Calibration, the Weblog, scriptForPI, and running the Pipeline

Luke Maud – ESO Garching - ALMA ARC Scientist

Before we begin

What is the Pipeline?

ALMA calibrate and image all user data

- Pre-pipeline:
  - People did the calibration and imaging with scripts, like you've learned
  - All ARCs and Arc-Nodes involved
- **Pipeline:**
  - Developed over the past decade or more
  - Now does the calibration and imaging along with heuristics, plots and useful information
  - **Must work for 1000s of datasets**
  - Pipeline are pythonic tasks/stages that are based on CASA core tasks and tools

Documentation and Guides

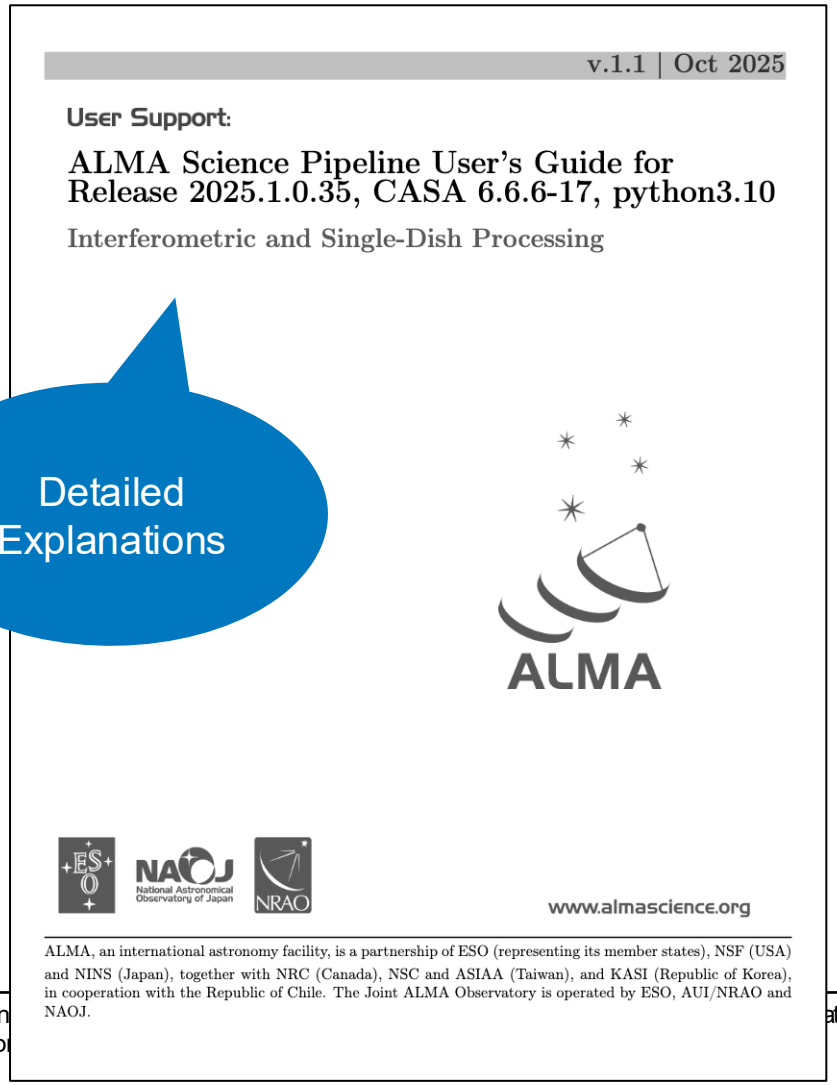
The focus for these sessions will be to use and utilize the ALMA Pipeline

Pipeline and CASA are both NA deliverables; they are responsible for documentation (with input from all Executives of ALMA, who provide developers and working group members)

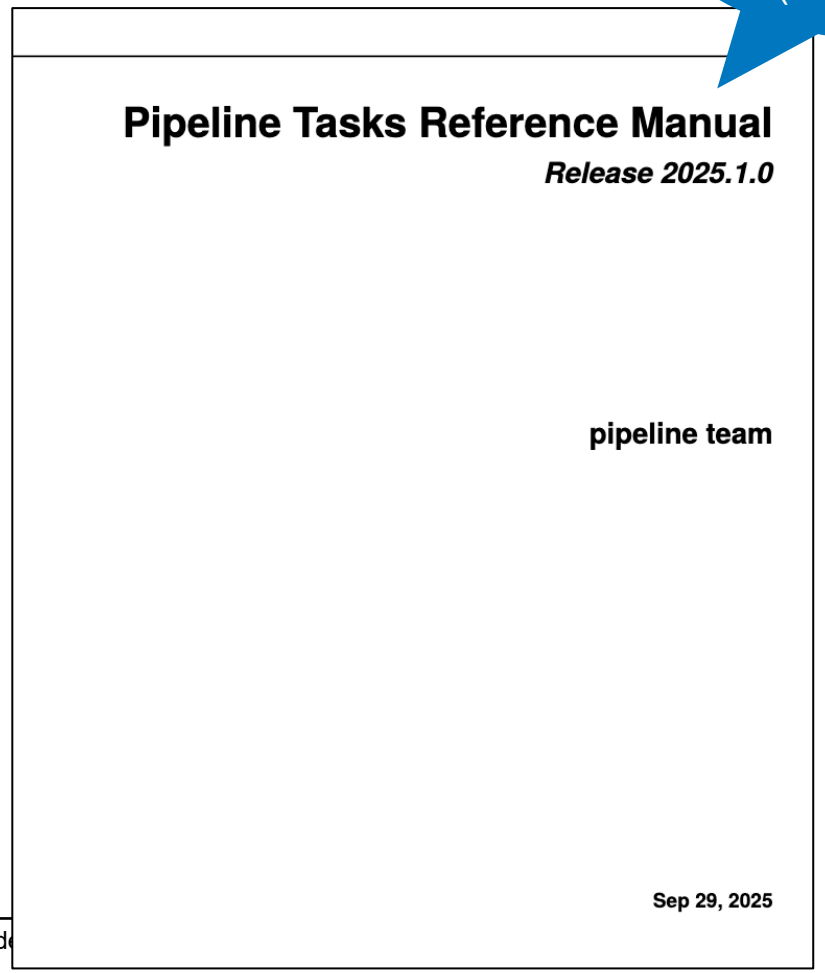
- Pipeline: https://almascience.eso.org/processing/science-pipeline
  - User Guide
  - User Task Reference Manual
  - Known Issues
  - Published Paper (The ALMA Interferometric Pipeline Heuristics – Dated 2023 version)
- CASA: https://casa.nrao.edu/casa_obtaining.shtml
  - Guides and Documentation

Documentation and Guides

The focus for these sessions will be to use and utilize the ALMA Pipeline



Detailed Explanations



Tasks, Parameters, brief overview (doc strings)

What do you want to get out of this tutorial?

It's possible we cover all of it, but maybe not…

Join at menti.com | use code 8707 2812

We'll use Mentimeter – there are **three choices per person**:

- To understand the weblog
- How to **restore** calibrated data
- How to do/re-do continuum subtraction (in Pipeline)
- How to use the pipeline self-calibration process
- How to totally **reprocess** data in the latest Pipeline
- How to use pipeline to re-image

Getting data to work with as a user

…this data looks interesting

As a PI or as an Archive user these are the steps:

- Go via the Archive and search out some data
  - You can download some weblogs without the data – *see already if there is any detection*
- Select what you want to download
  - On the download page **make sure you ticket the tarbal for RAW data**
- Run the downloaded script
  - Follow the prompts to untar everything
- Investigate the folder structure
- Check the QA and/or weblogs locally
- Start “processing”

Possible the data for these tutorials

We can use three datasets, some partially processed

All ACA data to run fast

- Datasets are:
  - 2013.1.00524.S – High Frequency (Nuclear Regions of Nearby LIRGs) – CGCG049
  - 2022.1.00974.S – Massive Star Forming Cores with Masers – G183.348
  - 2024.1.00408.S – Disk Population of Herbig Ae/Be – G304.664-00.966
    PI Miguel Vioque (*kindly allowed us to use these data* – will be published next year)
- I have prepared three tarbals for many different activities later:
  - Allow you to do a full Pipeline run
  - Allow you to do some step-by-step calibration stages
  - Allow you to do a findcont
  - Allow you to test a selfcal process
  - Allow you to do imaging

We'll come back and get data when we need it

Let’s all login and setup a work area

Setup as used for the Leiden School

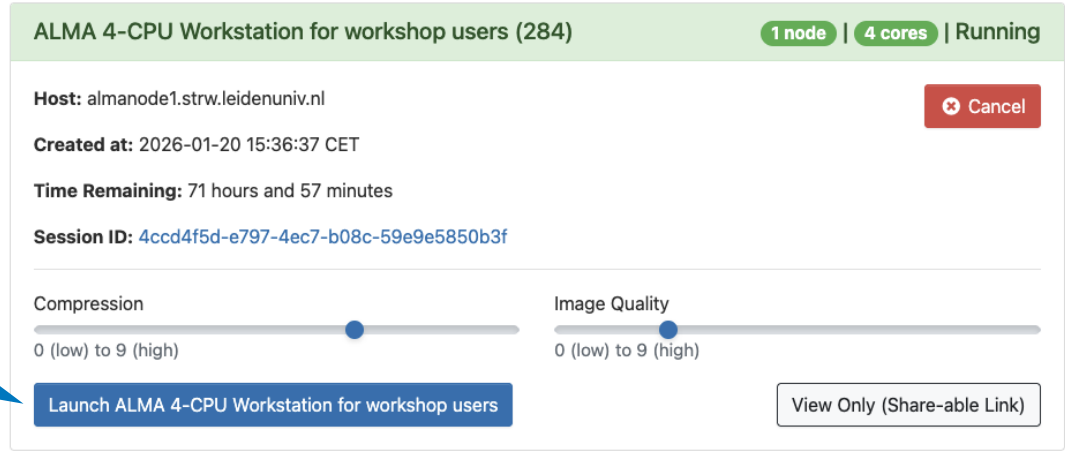
Setup a work area

Please follow exactly to login

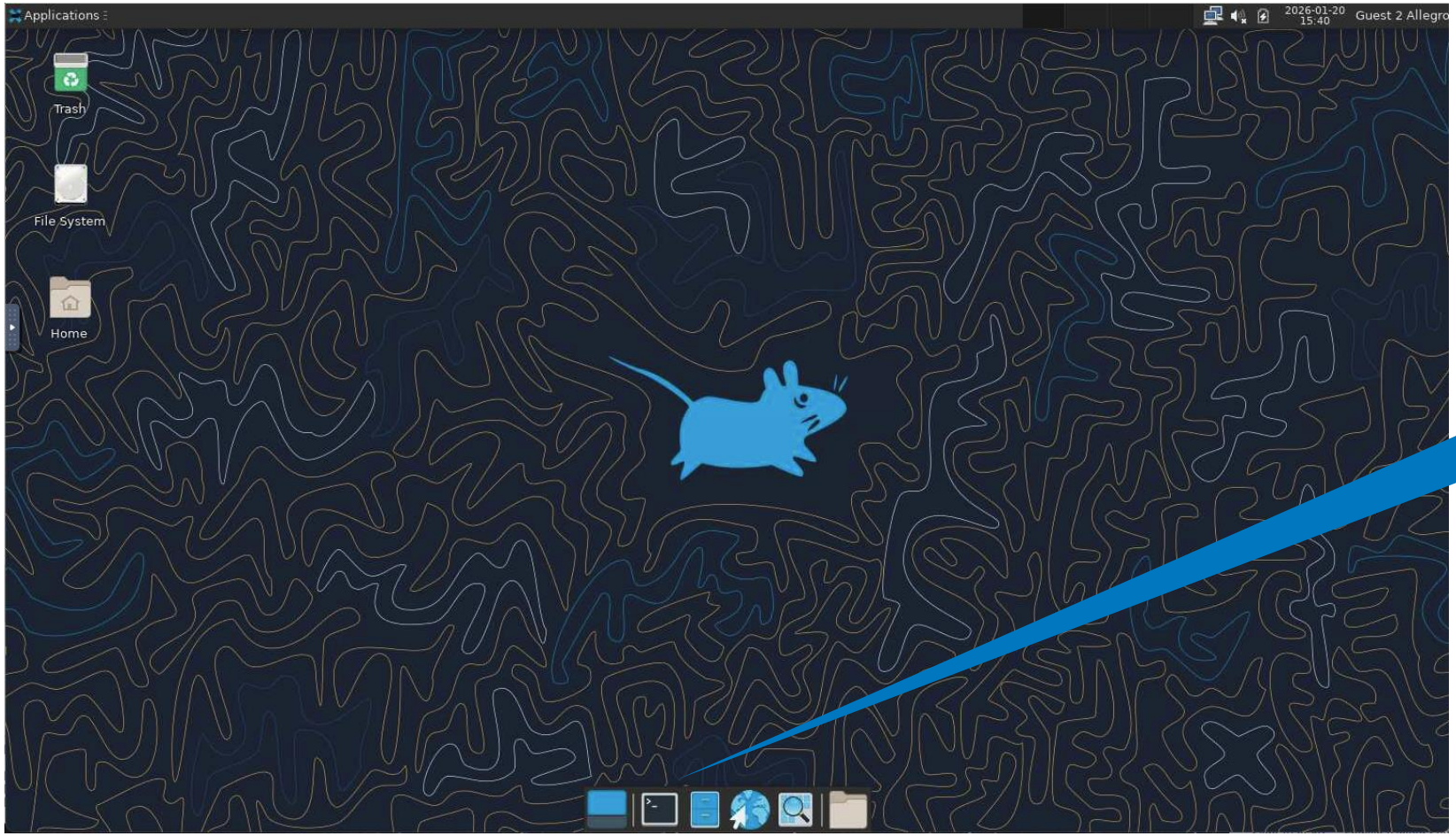
- Allegro instructions
  - Go to this weblink: https://almaportal.strw.leidenuniv.nl/
  - Enter your username (allegroguest<X>) and password (XXX)
- This should not be your first login:
  - if it is, use a 2FA app to scan the code here: https://home.strw.leidenuniv.nl/~alma/workshop/allegroguest<x>.png
  - else, use you app to enter the 2FA code
- Now use the **Interactive Apps** to launch a desktop

<X> is your assigned user number

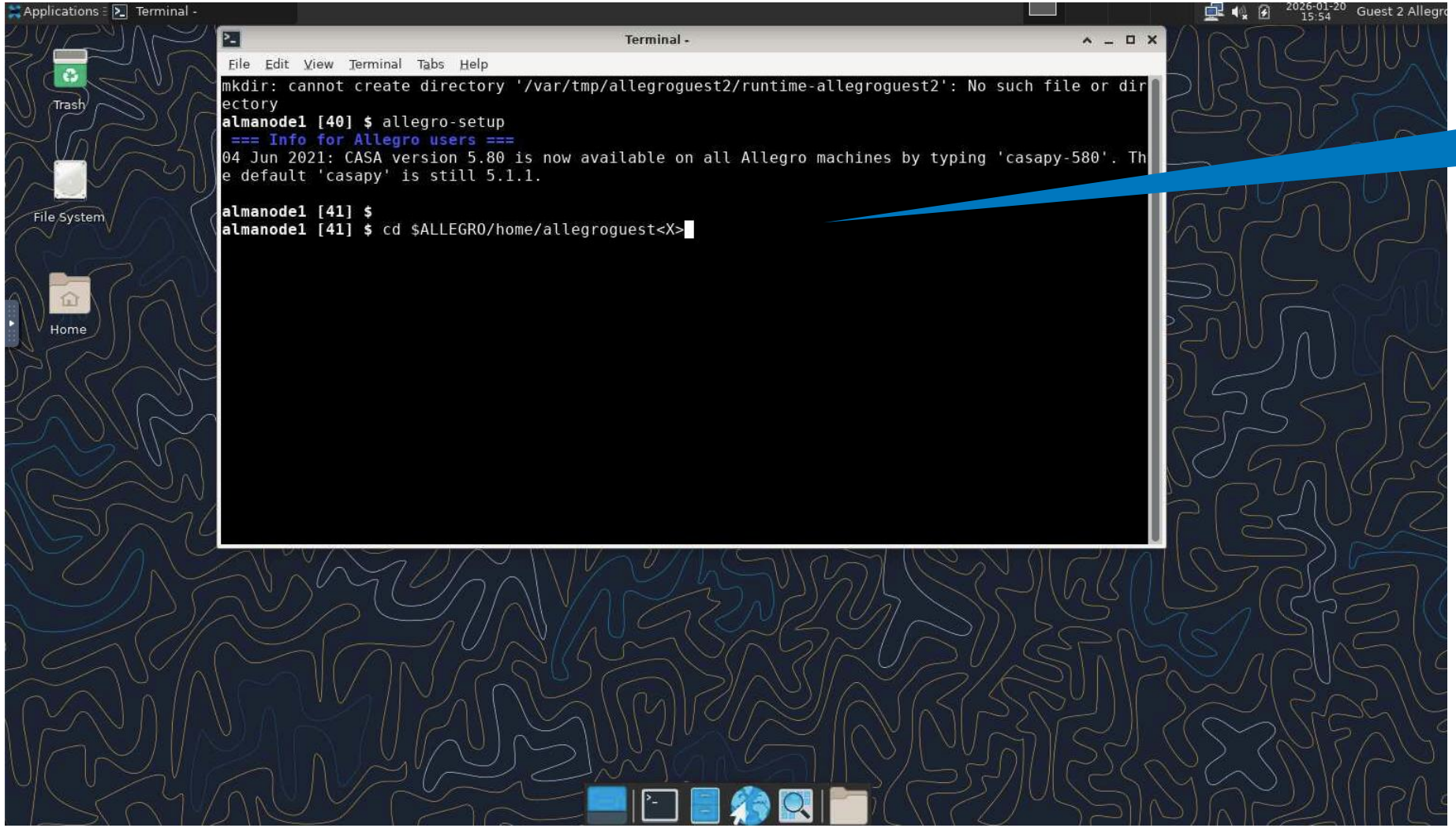
Click to launch the VNC



Are you on a desktop session?

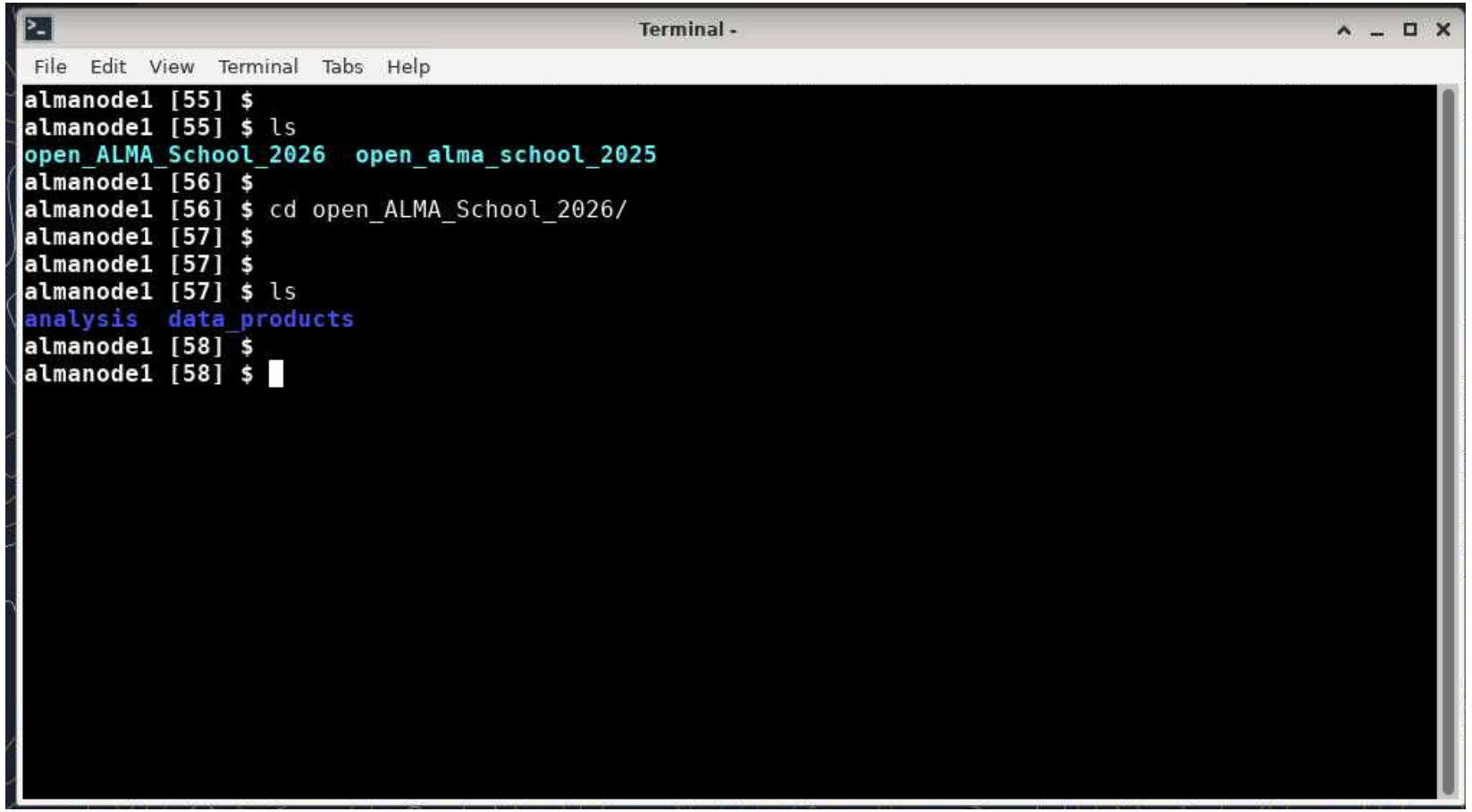


Setup commands

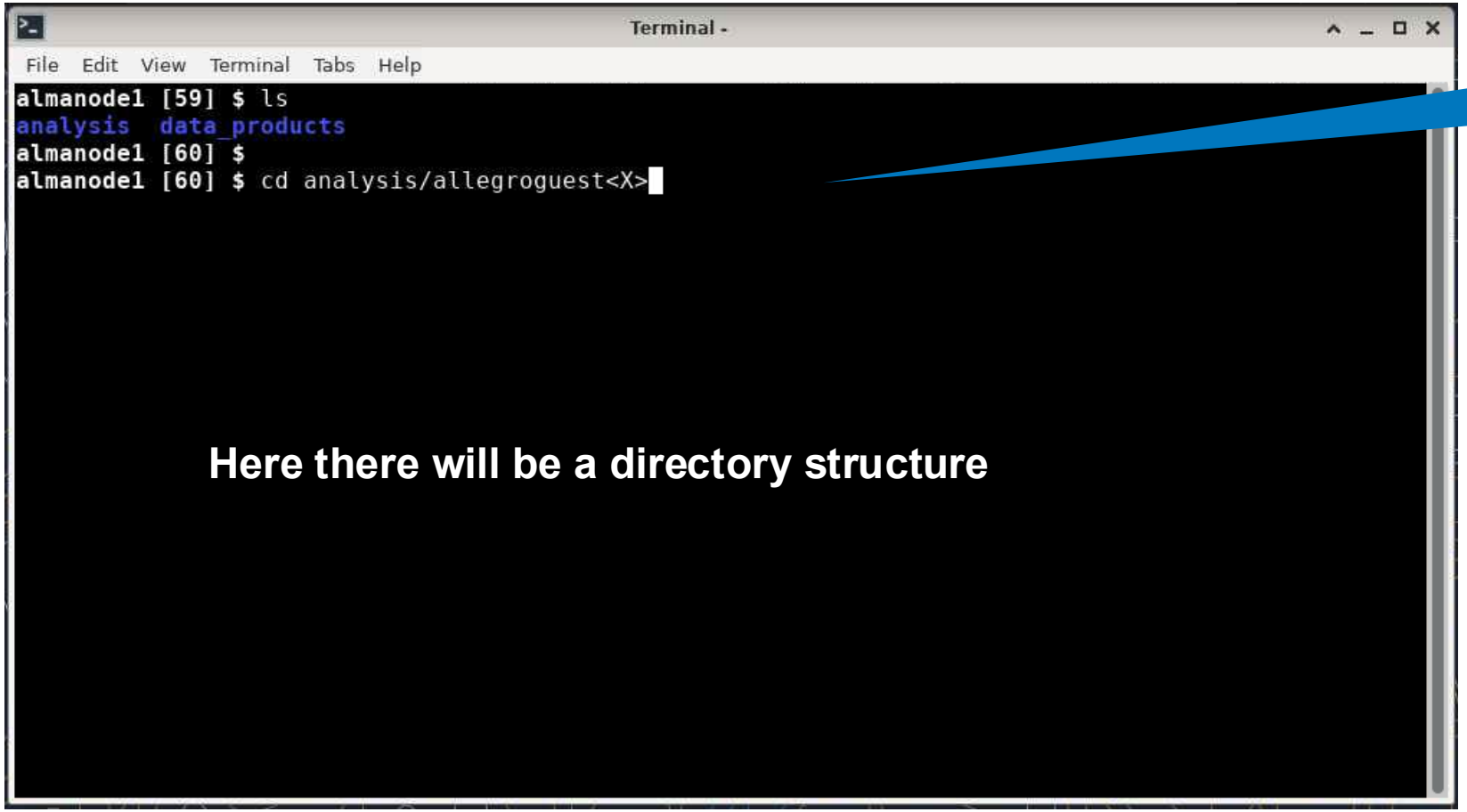


<X> is your assigned user number

Setup commands



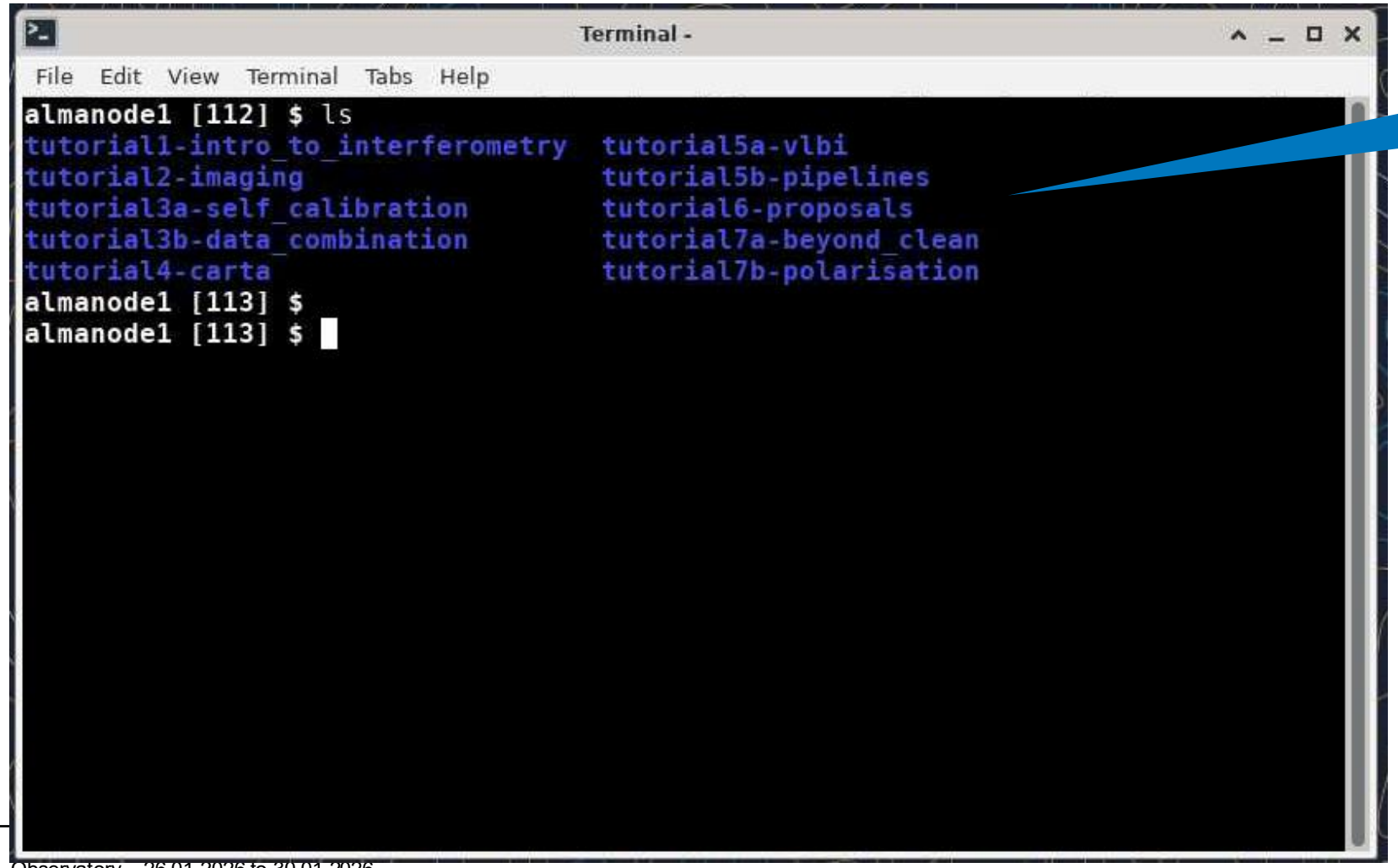
Setup commands



<X> is your assigned user number

Setup the working directories and copy data

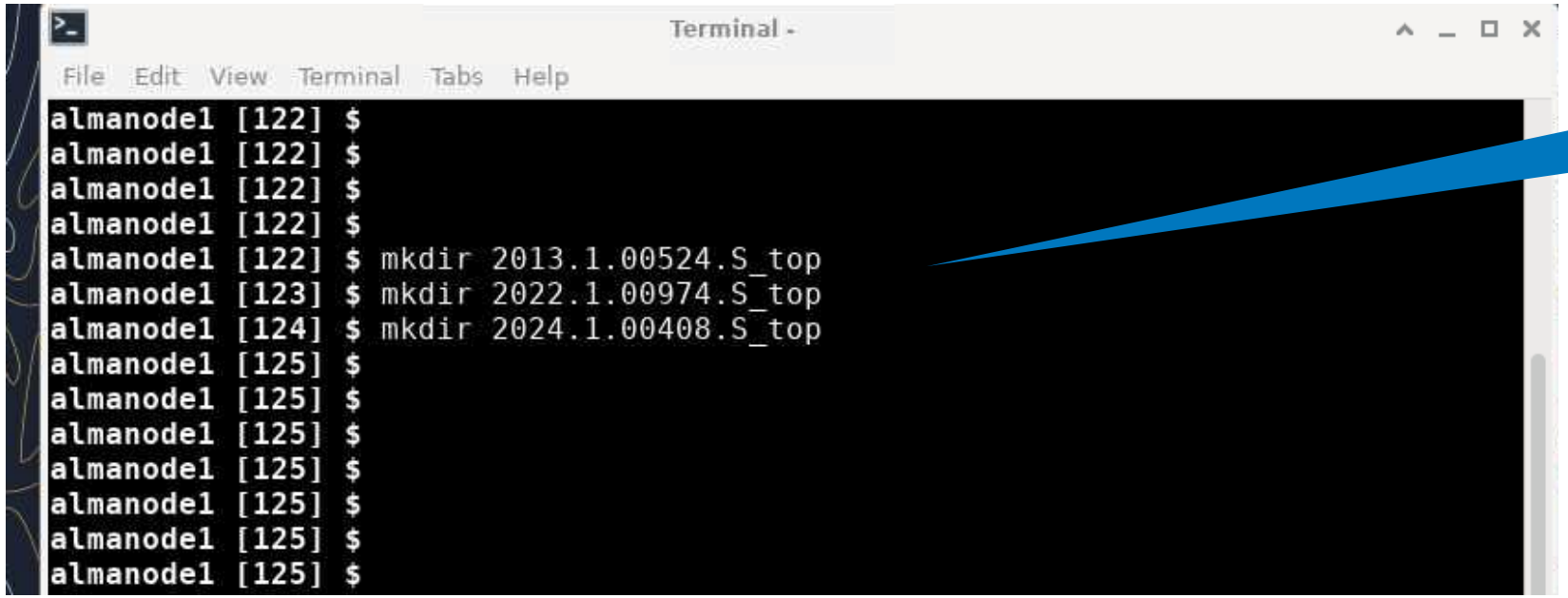
Please follow exactly or your directories will be messy and hard to navigate



These will be the top level directories

Setup the working directories and copy data

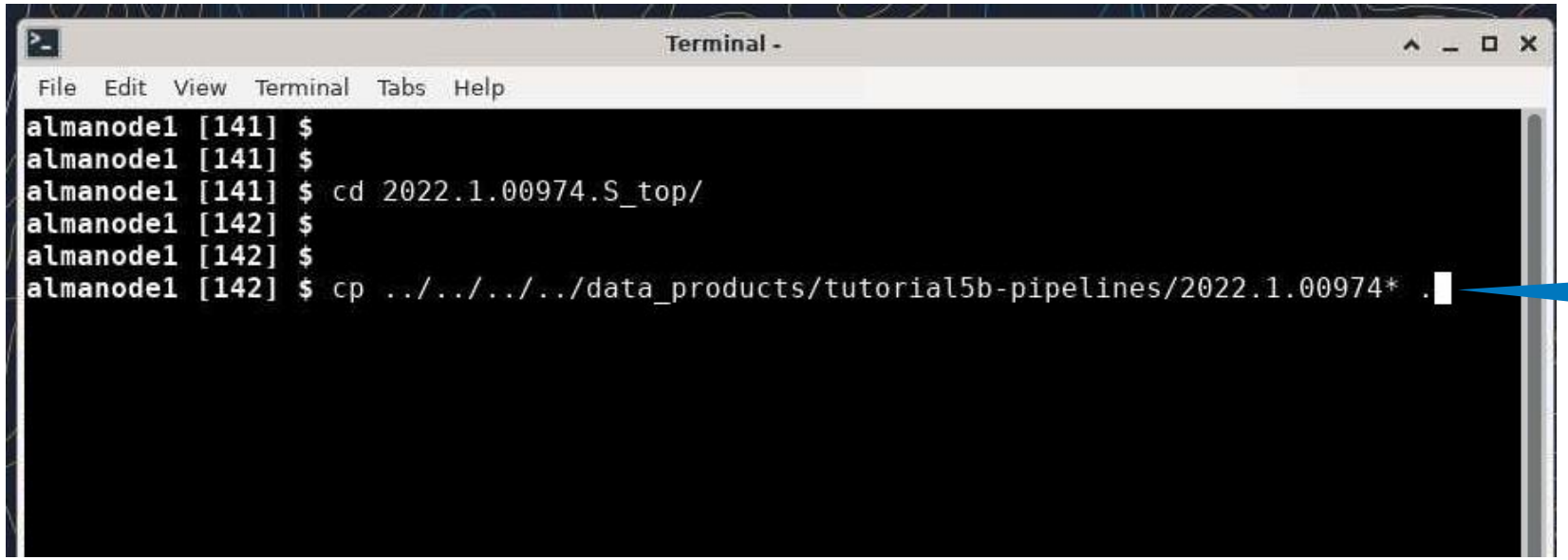
Please follow exactly or your directories will be messy and hard to navigate



Make these to transfer data into

Setup the working directories and copy data

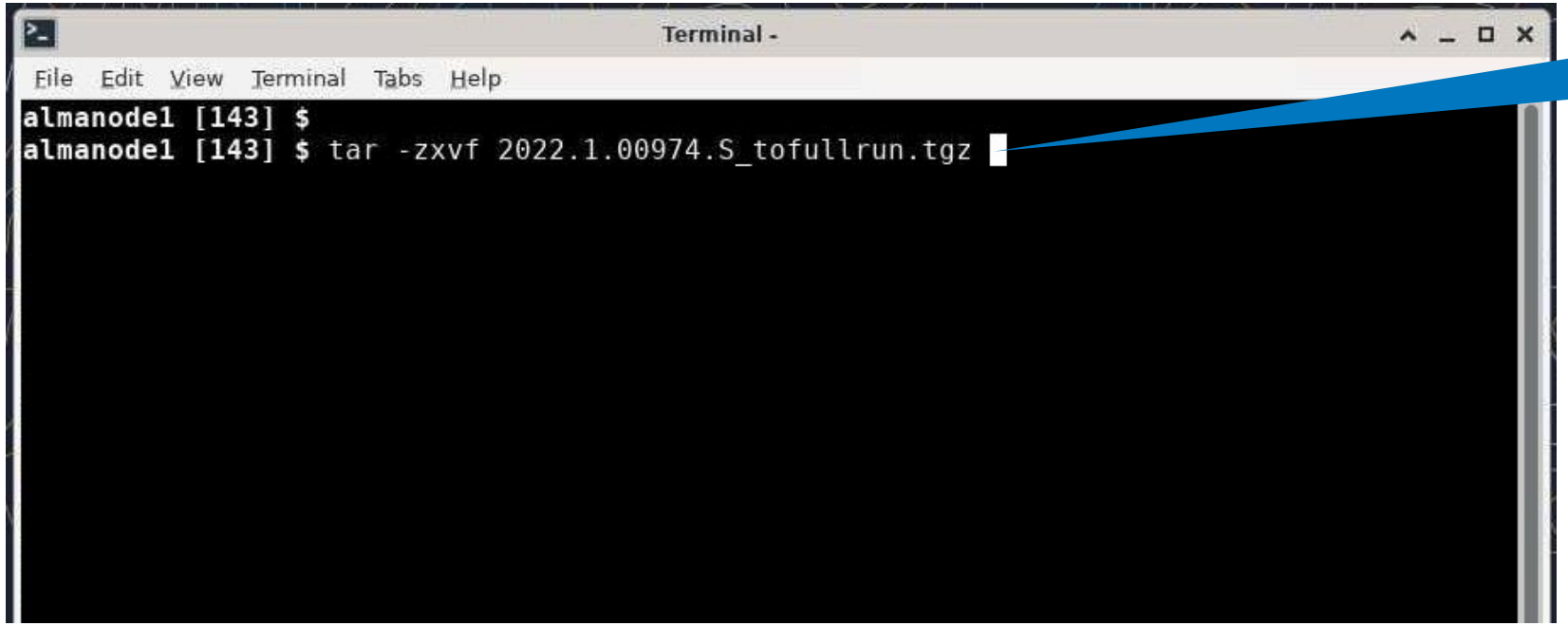
Please follow exactly or your directories will be messy and hard to navigate



Copy the relevant tarbals

Setup the working directories and copy data

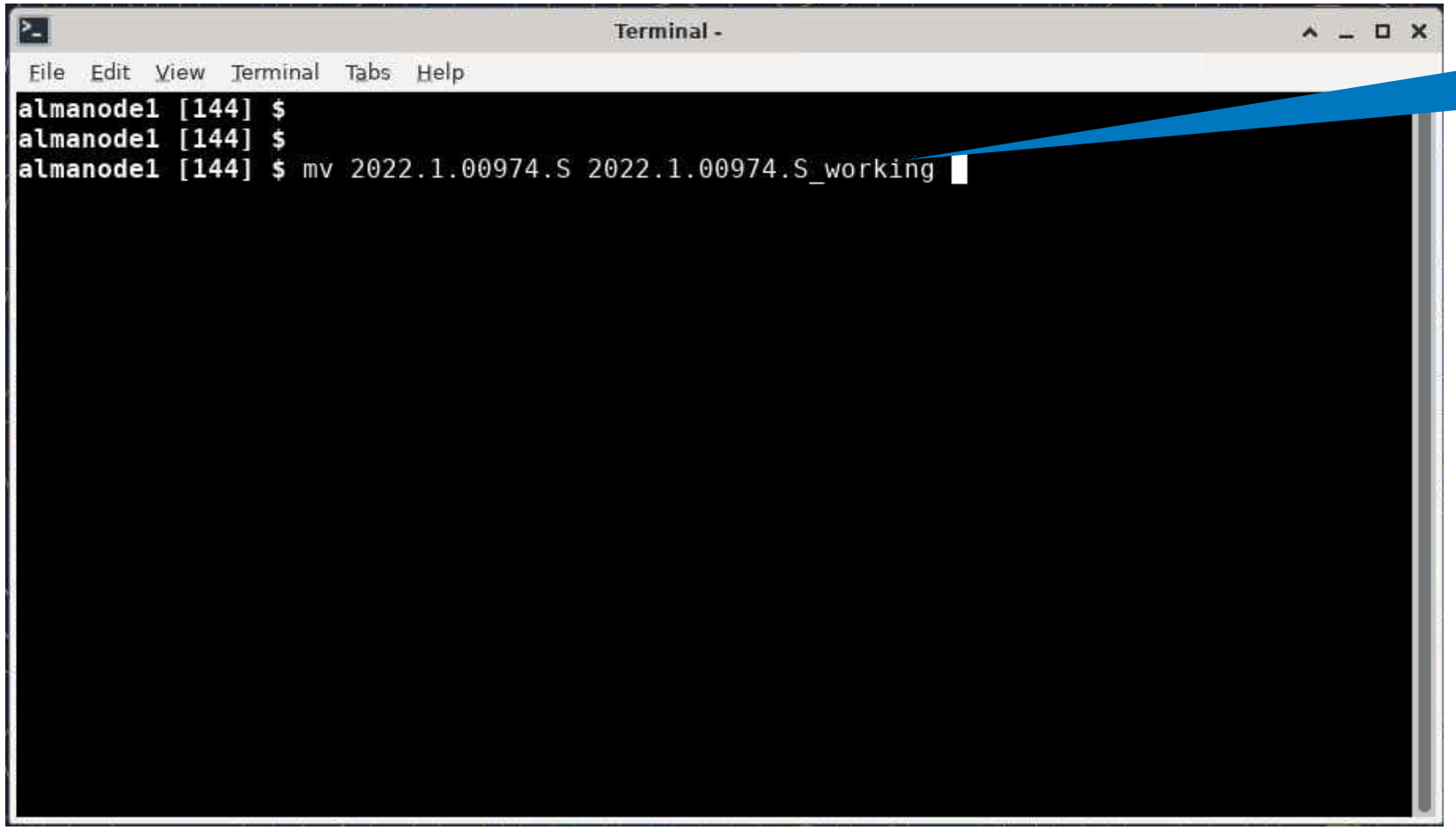
Please follow exactly or your directories will be messy and hard to navigate



For now only untar this

Setup the working directories and copy data

Please follow exactly or your directories will be messy and hard to navigate



rename

Setup on your own machine

Get data from archive

Follow relavant archive tutorials

- **Download and unpack all data as shown in other tutorials**
  - Follow the helper script that will do all of this
  - All ALMA data follow the same data structure

So, now let’s have a quick overview of the project

File structure

If you are using your selected Archive data file names will be different

Use your own Archive Download or the example data linked below

- Move down the entire directory tree
  - This is “project code” level,
  - Into the “science goal”
  - Then the “group observing unit set **GOUS**”
  - Finally, “member observing unit set **MOUS**”
- The tree depends on what the original PI (or user) aimed to do with ALMA
  - Can be different bands, different array configurations, difference sources all mixed into various parts

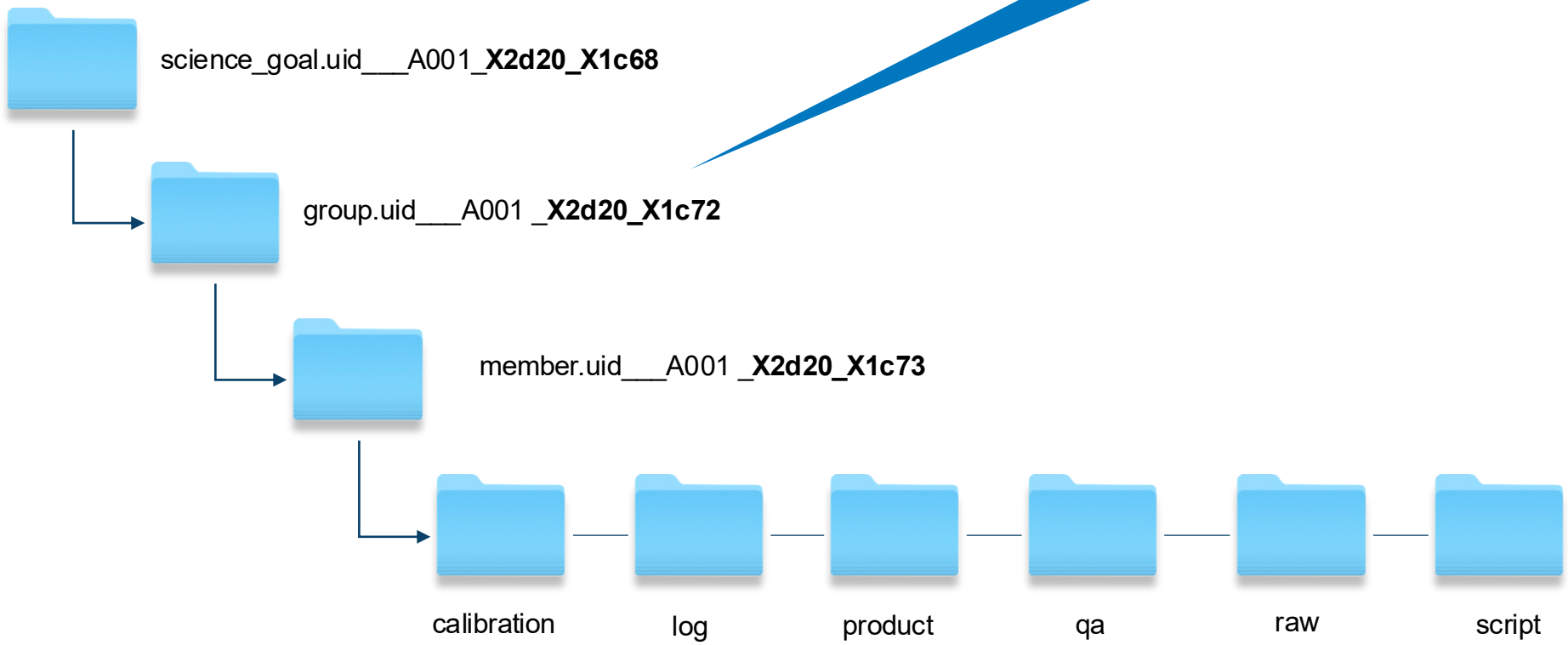
```
almanode1 [116] $
almanode1 [116] $ cd 2022.1.00974.S_working/
almanode1 [117] $
almanode1 [117] $ ls
science_goal.uid___A001_X2d20_X1c68
almanode1 [118] $ cd science_goal.uid___A001_X2d20_X1c68/
almanode1 [119] $
almanode1 [119] $ ls
group.uid___A001_X2d20_X1c72
almanode1 [120] $
almanode1 [120] $
almanode1 [120] $ cd group.uid___A001_X2d20_X1c72/
almanode1 [121] $
almanode1 [121] $ ls
member.uid___A001_X2d20_X1c73
almanode1 [122] $
almanode1 [122] $ cd member.uid___A001_X2d20_X1c73/
almanode1 [123] $
almanode1 [123] $
```

NOTE: *Older ALMA data may not have used the ALMA pipeline but manual calibration scripts*

File structure

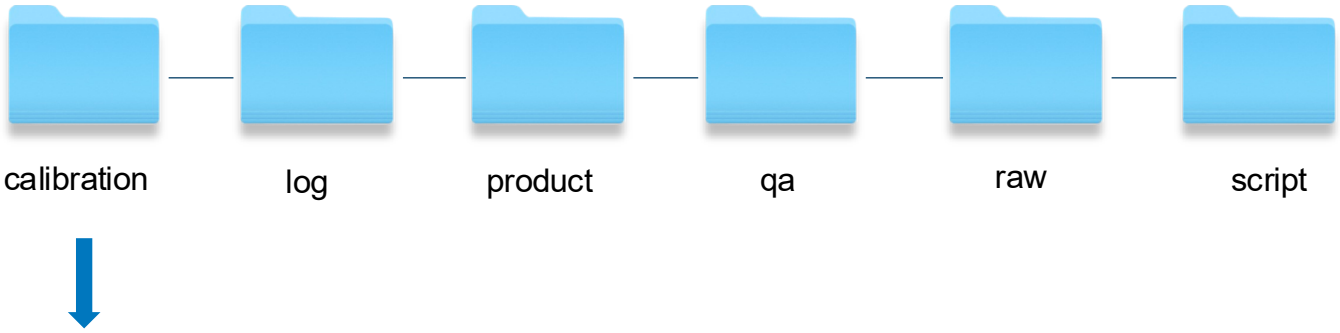
Use your own Archive Download or the example data linked below

If you are using your selected Archive data file names will be different



File structure

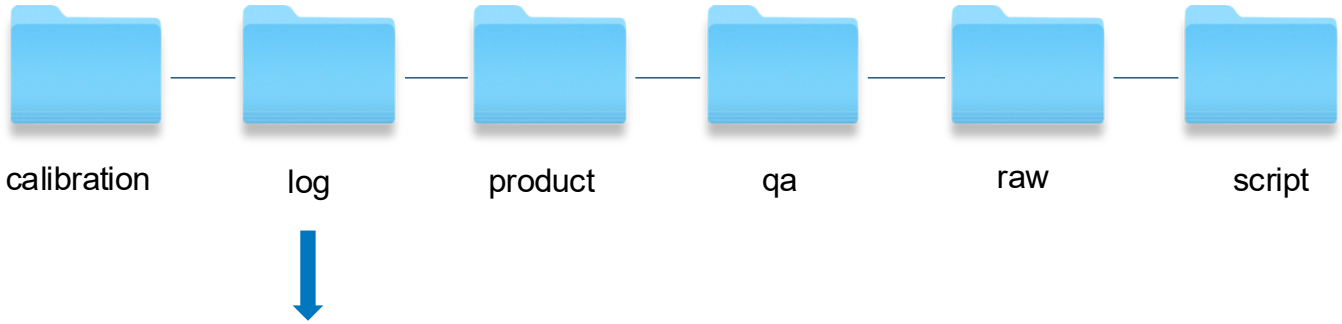
Use your own Archive Download or the example data linked below



- All calibration files done by ALMA
  - All as tarballs
  - Will be automatically untarred and moved in a **restore** process for Pipeline, or a re-run for old manually calibrated data
  - If run was with Pipeline there is a “calapply.txt” showing equivalent manual-like commands for reference only

File structure

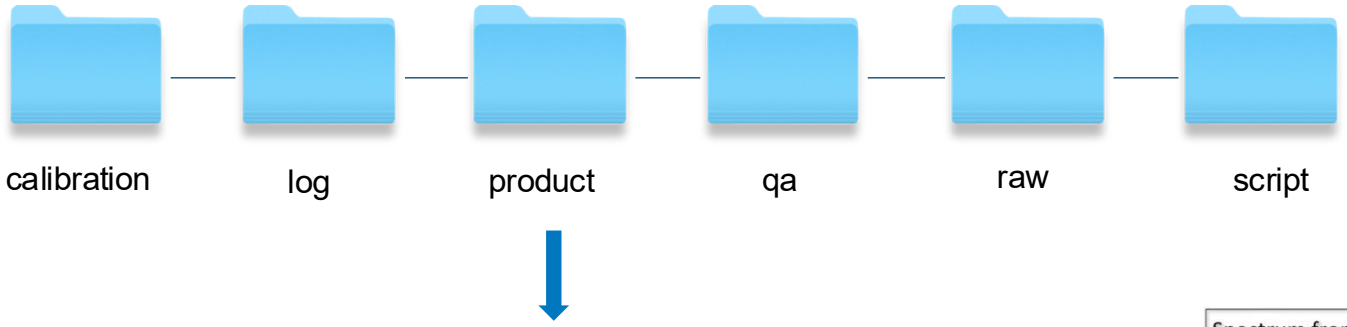
Use your own Archive Download or the example data linked below



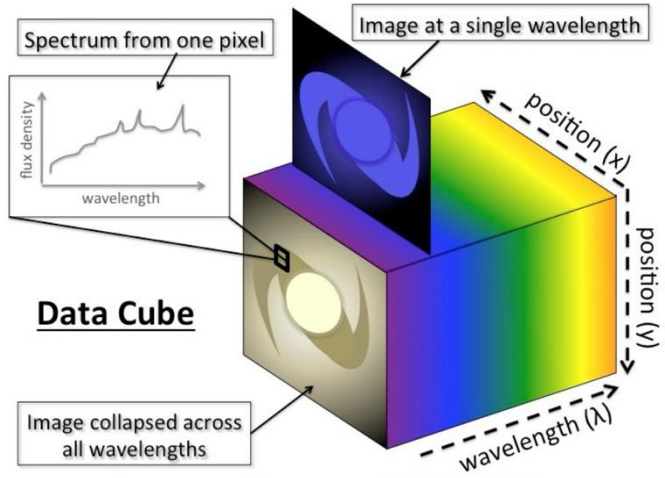
- Any logs made of the run
  - Might be calibration and imaging with ALMA pipeline in one log
  - Separate calibration and imaging
  - Other log files for manual calibrated data

File structure

Use your own Archive Download or the example data linked below

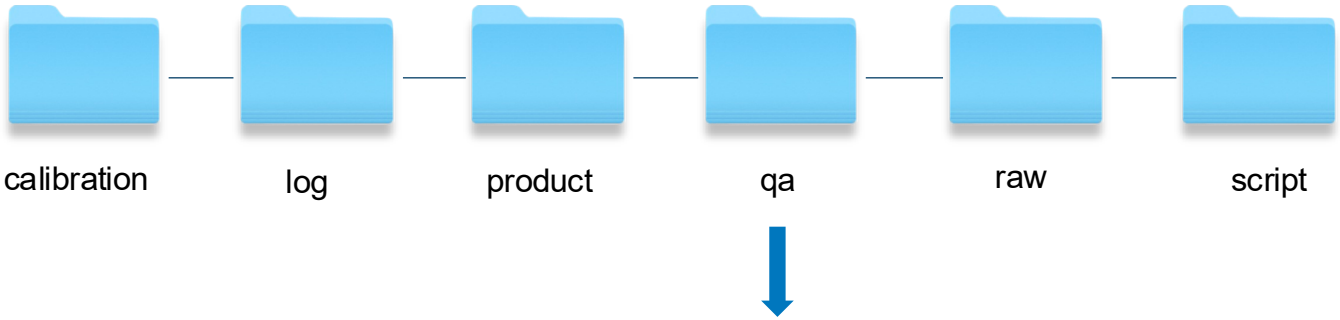


- All **image** products
  - FITS format images for the MOUS
  - Various Targets, SpWs
  - Types are MFS, CONT, CUBE (repBW)
  - Generally primary beam corrected and masks included
  - If Pipeline with selfcal, sometimes selfcal images



File structure

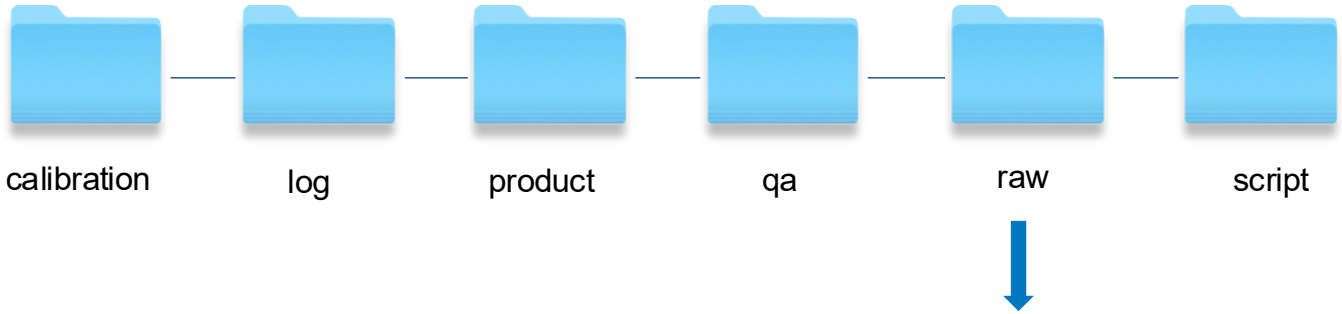
Use your own Archive Download or the example data linked below



- Reports from **quality assessment**
  - Newest data should include QA0 report
  - Tarball of the **weblog** from Pipeline processing
  - PNG or PDF for manually processed data

File structure

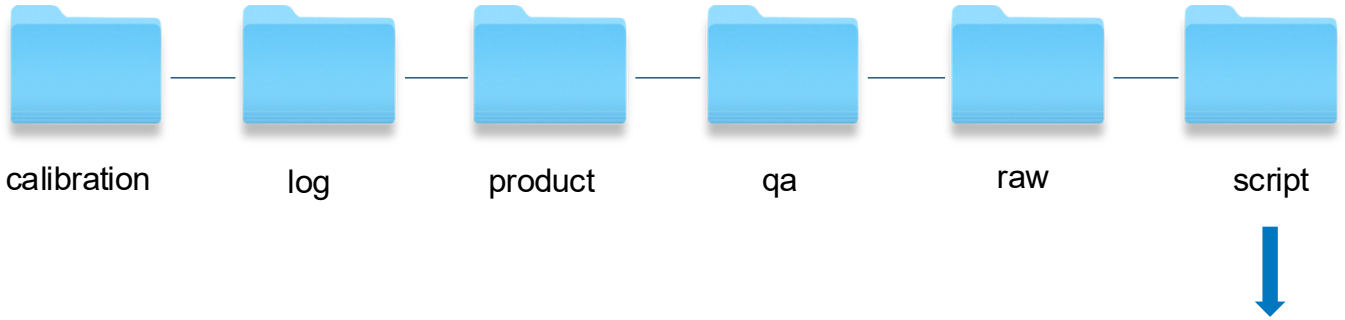
Use your own Archive Download or the example data linked below



- Data
  - The ASDMs (ALMA Science Data Model) are the raw data needed to be converted by CASA into measurement sets for calibration and imaging
  - These are **required** if you want to do any type of reprocessing, imaging, restore etc

File structure

Use your own Archive Download or the example data linked below



- Scripts
  - **scriptForPI.py** is a wrapper to automatically do everything to get your data in a state to image
  - **pipescirpt.py** and **piperestorescript.py** are Pipeline related scripts for **full run** and **restore** respectively
  - XML files are ancillary from pipeline runs by ALMA
  - Other .py could be from manual calibration or imaging

What is this weblog?

Overview of Quality Assurance

The ALMA project guarantee the quality of the data – so have many QA steps

- QA0:
  - Check if data are recorded correctly and all source/fields are included
  - Checks occur at the telescope in Chile
- **QA2:**
  - Data are calibrated and imaged
  - Must meet the requested **Beam Size** and **Noise RMS**
  - Up to October 2023 all data calibrated have always been checked by a human – data analyst (DA), and the data reduction manager (DRM)

Weblog

Interface to the whole calibration and imaging process

- In the recent years the majority (97% in the last 2 cycles) of ALMA data are calibrated and imaged bt the ALMA pipeline
- The ALMA pipeline is **primarily** run at **JAO** (Joint ALMA Observatory) in Chile, but also the ALMA Regional Centers (ARCs) in Europe, North America, East Asia
- A weblog is produced as part of the Pipeline processing
  - Serves as the **interface** for QA, for the DA to the data
  - Shows **everything** about how data was **processed**, including plots and tables for the data quality to be assessed
  - Latest versions have trustworthy scores on a traffic light scheme (**Green**, **Blue**, **Yellow**, **Red**)

NOTE: Older ALMA data may not have used the ALMA pipeline NO WEBLOGS

Weblog – why should you care

First look at data and processing – is everything what you wished for…

- Good calibration is the root of good imaging
  - Possible calibration error could lead to imaging errors
- Pipeline has continually evolved, better heuristics and processing logic
  - Old weblogs might be limited or hint at things that could be improved
- Is there something useful in the data
  - Images and spectra can be seen along with beam and noise parameters

Given ALMA's QA process, everything should be good, to a level not adversely effecting Science Output

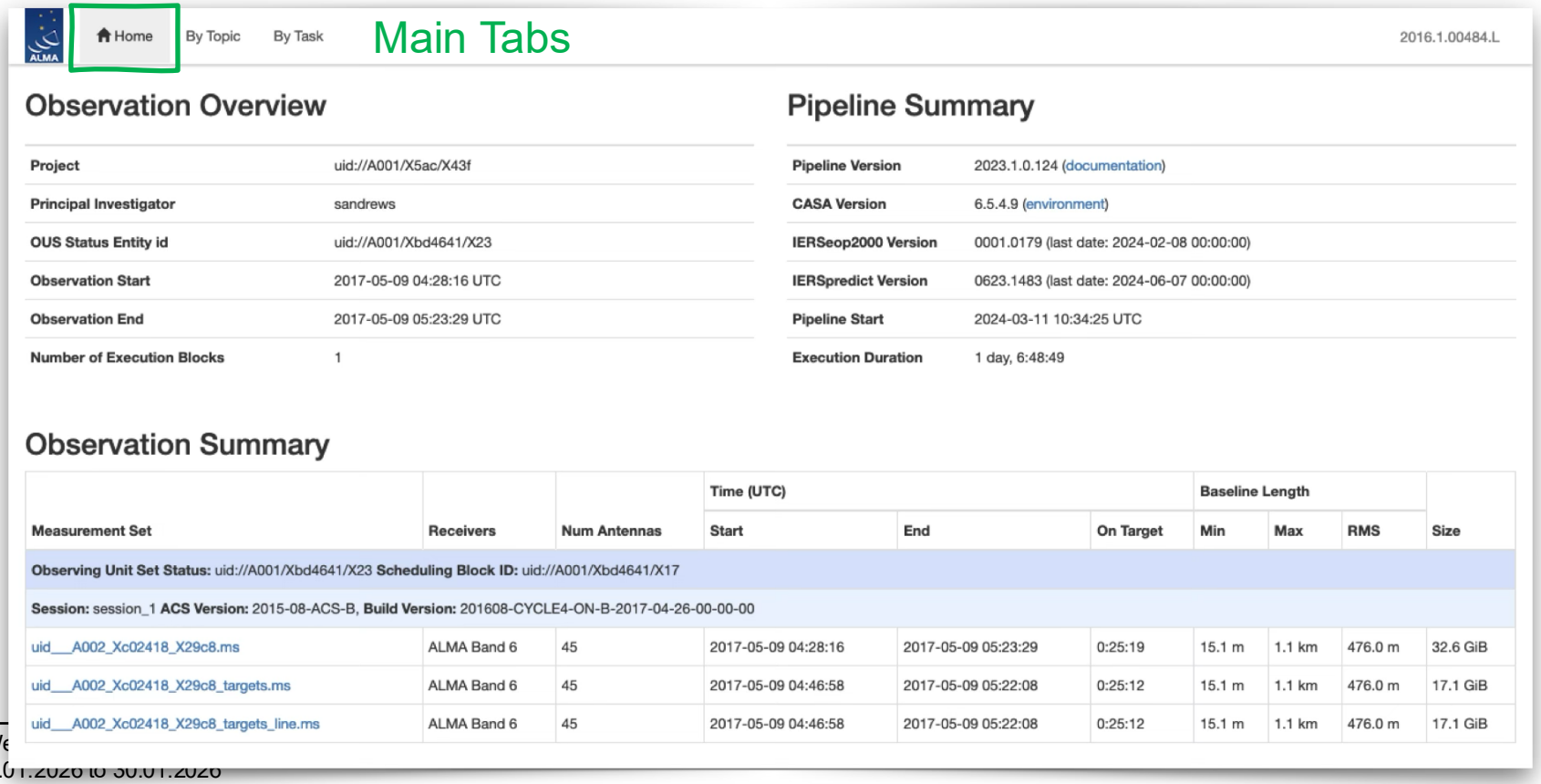
i.e. the point of these tutorials. - showing you how to redo

Weblog – take a look

Is there anything interesting inside?

- Main landing page – **HOME** – top taps govern main navigation
- All things in **BLUE** are a link

The weblog will show you everything you need to know, and more



Observation Overview

Project	uid://A001/X5ac/X43f
Principal Investigator	sandrews
OUS Status Entity id	uid://A001/Xbd4641/X23
Observation Start	2017-05-09 04:28:16 UTC
Observation End	2017-05-09 05:23:29 UTC
Number of Execution Blocks	1

Pipeline Summary

Pipeline Version	2023.1.0.124 (documentation)
CASA Version	6.5.4.9 (environment)
IERSeop2000 Version	0001.0179 (last date: 2024-02-08 00:00:00)
IERSpredict Version	0623.1483 (last date: 2024-06-07 00:00:00)
Pipeline Start	2024-03-11 10:34:25 UTC
Execution Duration	1 day, 6:48:49

Observation Summary

Measurement Set	Receivers	Num Antennas	Time (UTC)			Baseline Length			Size
			Start	End	On Target	Min	Max	RMS	
Observing Unit Set Status: uid://A001/Xbd4641/X23 **Scheduling Block ID:** uid://A001/Xbd4641/X17									
Session: session_1 **ACS Version:** 2015-08-ACS-B, **Build Version:** 201608-CYCLE4-ON-B-2017-04-26-00-00-00									
uid___A002_Xc02418_X29c8.ms	ALMA Band 6	45	2017-05-09 04:28:16	2017-05-09 05:23:29	0:25:19	15.1 m	1.1 km	476.0 m	32.6 GiB
uid___A002_Xc02418_X29c8_targets.ms	ALMA Band 6	45	2017-05-09 04:46:58	2017-05-09 05:22:08	0:25:12	15.1 m	1.1 km	476.0 m	17.1 GiB
uid___A002_Xc02418_X29c8_targets_line.ms	ALMA Band 6	45	2017-05-09 04:46:58	2017-05-09 05:22:08	0:25:12	15.1 m	1.1 km	476.0 m	17.1 GiB

Home | By Topic | By Task

Project Code 2016.1.00484.L

Observation Overview

Project	uid://A001/X5ac/X43f
Principal Investigator	sandrews
OUS Status Entity id	uid://A001/Xbd4641/X23
Observation Start	2017-05-09 04:28:16 UTC
Observation End	2017-05-09 05:23:29 UTC
Number of Execution Blocks	1

PI

Dates

Pipeline Summary

Pipeline Version	2023.1.0.124 (documentation)
CASA Version	6.5.4.9 (environment)
IERSeop2000 Version	0001.0179 (last date: 2024-02-08 00:00:00)
IERSpredict Version	0623.1483 (last date: 2024-06-07 00:00:00)
Pipeline Start	2024-03-11 10:34:25 UTC
Execution Duration	1 day, 6:48:49

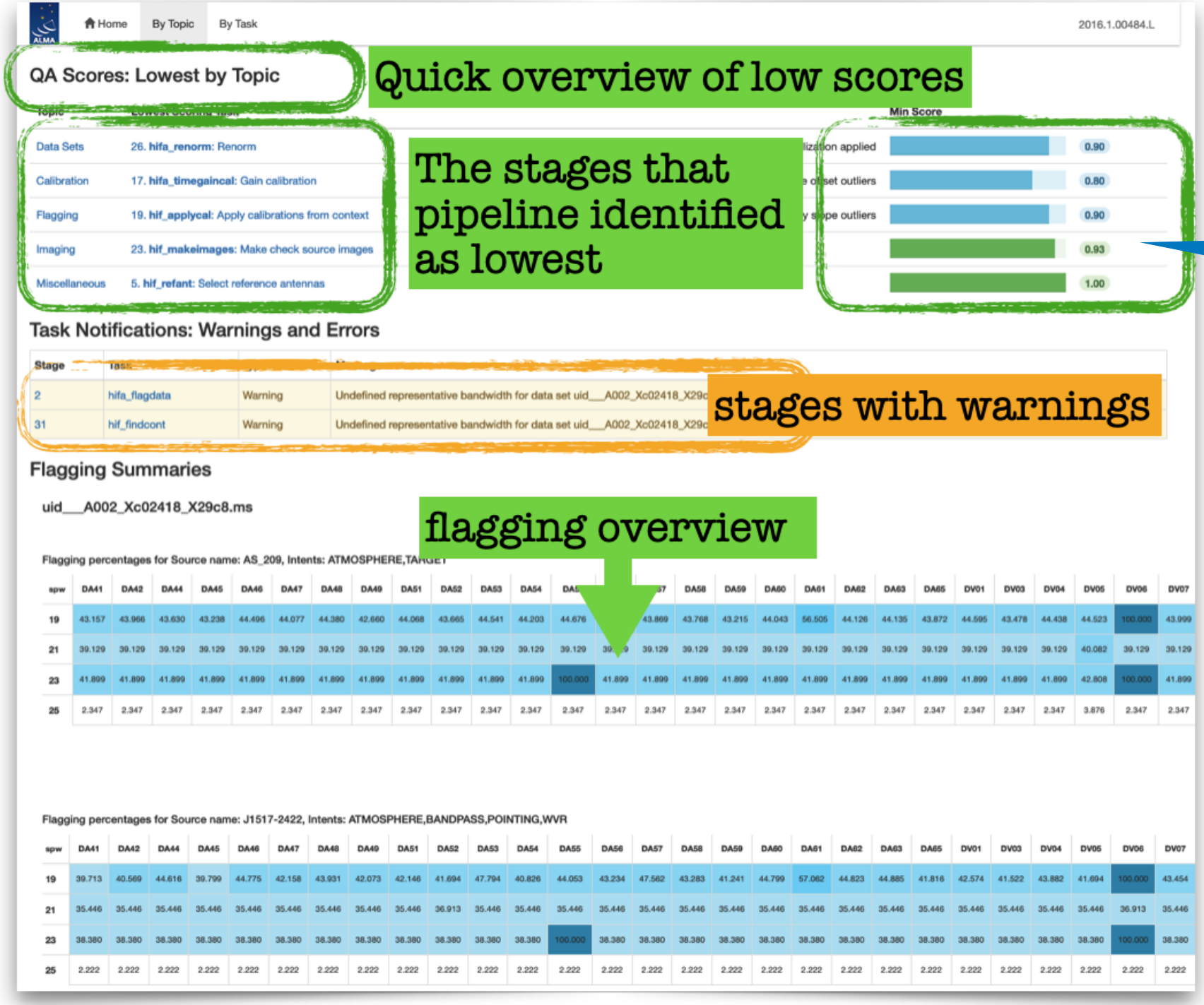
Versions

Observation Summary

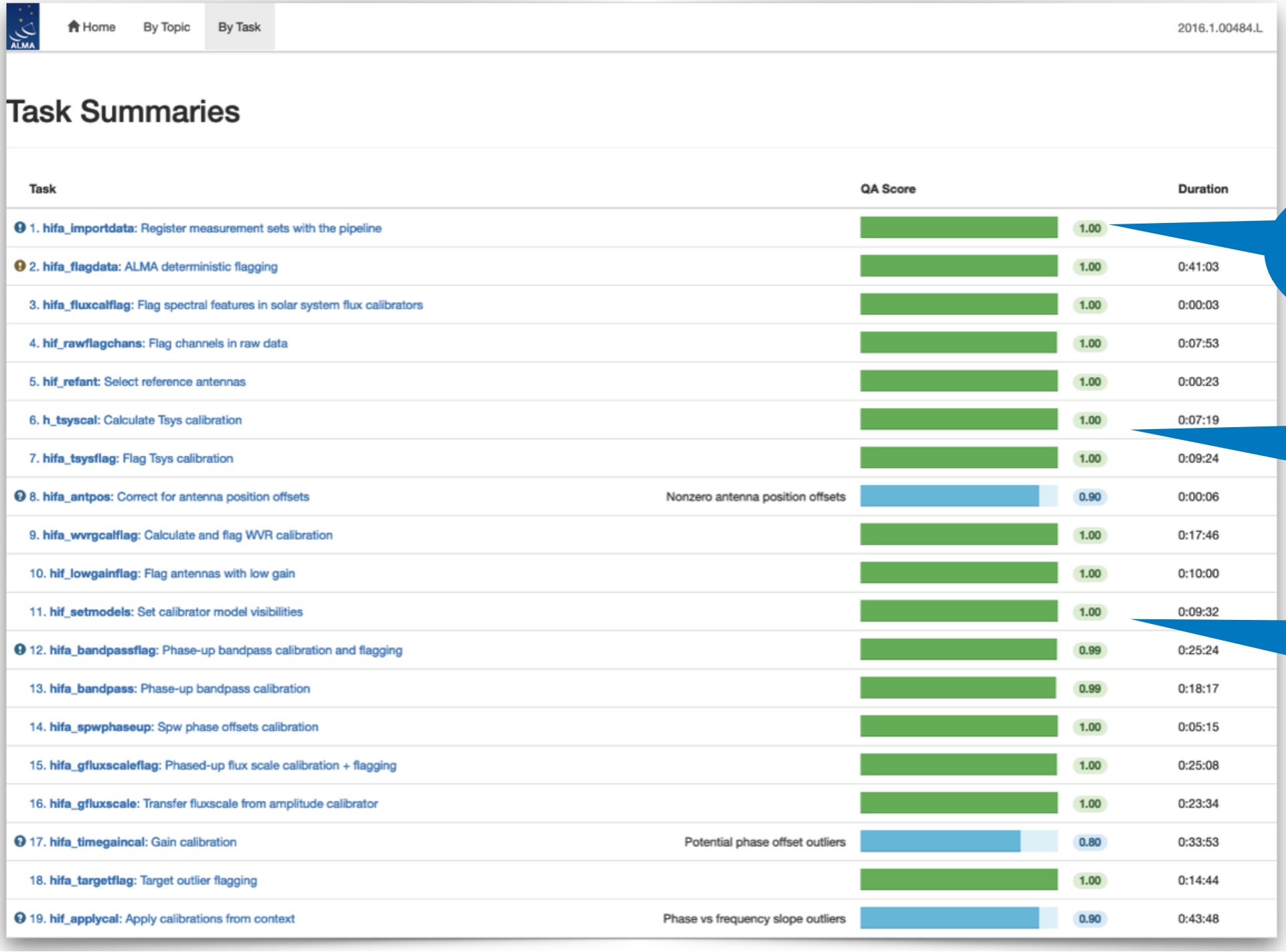
Measurement Set	Receivers	Num Antennas	Time (UTC) Start	End	On Target	Baseline Length Min	Max	RMS	Size
Observing Unit Set Status: uid://A001/Xbd4641/X23 **Scheduling Block ID:** uid://A001/Xbd4641/X17									
Session: [illegible]									
uid___A002_Xc02418_X29c8.ms			16	2017-05-09 05:23:29	0:25:19	15.1 m	1.1 km	476.0 m	32.6 GiB
[illegible]	ALMA Band 6	45	2017-05-09 04:46:58	2017-05-09 05:22:08	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
uid___A002_Xc02418_X29c8_targets_line.ms	ALMA Band 6	45	2017-05-09 04:46:58	2017-05-09 05:22:08	0:25:12	15.1 m	1.1 km	476.0 m	17.1 GiB

the Measurement Set

useful parameters



Overview of worst scores – in this one, nothing is yellow or red



”hifa” – ALMA specific interferometric stages

”h” – Pipeline stage can be used for any instrument, single dish, or interferometry

”hif” –specific interferometric stages

Weblog – take a look

Is there anything interesting inside?

- **Show of hands – have you looked at a Weblog in detail before ?**

- Open a browser and navigate to this weblog:

https://almascience.eso.org/arcdistribution/LeidenSchool/G183_QA2_delivered_weblog/html/

Links will be deleted mid-2026, you can download the weblog from the ALMA archive

Weblog – take a look

Is there anything interesting inside?

- Let's look for some things:
  - Names of the **traditional calibration** stages
  - What **stage** (in order) can you first see if molecular lines are detected in the source?
  - Did any Spectral Windows (SpWs) have **a low signal to noise**?
  - Was **self-calibration** performed?

Weblog – take a look

Is there anything interesting inside?

- Let's look for some things:
  - Names of the **traditional calibration** stages

Note, all stages in Pipeline are required to do calibration, they are running flagging and heuristic setups, NOT, just these few "traditional" stages

h_tsyscal (combined with hifa_tsysflag): Tsys gaintable and flags outliers

hifa_wvrgcalflag: makes the WVR solutions and flags bad antennas

hifa_bandpass: makes the bandpass solutions

hifa_gfluxscale: corrects the flux scaling using an amplitude calibrator

hifa_timegaincal: does all temporal gains, amp and phase

hif_applycal: applies all previously made solutions

Weblog – take a look

Is there anything interesting inside?

- Let's look for some things:
  - What **stage** (in order) can you first see if molecular lines are detected in the source?

hif_applycal

- **hif_applycal**: sorts and applies all calibrations to the calibrators and targets
- Many plots are made of phase, amplitude, uvdistance, frequency
- Some amp vs. freq. could show molecular lines in the target, but it's a average of the visibilities
- Take care as more noise (positive amplitude noise) near ATM features (low transmission) look line lines – but are not

This stage is preferable though

Weblog – take a look

Is there anything interesting inside?

- Let's look for some things:
  - What **stage** (in order) can you first see if molecular lines are detected in the source?

hif_findcont

- **hif_findcont**: this is the first stage to **actually run target imaging**
- It makes some dirty cubes, behind the scenes, and checks channel by channel for molecular lines
- This is required so Pipeline knows later what to use for continuum subtraction

Weblog – take a look

Is there anything interesting inside?

- Let's look for some things:
  - Did any Spectral Windows (SpWs) have **a low signal to noise**?

YES

  - **hifa_spwphaseup**: this stage governs how calibrations proceed, if there is low SNR a SpW can be mapped to another one, or if many SpWs have low SNR they can all be combined
  - **SpW 28** is low SNR (16.5 where 32.0 is required and is mapped to 30)
  - Since PL2025 (released Oct 2025) all INTENTs (i.e. calibrator fields) can be assessed for mapping and combine

Weblog – take a look

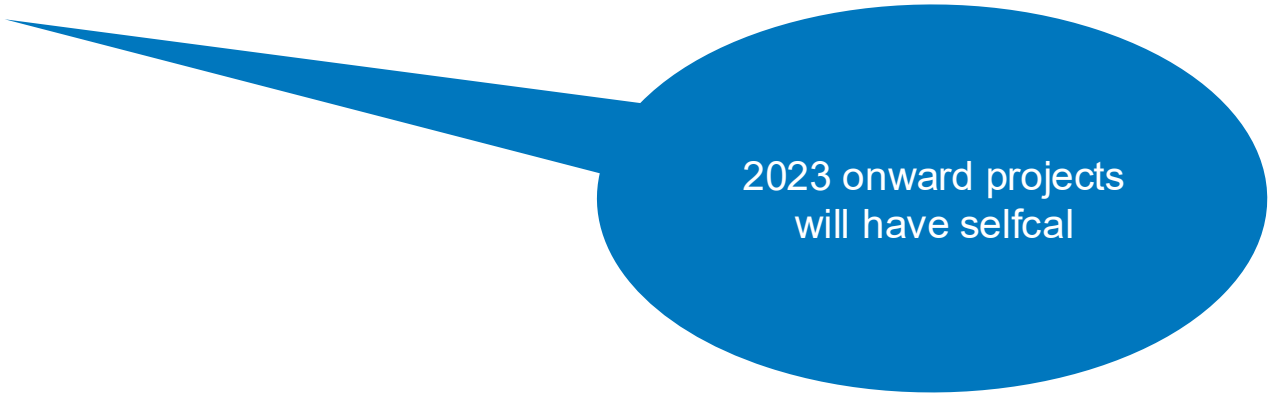
Is there anything interesting inside?

- Let's look for some things:
  - Was **self-calibration** performed?

NO

  - **hif_selfcal**: global interferometric task for self-calibration **available from Cycle 10 only**

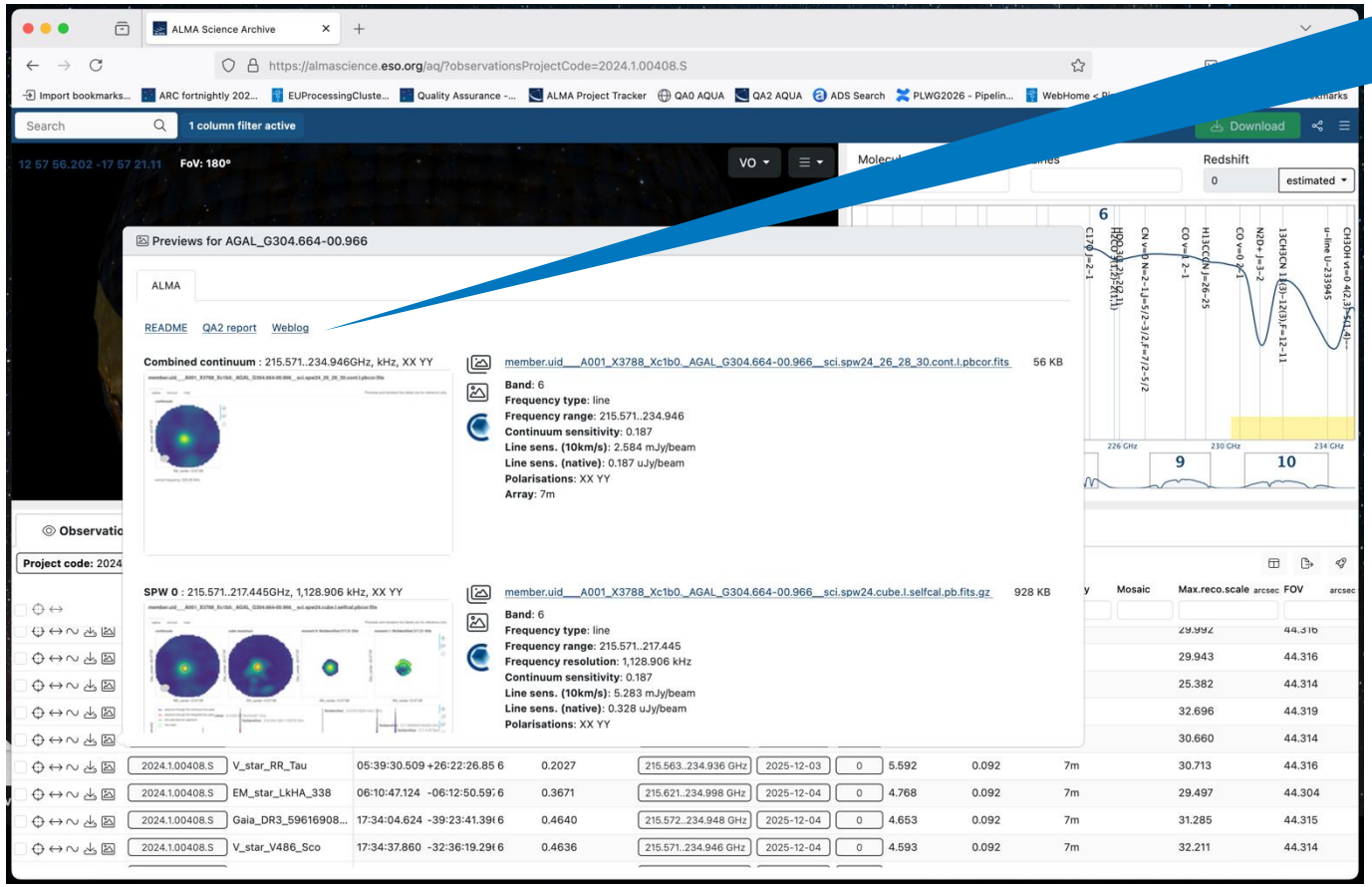
G183 is older data, 2022 project, the Pipeline used does not have self-cal, or latest heuristics - we can reprocess!!!



Weblog – other ways to look

Is there anything interesting inside?

- ALMA Archive is now providing direct access to some weblogs



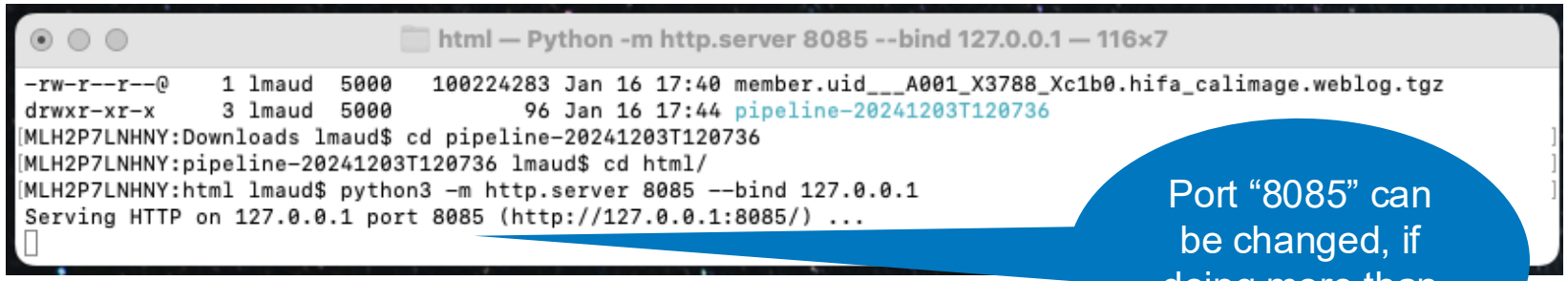
Click to download from Archive

Weblog – other ways to look

Is there anything interesting inside?

- ALMA Archive is now providing direct access to **some** weblogs
  - Click to download
  - Untar

  - *Run* **ls –ltr** *to see last file name*
  - Move to the “html” folder
  - Open via local server with Python



Port “8085” can be changed, if doing more than one weblogs

What are our options for processing?

Options on what to

Right into science, or do I need to reconsider the calibration or imaging?

- How to **restore** calibrated data
- How to do/re-do continuum subtraction (in Pipeline)
- How to use the pipeline self-calibration process
- How to totally **reprocess** data in the latest Pipeline
- How to use pipeline to re-image

https://casa.nrao.edu/casa_obtaining.shtml

What to keep in mind

Which CASA? Do I need a pipeline version?

- There is a compatibility matrix for both CASA and the ALMA Pipeline

PROJECT CODE	CYCLE	CASA	Release	
<=2013	0,1,2	4.4.0, 4.3.1, 4.2.2, 4.2.1, 4.2.0, 3.4, 3.3.	Redhat, Mac OS 10.7, 10.8	later some cal PL
2015	3	4.6.0, 4.5.2, 4.5.1	RH 5	
2016	4	4.7.2,.4.7.0-1, 4.7.0	RH6/7, MacOS 10.10, 10.11	image PL too
2017	5	5.4.0-70, 5.4.0-68, 5.3.0, 5.1.1	RH6/7 MacOS 10.12	
2018	6	5.6.1-8, 5.4.0-70	RH6/7, Mac OS 10.13, 10.14	
2019	7	6.1.1-15, 5.6.1-8	RH 6-9, Mac OS 10.15	
2021	8	6.4.1-12, 6.2.1-7	RH 6-9, Mac OS 11	
2022	9	6.4.1-12 + patch	RH 6-9, Mac OS 11	
2023	10	6.5.4-9	RH 6-9, Mac OS 11	
2024	11	6.6.1-17	RH 6-9, Mac OS 12	
2025	12	6.6.6-17	RH 6-9, Mac OS 13, 14	

Most manually calibrated, no Pipeline, no weblog

Pipeline here onwards is really very good

Pipeline has best lowSNR heuristics

Pipeline now has selfcal, these are best versions

https://almascience.eso.org/processing/science-pipeline

What to keep in mind

Which CASA? Do I need a pipeline version?

- There is a compatibility matrix for both CASA and the ALMA Pipeline

CASA versions accepted for ALMA data processing:

CASA version	Pipeline branch and version	Pipeline Documentation	Description	used in operations	tarball for most modern OS available for each	versions that can be used to restore these data[1]
6.6.6-17	2025.1.0.35	User's Guide Reference Manual 2025 Pipeline 2025 known issues	Cycle 12	2025-09-29	casa-6.6.6-17-pipeline-2025.1.0.35-py3.10.el8.tar.xz casa-6.6.6.17-pipeline-2025.1.0.35-14.0-arm64-py310.dmg	most recent
6.6.1.17	2024.1.0.8	User's Guide Reference Manual 2024 Pipeline 2024 known issues	Cycle 11	2024-09-30 ~ 2025-09-29	casa-6.6.1-17-pipeline-2024.1.0.8-py3.8.el8.tar.xz casa-6.6.1.17-pipeline-2024.1.0.8-12.0-py38.dmg	most recent
6.5.4.9	2023.1.0.124	User's Guide Reference Manual 2023 Pipeline 2023 known issues	Cycle 10	2023-09-30 ~ 2024-09-29	casa-6.5.4-9-pipeline-2023.1.0.124-py3.8.tar.xz casa-6.5.4.9-pipeline-2023.1.0.124-11.0-py38.dmg	most recent
6.4.1.12	2022.2.0.68	User's Guide Reference Manual 2022.2 Pipeline 2022 known issues	Cycle 9 patch	2023-04-18 ~ 2023-09-30	casa-6.4.1-12-pipeline-2022.2.0.68-py3.6.tar.xz casa-6.4.1.12-pipeline-2022.2.0.68-11.0-py36-py36.dmg	most recent
6.4.1.12	2022.2.0.64	User's Guide Reference Manual 2022.2 Pipeline 2022 known issues	Cycle 9	2022-09-27 ~ 2023-04-18	casa-6.4.1-12-pipeline-2022.2.0.64-py3.6.tar.xz	most recent
6.2.1-7	2021.2.0.128	User's Guide Reference Manual 2021.2 Pipeline 2021 known issues	Cycle 8	2021-10-01 ~ 2022-09-26	casa-6.2.1-7-pipeline-2021.2.0.128.tar.xz casa-6.2.1.7-pipeline-2021.2.0.128-10.15-py36.dmg	most recent
6.1.1-15	2020.1.0-40	Users Guide 2020.1 Reference Manual 2020.1 Pipeline 2020 known issues	Cycle 7 reprise	2021-05-10 ~ 2021-10-01	casa-6.1.1-15-pipeline-2020.1.0.40.tar.xz casa-6.1.1.15-pipeline-2020.1.0.40-10.15.dmg	casa-6.1.1-15-pipeline-2020.1.0.40 or most recent
5.6.1-8	Pipeline-CASA56-P1-B, r42866	Users Guide 5.6.1	Cycle 7	2019-10-01 ~ 2021-05-10	casa-pipeline-release-5.6.1-8.el7.tar.gz casa-pipeline-release-5.6.1-8-10.14.dmg	casa-pipeline-release-5.6.1
5.4.0-70	Pipeline-CASA54-P1-B, r42254 and r42866	Users Guide 5.4 Reference Manual 5.4	Cycle 6	2018-10-01 ~ 2019-10-01	casa-release-5.4.0-70.el7.tar.gz	casa-pipeline-release 5.4.0 or 5.6.1

Older means then old CASA and old Pipeline

Options on what to do – (1)

Get to the imaging as fast as possible

To get to imaging we need to **restore** the previous calibration **done by ALMA**

- **scriptForPI.py** is the wrapper that does all the work
- You can only do a restore with **compatible versions** of CASA and Pipeline
  - Likely need older OS machine for anything but latest data – **this can be a problem**
  - Can use the **CalMS** service
  - Pipeline from circa 2021-2022 with CASA 6.2.1 **can** restore using **newer** versions (*care with options for e.g. dosplit, docontsub)*

Options on what to do – (1)

Get to the imaging as fast as possible

To get to imaging we need to **restore** the previous calibration **done by ALMA**

- For a **Pipeline** run, **you only restore** existing calibration tables. You can flag before imaging, but you have **no access to any of the calibration process**
- **Nuance** – **manually calibrated data** runs the calibration script, again, in CASA
  - This remakes and applies calibration

- After a restore, you **can use the latest** CASA and Pipeline for imaging
  - If using Pipeline stages for imaging (very) old data, there may be some warnings or necessity to hard code inputs

Options on what to do – (2)

Improve the calibration and flagging

This means you want to **reprocess** the data, effectively **redoing** the QA2 from ALMA

- **scriptForPI.py** can help with this for **Pipeline** processed data
- You might have to be careful of some required input files (e.g. getting correct source reference fluxes)
  - If you use the older CASA and Pipeline version, **helper** scripts and files need to be copied
  - When using the scriptForPI.py it will do this copying
- *For manual calibration you can edit the calibration script, but you are fixed to the old CASA versions due to format structure and options*

- Best option is to **reprocess with the latest ALMA Pipeline**

Options on what to do – (3)

Want to manually improve the continuum selection

This means you are happy with the previous calibration but want to edit the continuum ranges

- **hif_findcont** is part of the **imaging** stages
  - You can look at the existing weblog and judge the continuum and lines
  - You can investigate the existing images to judge the continuum and lines
- You can made a new, or edit an existing **cont.dat** file
  - Care must be taken as the units in frequency are LSRK
  - You can start from restored data, or Pipeline calibrated data (without imaging), or even add the cont.dat for a full Pipeline reprocessing (*findcont will be skipped in this case*)

Options on what to do – (4)

Only try to do self-calibration in Pipeline

This means you are happy with the previous calibration but want to do automatic self-calibration

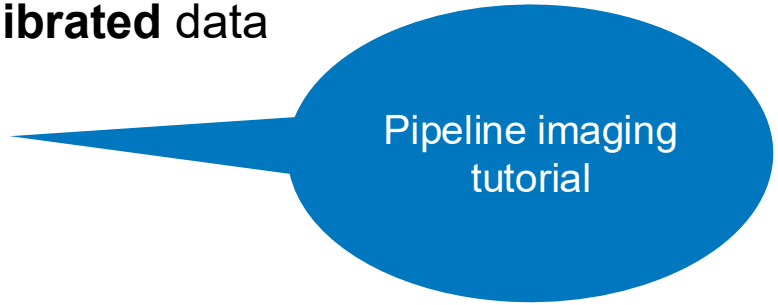
- **hif_selfcal** is part of the **imaging** stages
  - You can use restored data to image and do self-calibration
  - Data earlier than 2021 might not work properly – self-cal needs some metadata that might not exist
    - You can test this first and see if selfcal completes
    - another option is to **reprocess** with the latest ALMA Pipeline
- You can edit some self-cal options as to not be limited by “PI” requirements
  - E.g. the beam may change during self-calibration, it is a problem for QA2, but maybe not for your science

Options on what to do – (5)

Image only with the Pipeline

Not really interested in the calibration parts

- You **must** have run a **restore** or **reprocessing** to have the **calibrated** data on disk
- You can use the **calMS** service and download the copy of the **calibrated** data
- You have to follow (or script) the set pipeline imaging tasks



Move to TUTORIAL - *restore*

The Pipeline Context System

Wait…what? Context….

What is the Pipeline Context?

The Pipeline tracks where it is, and what was done

The ALMA Pipeline must store information about the data processing and imaging

- The **context** is not an official interface, it is used internally by the pipeline
  - i.e. the Pipeline developers and working group can change parameters with various releases
- The **context** is a Domain Object
  - Stores metadata about the dataset, spectral setup, fields etc.
  - Stores information about running stages
  - Stores information about calibration tables

The Context and the Pipeline

There is a sequence, and messy things happen if it is broken

The ALMA Pipeline, is a **Pipeline**, to run, stage-by-stage, automatically

- The **context** will correctly store stage-by-stage everything required for calibration and imaging to proceeds and for each stage to have access to information it needs
- The Pipeline has to be run in order, which in order will store to the **context**
  - Repeating (some) stages will likely result in appending successive information over-and-over
  - e.g. repeating the **hifa_gfluxscale** will continue to apply previous known corrections as told by the context – this is invalid
- Exiting Pipeline without correct context **save points** will render the data state and the context **incompatible**
  - Without saving, the context thinks processing (or imaging) is not as far as it is, stages will be saved overlapping to the same number
  - If in doubt begin processing/imaging again in a clean/new directory

For example, in manual running of Pipeline stages

Pipeline Context from the User perspective

Following good practices will save time

Users don’t really need to know what the context is, or what’s inside, but

- Users **must** initiate the context
- Users **must** save the context at appropriate points
- Users **can restore** a context and **continue** a processing or imaging job, provided saving rules were followed
- Users can move/copy the context and all data run to a given saved point to another directory (with another name) as to restart a Pipeline run from that point again

Pipeline Context from the User perspective

A hypothetical example

Run session 1:

```
h_init()
hifa_importdata(vis='uid___A002_X123_Xabc'
h_save()
> exit
```

In CASA with Pipeline, you only **imported** data, then had to stop the job and leave

Run session 2:

```
h_resume("pipeline-
20260119T173936.context")
hif_mstransform()
hifa_flagtargets()
hifa_imageprecheck()
hif_makeimlist(specmode='mfs')
hif_fincont()
hif_uvcontsub()
hif_makeimages()
h_save()
```

Now you **resume** the session, and don't need to import again the context knows you did it

Here you continue with some imaging commands for continuum subtractions and MFS imaging

Pipeline Context from the User perspective

Following good practices will save time

Experience users can save to multiple context files, but this becomes very complex

We will do this later

- My advice for interactive sessions:
  - Copy a dataset/directory before starting so you can always restart
  - Make a habit to repeat **h_save()** regularly after tasks
  - If a mess happens or stages fail, start fresh
- If you are doing a calibration or long run, also copy a full directory out at various intervals
  - If you are halfway through a long calibration or imaging and are worried you migh make a mistake or something might crash
  - This moves the entire data, context, calibration and images into another directory which can be **resumed** at the same spot again

Back to TUTORIAL