EUROPEAN ARC
ALMA Regional Centre

ALMA Science Data Products (and their generation)

Dirk Petry (ESO)

Outline:

- ALMA Quality Assurance (QA) Purpose
- QA Procedure
- QA Products

Relevant documents:

Lundgren, A., 2013, ALMA Cycle 2 Technical Handbook Version 1.1
Petry, D. et al., 2014, Proc. SPIE, Volume 9152, id. 91520J
Petry, D. et al., 2014, ALMA QA2 Data Products for Cycle 2, Version 2.2

EUROPEAN ARC
ALMA Regional Centre

ALMA QA

"The goal of ALMA Quality Assurance (QA) is to deliver to the PI a reliable final data product that has reached the desired control parameters outlined in the science goals, that is calibrated to the desired accuracy and free of calibration or imaging artifacts."

i.e. ALMA performs science-goal-oriented service data analysis

ALMA QA happens on 4 levels:

QA0: near-real time verification of weather and hardware issues carried out on each SB execution (execution block, EB) immediately after the observation.

QA1: verification of longer-term observatory health issues like absolute pointing and flux calibration.

QA2: offline calibration and imaging (using *CASA*) of a completely observed MOUS. Performed by expert analysts distributed at the JAO and the ARCs with the help of a semi-automatic *CASA* pipeline. Results are archived and given to the PI.

QA3: (optional) PIs may request rereduction, problem fixes, possibly reobservation

ALMA QA

CASA (Common Astronomy Software Applications)
is the designated data analysis package for ALMA and the JVLA.

Used for all offline processing of ALMA data.

CASA is developed by NRAO, ESO, and NAOJ (under NRAO management);
for details see http://casa.nrao.edu
and e.g., Petry et al., 2012, "Analysing ALMA data with CASA", ADASS XXI, ASP conf., 461, 849

Latest release is CASA 4.3.1 .

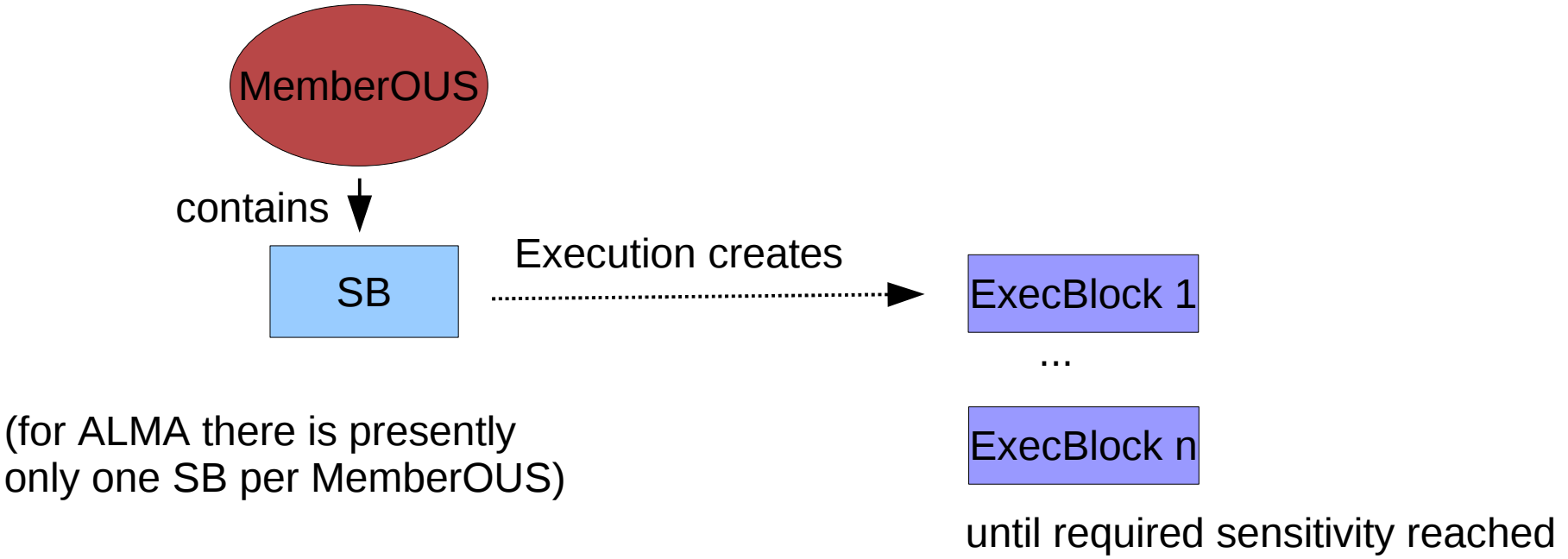
The ALMA pipeline is an optional add-on of CASA.

Latest release with ALMA pipeline is CASA 4.2.2 .

ALMA QA2

Science-goal-oriented service data analysis

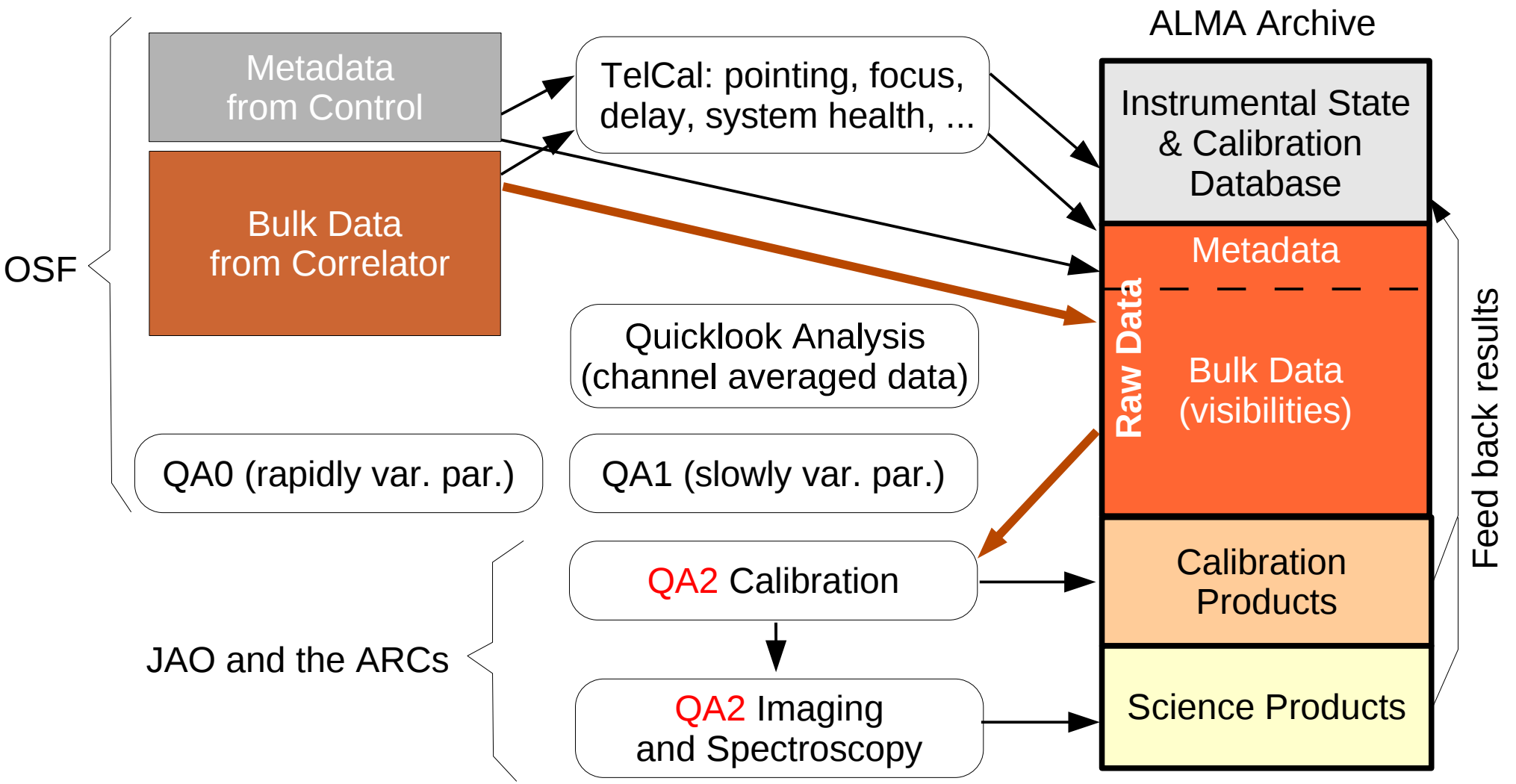
- ***PI defines science goals in proposal using the Observing Tool (OT)***
  ⇒ Scheduling Blocks (“SBs”) in Observation Unit Sets ("OUSs")
- ***SB = prototype of an atomic (ca. 0.5 h) observation to reach a science goal***
- ***Exec Block = actual execution of an SB (may need several to reach science goal)***



- ***QA2 verifies that the n ExecBlocks combined reach the memberOUS science goal***

EUROPEAN ARC
ALMA Regional Centre

ALMA Data Flow

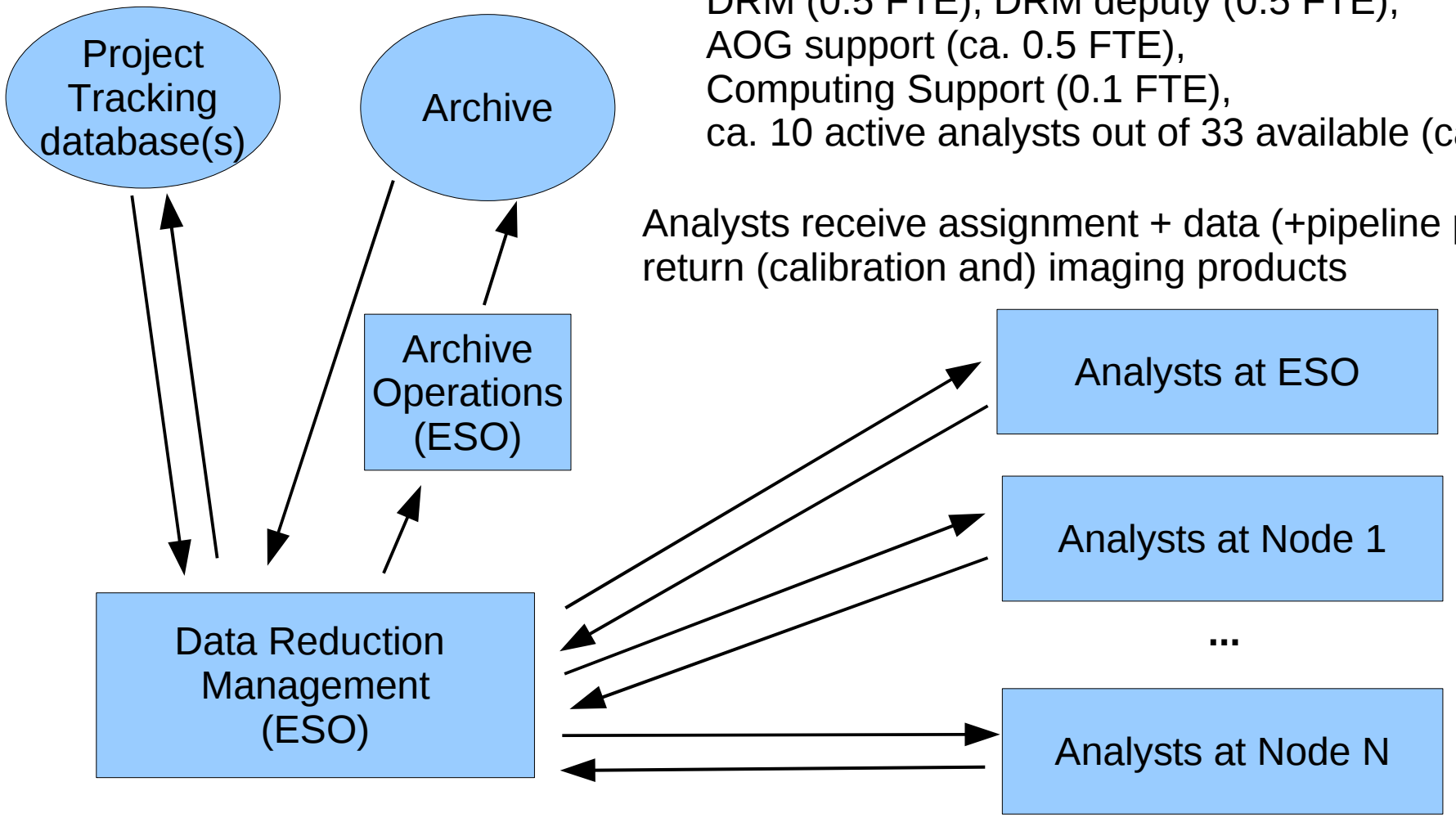


EUROPEAN ARC
ALMA Regional Centre

ALMA QA2 at the EU ARC

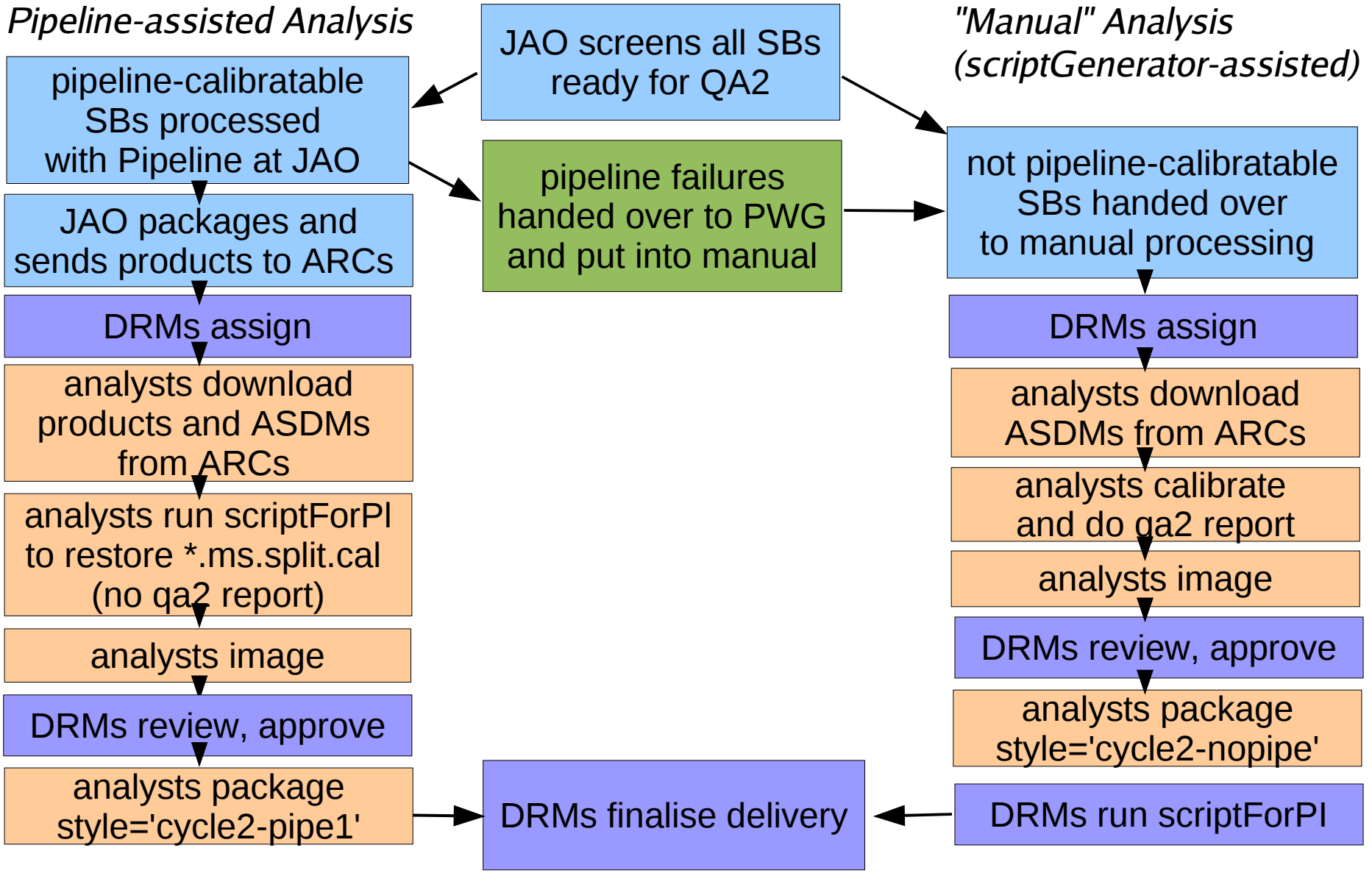
Personnel:
DRM (0.5 FTE), DRM deputy (0.5 FTE),
AOG support (ca. 0.5 FTE),
Computing Support (0.1 FTE),
ca. 10 active analysts out of 33 available (ca. 5 FTE)

Analysts receive assignment + data (+pipeline products),
return (calibration and) imaging products



EUROPEAN ARC
ALMA Regional Centre

QA2 Workflow

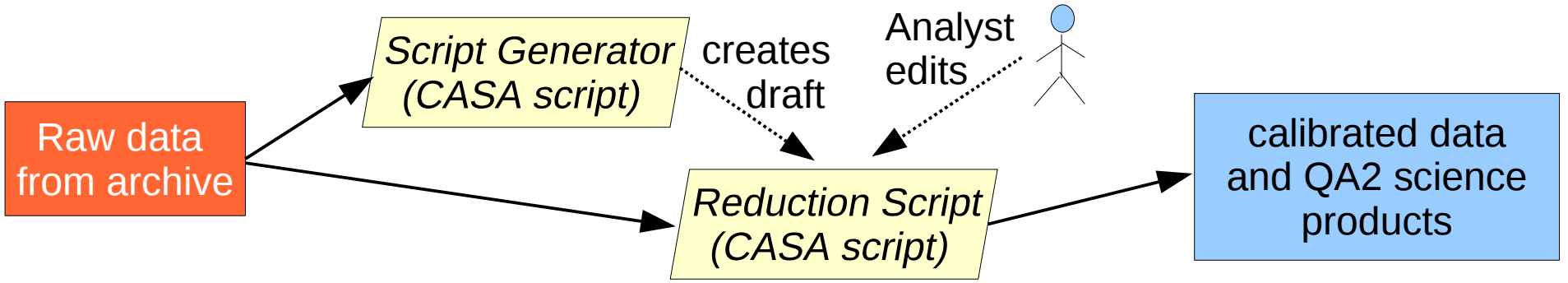


Script Generator assisted analysis

- **Before a fully automated Pipeline can be commissioned, *the manual data analysis has to be fully understood!***

- **A large team of ALMA scientists works together to develop the *best practices* to perform a robust standard calibration of ALMA data.**

- **These best practices are automated using a system of Python scripts called the "*Script Generator*"**



- **The Script Generator *evaluates* a raw dataset (imported MS) and *writes a draft for a CASA data reduction script* (one each for calibration and imaging)**

- **The data analyst then *edits the draft scripts* where necessary and *runs them* (typically in small steps) *iterating* until confindent that best calibration achieved**

EUROPEAN ARC
ALMA Regional Centre

EU ARC QA2 Statistics

Overview: Status 10 April 2015

```
Total processed (including QA2_FAIL) since January 2013:
   307 (pass) +  88 (fail) = 395  SBs
   556 (pass) + 174 (fail) = 730  EBs

Total delivered: 307 SBs (556 EBs)
    Manual cal.: 212 SBs
  Pipeline cal.:  95 SBs

Cycle 2 Net analysis time  median:          6 working days
Time from observation to delivery median: 34 working days
(includes delays due to work at JAO)

Analysis effort 2013+2014 amounts to ca. 2000 working days
        (assuming 75% efficiency, this is ca. 7.5 FTE years)

Since Pipeline became available in October, ca. 75% of the
calibration is done by pipeline
```

QA2 Science Data Products

Data Products Basics

- We don't deliver Calibrated Data, we deliver *Raw Data, Cal. Scripts and Tables*.

  Users need to run CASA to generate the Calibrated MeasurementSets.

  The resulting calibrated data is science-ready.

- Furthermore, we deliver *Imaging Products*
  (in Early Science provided on a best effort basis, *not necessarily science-ready*)

  a) for Line Observations:
  - continuum-subtracted (where needed) image cubes at the requested resolution
  - a continuum image for all line-free channels (where possible)

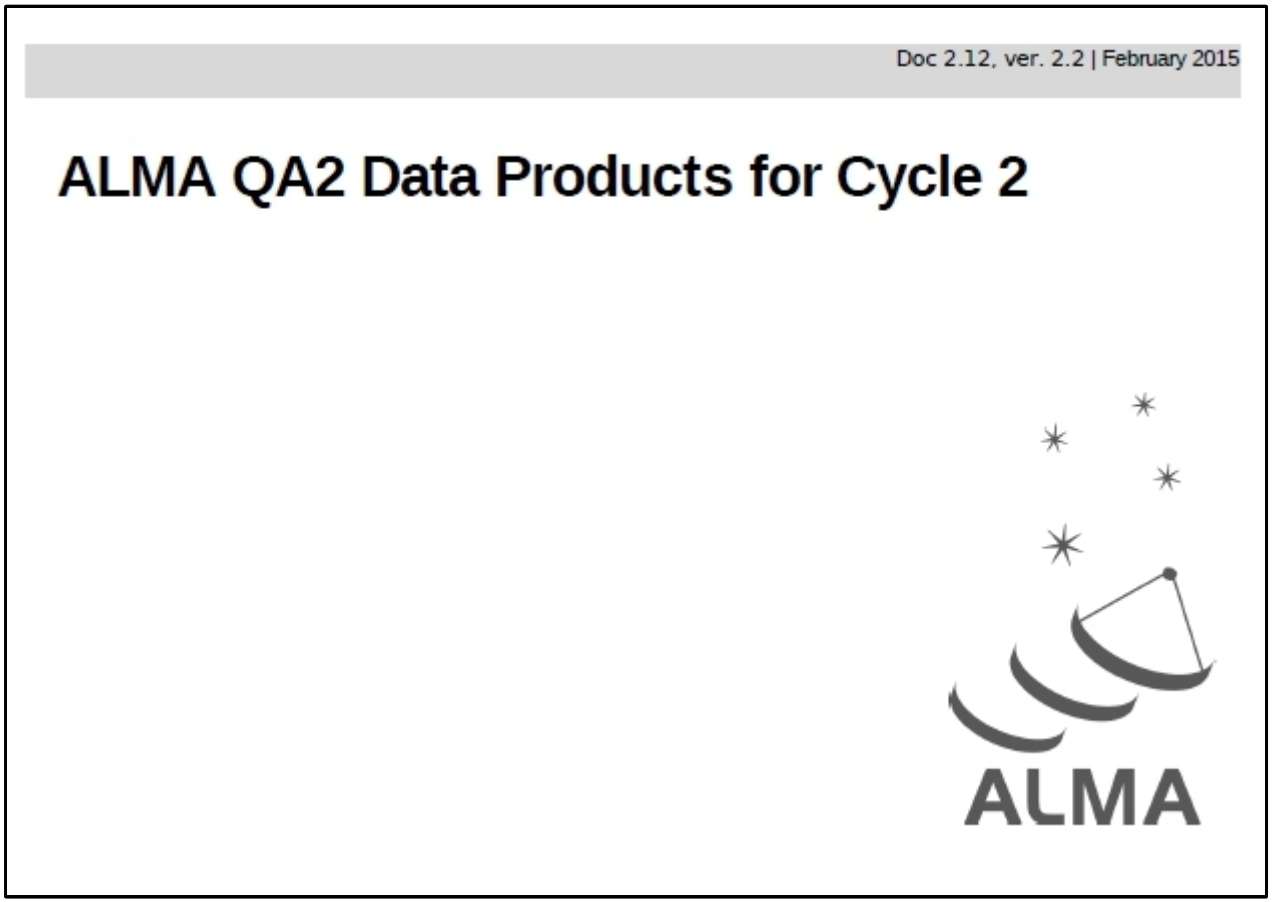
  b) for Continuum Observations:
  - continuum image combining all SPWs

  Main purpose: to measure achieved RMS+beam and to provide images for archive

QA2 Science Data Products

Details

Described in



... see www.almascience.org (the ALMA Science Portal)

QA2 Science Data Products - Details

Delivery contains **2 parts**

a) The Products Package (a tar ball containing a mix of files, typically a few 100 MB)

download this first

b) The RAW data (ASDM format, typically several 10 GB in Cycle 2)

only needed if you would like to work with the visibility data,
e.g. for improving details of the calibration and the imaging

QA2 Science Data Products - Details

When untarred, the *Product Package* **standard directory structure** contains

```
|-- project_id/
| |-- sg_ouss_id/
| | |-- group_ouss_id/
| | | |-- member_ouss_id/
| | | | |-- README ..................... important summary of the contents
| | | | |-- product/ ..................... all the imaging products
| | | | |-- calibration/ ..... calibration and flagging tables
| | | | |-- qa/ ..... diagnostic plots generated during calibration
| | | | |-- script/ ..... the scripts necessary to regenerate the cal. MS
| | | | |-- log/ ..... CASA log files from QA2 calibration and imaging
| | | | |-- raw/ ..... only present when you have unpacked the raw data
```

QA2 Science Data Products - Details

Imaging products

- 4D Cubes: RA, Dec (from ca. 256x256 up to a few 1000 x a few 1000 pixels),
  Frequency (up to a few 1000 channels),
  Stokes (up to 4 channels)
  (Freq and Stokes axis can be degenerate)

  for line and continuum images of the science target
  and the bandpass and phase calibrators

  stored as FITS files following the latest FITS standard (3.0)

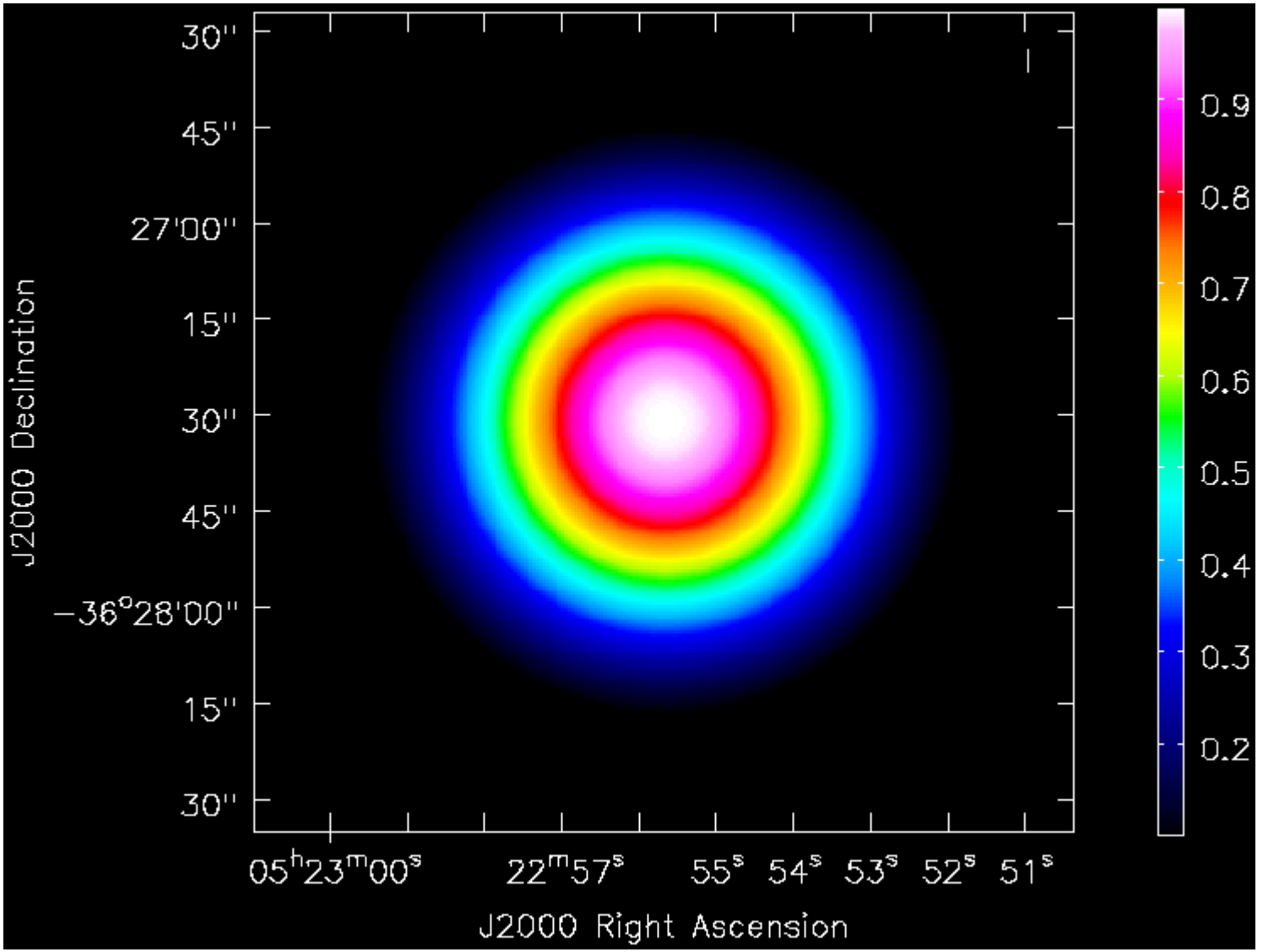
- Science target images are corrected for the primary beam
  (sensitivity variation across the FOV)
  The PB correction is included as a separate image

- The clean mask(s) for the science targets

QA2 Science Data Products - Details

Example of a **Primary Beam** image.

Multiplying the corrected image by the PB image restores the uncorrected image.



QA2 Science Data Products - Details

Data Reduction Scripts *(ASCII files: Python or XML)*

a) for pipeline-calibrated data
- the Python scripts needed to restore the calibrated MS or rerun the entire pipeline
- the pipeline processing request file (PPR)
- the flux equalisation script (if necessary) and the imaging script

b) for analyst-calibrated data
- CASA reduction scripts including: calibration scripts, flux equalization script (if necessary), and imaging script

In both cases, there is a master script "scriptForPI.py" which is the one the user needs to run to perform the generation of the calibrated visibilities.

QA documentation

a) for pipeline-calibrated data
- The *Pipeline Weblog* – a system of webpages containing all the diagnostic plots and other information generated by the pipeline.

b) for analyst-calibrated data
- QA Reports for all the imaged data (png images and pdf files).

QA2 Science Data Products - Details

Calibration and flagging tables

a) for pipeline-calibrated data
- Calibration tables (Tsys, WVR, Bandpass, Gain, Amplitude)
- Flagversions tables
- Calibrator fluxes (flux.csv)
- Pipeline metadata (*calapply.txt and *flagtemplate.txt)

b) for analyst-calibrated data
- Calibration tables (Tsys, WVR, Bandpass, Gain, Amplitude)
- Flagversions tables

These products together with the calibration scripts serve to recreate the calibrated visibilities (MS format) from the raw data. This also requires the installation of the right version of CASA.

QA2 Science Data Products - Details

QA2 Pass Criteria

A dataset passes QA2 if the achieved sensitivity (noise RMS) is less then

1.10 x requested RMS (Bands 3-6)

1.15 x requested RMS (Bands 7-8)

1.20 x requested RMS (Band 9)

and the angular resolution (synthesized beam area) is within

20% of the requested value

If the products still reach the science goal described in the proposal, these criteria can be relaxed to

RMS < 1.4 x requested RMS (Bands 3-6)

< 1.5 x requested RMS (Bands 7-8)

< 1.6 x requested RMS (Band 9)

and

synth. beam area within 50% of requested value.

QA2 Science Data Products - Details

Tolerances and Precision in the Cycle 2 imaging products

Angular Resolution

The pixels size is chosen to be 1/5 of the synth. beam major axis

Astrometric Accuracy

10% of the synth. beam diameter

Spectral Resolution

The requested resolution is always achieved.
Spectral axes are given in LSRK or BARY reference frame.
Note: depending on the obs. mode, the *true* spectral resolution may be = 2 x native channel width due to Hanning smoothing
Note 2: ALMA does not Doppler track during the observation, but Doppler tracking is performed during the offline analysis process.

Absolute Flux Accuracy

5% (Band 3 and 4), 10% (Bands 6 and 7), 20% (Bands 8 and 9)

Conclusions

- The ALMA QA2 system offers the ALMA PI with the valuable service of providing ***fully calibrated, science-ready data*** and (still on a best effort basis) ***comprehensive imaging products*** and the ***data reductions scripts*** to obtain them

- The imaging products serve two purposes:
  a) enabling the QA2 team to **judge whether the science goal was reached**
  b) giving the PI a **good starting point on the way to obtain the final images** and a valuable basis for archive researchers

- The QA criteria are stringent and so far the user community has been **enthusiastic about the achieved data quality**

- The QA2 process is labour intensive but **ALMA is keeping up with the data flow**.

- Since January, the **ALMA pipeline does ca. 75% of the calibration** work. All Imaging and the rest of the calibration work is done by analysts. Work is underway to also automate at least some of the imaging.

- The ALMA data products structure is prepared for upcoming new data features and will not change much over the next years.