



NORTH AMERICAN ARC  
ALMA Regional Center

North American  
ALMA Science  
Center



## Data Reduction Workshop Lessons Learned

NAASC Memo #109

Adam Leroy and Jim Braatz

Date: April 15, 2013

### ABSTRACT

This document summarizes “lessons learned” based on the first two NAASC data reduction workshops, held 28 Feb – 1 Mar 2013 and 1-2 December 2011. It is primarily intended to inform future workshops along these lines and so focuses on suggested improvements to the existing material. This document replaces the “lessons learned” from the first workshop (NAASC Memo #109).

## WORKSHOP SUMMARY

The most recent NAASC Data Reduction Workshop was a 2-day program held on 28 February – 1 March, 2013. It aimed at teaching novice CASA users how to calibrate, image, and analyze ALMA data. The conference website is here:

[https://science.nrao.edu/facilities/alma/naasc-workshops/alma\\_dr](https://science.nrao.edu/facilities/alma/naasc-workshops/alma_dr)

and the agenda was broken into two days as follows:

### Day 1: CASA Basics and Calibration

On Day 1 we introduced CASA syntax and the CASA approach to data processing. We gave an overview of the Measurement Set, and discussed the steps needed to calibrate ALMA data, including bandpass, phase, and amplitude calibration. We also described the tools needed for data inspection.

Time	Presentation
8:30-9:00	<a href="#">Welcome, Logistics, ALMA Status</a>
9:00-10:00	<a href="#">Intro to CASA and the ALMA Measurement Set</a>
10:00-10:30	<a href="#">Intro to Interferometric Calibration</a>
10:30-10:45	Coffee
10:45-11:15	<a href="#">Overview of Calibration in CASA</a>
11:15-12:00	<a href="#">Data Inspection in CASA</a>
12:00-1:00	Lunch
1:00-1:45	<a href="#">Bandpass Calibration</a>
1:45-3:00	<a href="#">Phase and Amplitude Calibration</a>
3:00-3:15	Coffee
3:15-5:00	Hands-On Calibration Examples

## Day 2 : Self-Calibration and Imaging; ALMA Archive and Science Data

On Day 2 we introduced the science data products provided by ALMA, and then focused on producing sky images from calibrated ALMA  $u-v$  data. Topics included imaging, cleaning, self-calibration, and using the CASA viewer and other image tasks.

Time	Presentation
8:30-9:00	<a href="#">What's in the ALMA Archive/Data Package?</a>
9:00-9:30	<a href="#">Imaging Interferometric Data</a>
9:30-10:15	CASA Imaging Example
10:15-10:30	Coffee
10:30-11:00	Inspecting Image Cubes
11:00-12:00	Hands-On Imaging
12:00-1:00	Lunch
1:00-1:30	<a href="#">Self-Calibration</a>
1:30-3:00	Hands-On Self-Calibration Example
3:00-3:15	Coffee
3:15-5:00	Special Topics, Mosaics, Band 9

Registration and the web page were handled nicely by Dongchan Kim (for both workshops) and the CIS staff, especially Dan Klopp and David Halstead, provided fantastic support, setup, and breakdown for the event. NAASC Staff Aaron Evans, Carol Lonsdale, and Alison Peck helped promote and organize the event and Jim Braatz, Jennifer Donovan-Meyer, Mark Lacy, Adam Leroy, Juergen Ott, Nuria Marcelino, and Kim Scott presented, prepared materials, and offered hands-on training. Initial registration was small in number, but in the end twenty people attended – the limit for the room 230 venue. Each was set up with their own “thin client” and an account on the NAASC cluster; two accounts shared each node and each account had its own copy of data and reduction scripts in its directory.

## SCHEDULING CONSIDERATIONS

- *Edgemont Road 230 is viable but limits the workshop to ~20 participants.* If the auditorium is otherwise occupied, ER 230 represents a fine venue for a smaller workshop. If the auditorium is available then participation could instead be ~30.
- *Avoid overlap with the colloquium.* This was only a minor factor for the ~20 attendants at the 28 Feb workshop but caused significant issues at the first workshop. It is worth avoiding Thursday afternoon or catering to a smaller crowd.
- *Avoid major proposal deadlines.* The second workshop coincided with the HST proposal deadline, which limited attendance and person-power available and represented a source of distraction for participants.
- *Publicity effort pays off: target postdocs, graduate students, nearby institutions, and PIs; emphasize the availability of archival and SV data.* Registration was initially low. Future workshops should advertise early and often. Fruitful avenues include: investigators on successful proposals and geographically close institutions. The advertisement should explicitly target students and postdocs and note that the skills taught are directly relevant to exploitation of ALMA SV and archival data.

## CONTENT CONSIDERATIONS

- *Mixing hands-on work with presentations worked very well:* We mixed short presentations with slide-based walkthroughs of scripts that they executed and played with on their own computers. This worked very well and the feeling was that the more they are typing and following on their own, the better.
- *Add more walked-through hands-on examples.* The attendees found these most helpful. We could add walk-throughs of more sophisticated imaging (e.g. a mosaic) and give a walk-through of a more involved self-cal script.
- *It was good to have the presentations in PDF format available on the web site beforehand:* Many people opened them and followed along.

- *Give them scripts:* They appreciated having the full scripts that we went through. It let them decide how much typing/refining to do rather than forcing them to type everything new (avoiding snafus from typos).
- *The slides should be linked very clearly to the scripts* (e.g. see the “bandpass” presentation). Otherwise, in the walk-throughs it may not be clear what command was being run on a given slide. Put the commands on the slide to clean this up.
- *Keep the hands-on and walk-through scripts completely autonomous so that executing the commands in one session does not require successful completion of the previous session.* This avoids them messing up the afternoon tutorial if they hit a snag in the morning, miss a session, or don’t finish part of the script.
- *Make a tarball of data+scripts available for download:* At both workshops, this was a common request. That means keeping the data somewhat manageable in size and using public data.
- *Re-evaluate calibration content in light of the pipeline:* The user experience and needed knowledge will change as the pipeline comes on. Look into this for the next workshop.
- *Find a test data set with more antennas but still TDM/small:* Currently the scripts use Science Verification data with 8-14 Antennas. To more accurately reflect a realistic experience, the scripts should be updated to use ~30+ antenna data sets when this becomes available. They need to be as small as possible (TDM, smoothed in time) to keep the experience real-time and make it possible to download the data.
- *Emphasize interaction with the archive and data exploration/image analysis more as the archive develops:* With the maturation of the pipeline and delivery of calibrated data, data and archive exploration are going to become more important parts of improving ALMA’s impact. Future workshops should continue to slowly shift emphasis in this direction (e.g., lead an archive walkthrough, continue to refine the viewer tutorial, Splatalogue, image analysis demos).

## STAFFING CONSIDERATIONS

- *About one staff present per five attendees makes sense.* For a hands-on workshop this ratio seemed to work well. The second workshop was slightly overstaffed in this respect. A few attendees tend to benefit from extra attention, especially those who are not already familiar with the computing environment.
- *Three to four staff to give presentations worked well.* Most staff gave two presentations and three would have been viable. The presentations are in pretty good shape. It takes 1-2 days per presentation, per talk, to update the slides to reflect new scripts.
- *Scripts need to be tested thoroughly before the event.* This took about two full days for each of two testers for the Feb 2013 workshop.
- *Script generation took about a week.* This involved both new material generation and clean-up/updating to reflect new versions of CASA and new best practices. Based on our “content suggestions,” this will probably still be necessary in future presentation.
- *The rough cost of a workshop is about 40-50 employee-days, less if it begins repeating rapidly.* This assumes a week of CST time. The figure might be as low as ~20-30 if materials become thoroughly recyclable (which would be the case if we held the workshops very close together). In the immediate future pipeline and data evolution probably push us more towards the ~40-50 number as we will probably have to keep revising material.