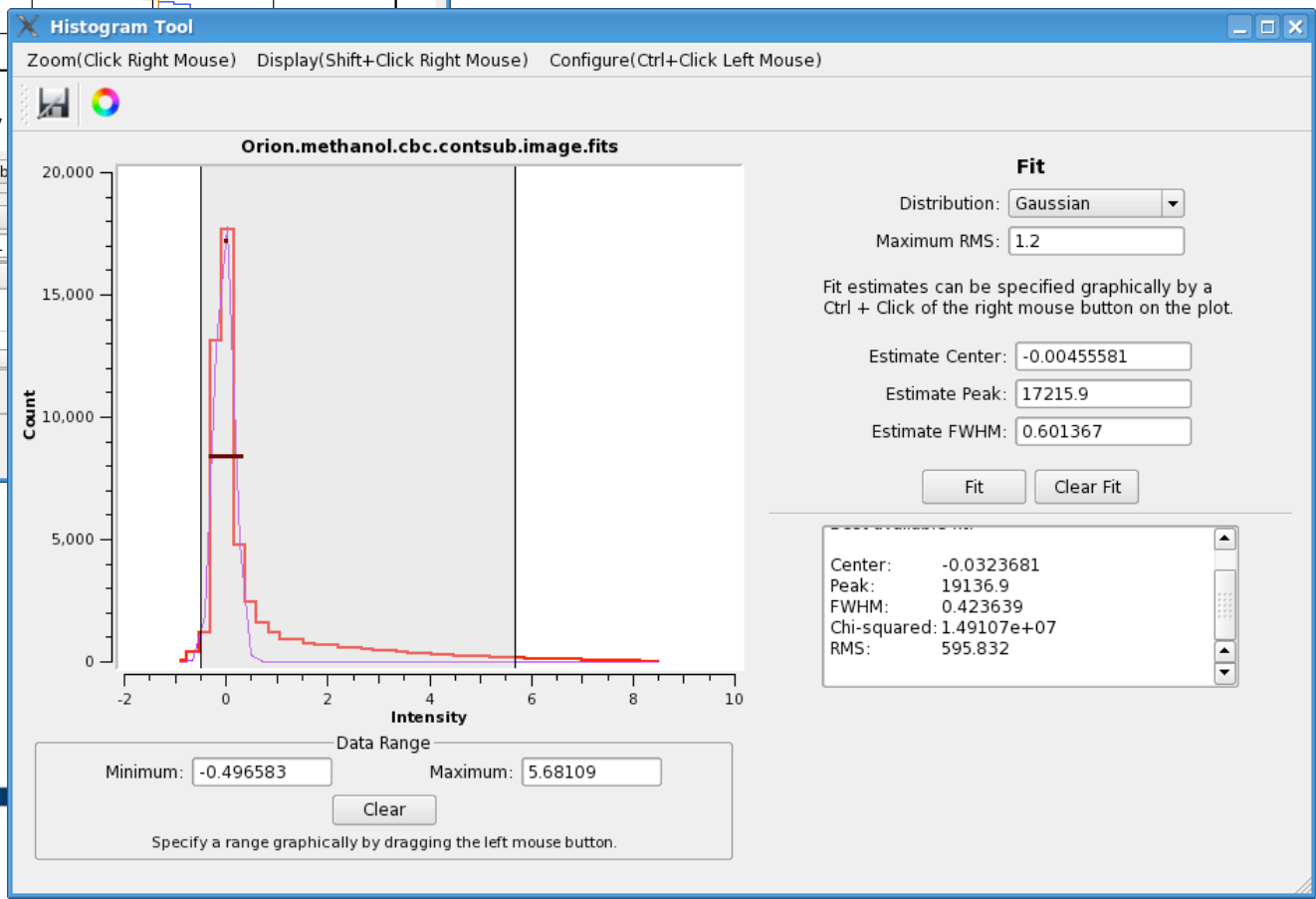
CASA Viewer

- Good "traditional" image viewer application for N-D images
  - Selected after survey by ESO for 3D data
- Steady development (1+ FTE)
- Weaknesses
  - Not tested with very large images (although tiles should help)
    - Should have a client/server (cluster) architecture
  - Sophisticated analysis tasks
    - Often in the library, "only" needs to be made visible in application (AIPS++ had many of these)
  - Model fitting (e.g., of physical models)
  - 3D displays (e.g., isosurfaces, volume rendering, stereoscopy)
  - VO interface

Looking for a collaboration for a "large data" visualization application


Viewer Display Panel
Data
Display Panel
Tools
View
Orion.methanol.cbc.contsub.image.fits
6.03962 km/s
20"
25"
30"
35"
40"
45"
50"
−5°22'55"
J2000 Declination
05h35m15s.5
14s.5
14s.0
13s.5
J2000 Right Ascension
Animator
Channels
Rate:
10
Jump
18
41
0
40
Position Tracking
Orion.methanol.cbc.contsub.image.fits
+0.00695769 Jy/beam Pixel: 16 54 18 0
05:35:15.258 -05.22.33.360 6.03962 km/s (lsrk/radio velocity) I
Regions
properties
stats
fit
file
Orion.methanol.cbc.contsub.image.fits
Frequency
2.297544e+11Hz
Velocity
6.03962km/s
Stokes
I
BrightnessUnit
Jy/beam
BeamArea
18.4641
Npts
1248
Sum
2.850861e+03
FluxDensity
1.544001e+02
Mean
2.284344e+00
Rms
3.335089e+00
Std dev
2.430911e+00
Minimum
-7.404813e-01
Maximum
8.424271e+00
region count
1
next


Spectral Profile - Orion.methanol.cbc.contsub.image.fits
Rectangle Region Profile
Orion.methanol.cbc.contsub.image.fits
(Jy/beam)
radio velocity
Bottom: radio velocity [km/s]
Top: radio velocity [km/s]
Left: Jy/b
Curve: Orion.methanol.cbc.contsub.image.fits
[km/s]
To specify the range graphically, shift+click the left mouse button on the graph and then drag.
Min: -5.42974
Max: 20.0577
Gaussian Count: 1
Peak
MultiFit
Polynomial
Fit
Order: 0
Output
Save
File Name...
Set Position
Spectral-Line Fitting
Line Overlays
Fit(s) succeeded: 1
Fit(s) converged: 1
Fit(s) valid: 1

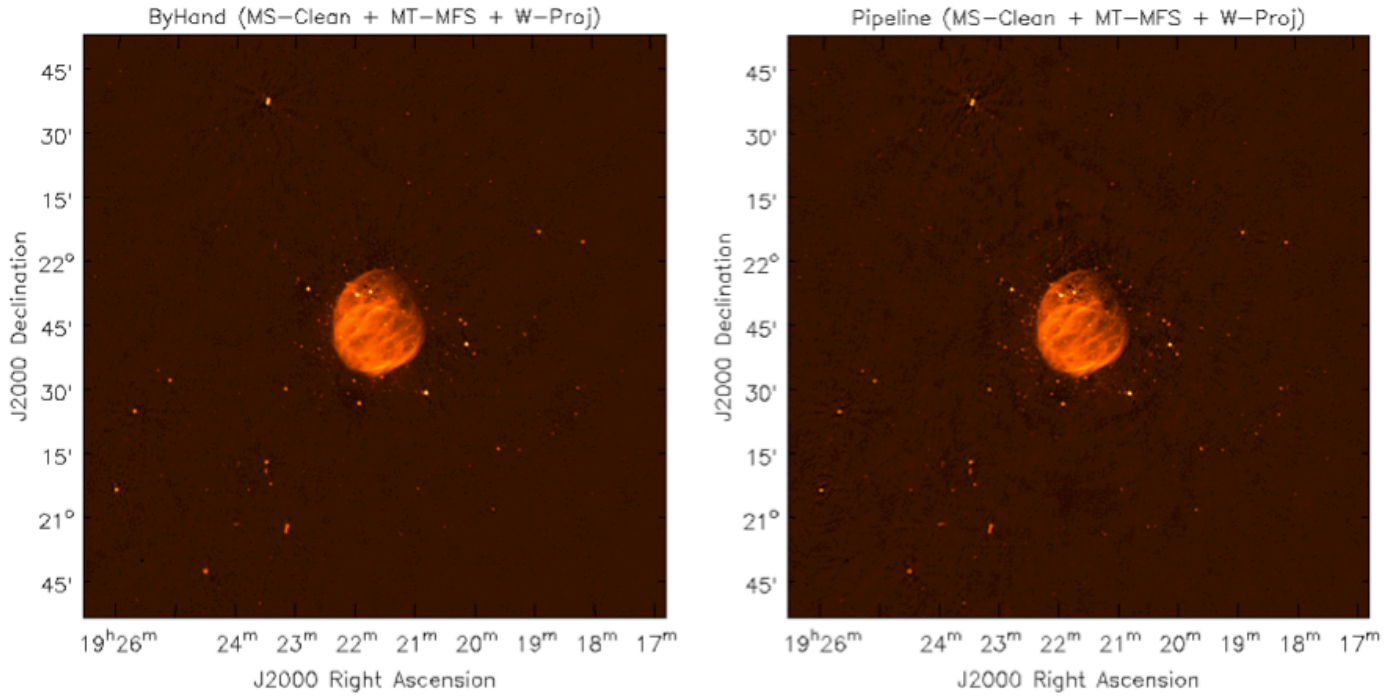

Histogram Tool
Zoom(Click Right Mouse)
Display(Shift+Click Right Mouse)
Configure(Ctrl+Click Left Mouse)
Orion.methanol.cbc.contsub.image.fits
Count
Intensity
Data Range
Minimum: -0.496583
Maximum: 5.68109
Clear
Specify a range graphically by dragging the left mouse button.
Fit
Distribution: Gaussian
Maximum RMS: 1.2
Fit estimates can be specified graphically by a Ctrl + Click of the right mouse button on the plot.
Estimate Center: -0.00455581
Estimate Peak: 17215.9
Estimate FWHM: 0.601367
Fit
Clear Fit
Center: -0.0323681
Peak: 19136.9
FWHM: 0.423639
Chi-squared: 1.49107e+07
RMS: 595.832


NRAO

Pipelines

- NRAO has traditionally put only modest resources into pipelined image production
  - E.g., surveys; degraded resolution continuum primary beam images
- ALMA has always had requirements for a pipeline (CASA based)
  - Will be in use for Cycle 1 (starting in June)
- VLA calibration/flagging pipeline in production since January (CASA based)
  - Imaging will be added in a later step
  - Currently being merged with ALMA pipeline
- GBT: Spectral & Mapping pipeline (Imaging moving to CASA)
- VLBA: AIPS-based pipeline available, being refurbished
- Pipeline products not yet ingested into the Archive (September)
- Pipelined images will be routinely produced and archived, except for the VLBA, in the coming year – a significant change for the radio community

VLA Pipeline Imaging comparison



Left: L-band image of G55.7+3.4 produced from data flagged and calibrated by hand; the rms noise is 11.5 μJy/beam. Right: an image made from data flagged and calibrated by the VLA calibration pipeline; the rms noise is 12.2 μJy/beam. Differences in the source structure and/or source flux density are dominated by the uncertainty in the deconvolution process, not the calibration and flagging (images provided by Urvashi Rao).

GBT Pipeline Image

Image of Proprietary Data Deleted

Courtesy: Youngmin Seo (Steward)

VO & NRAO (Mildly obnoxious)

- Modest engagement to date
- NRAO hosts some VAO standards and technology development activities (notably Tody); < 2 FTE, entirely VAO funded
- Little VO compatibility in NRAO developed software
  - Spectral line catalogue – splatalogue – uses SLAP
- Not coincidentally (chicken v. egg?) little interest from our community
  - I have ***never*** been asked to make VO a priority by any NRAO user not directly associated with VO projects
  - But I ***often*** get the opposite comment: don't waste effort on VO
- Why?
  - Original sin: fundamental data output (3D+ images) not represented in VO
  - Conservatism of radio astronomers
  - Lack of pipelined images
  - Insufficient outreach

Now or never? (Still mildly obnoxious)

- ALMA & NRAO telescopes provide data to a lot of users
  - NRAO + ALMA = 2k proposals per year
  - NRAO = 3.5k users (5 years)
- Pipeline efforts are young, no ingrained habits yet, will bring non-traditional radio astronomers
- VAO has some funding, NRAO can provide some modest in-kind contributions
  - Minimum: VO enable CASA views, NRAO and/or ALMA Archive
- VAO funding must be spent by Oct 1 2014

- Time is tight: "the best is the enemy of the good" (Voltaire)

Shameless plug

- NRAO is recruiting a Head of Software
  - https://careers.nrao.edu/applicants/Central?quickFind=50910

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc

www.nrao.edu