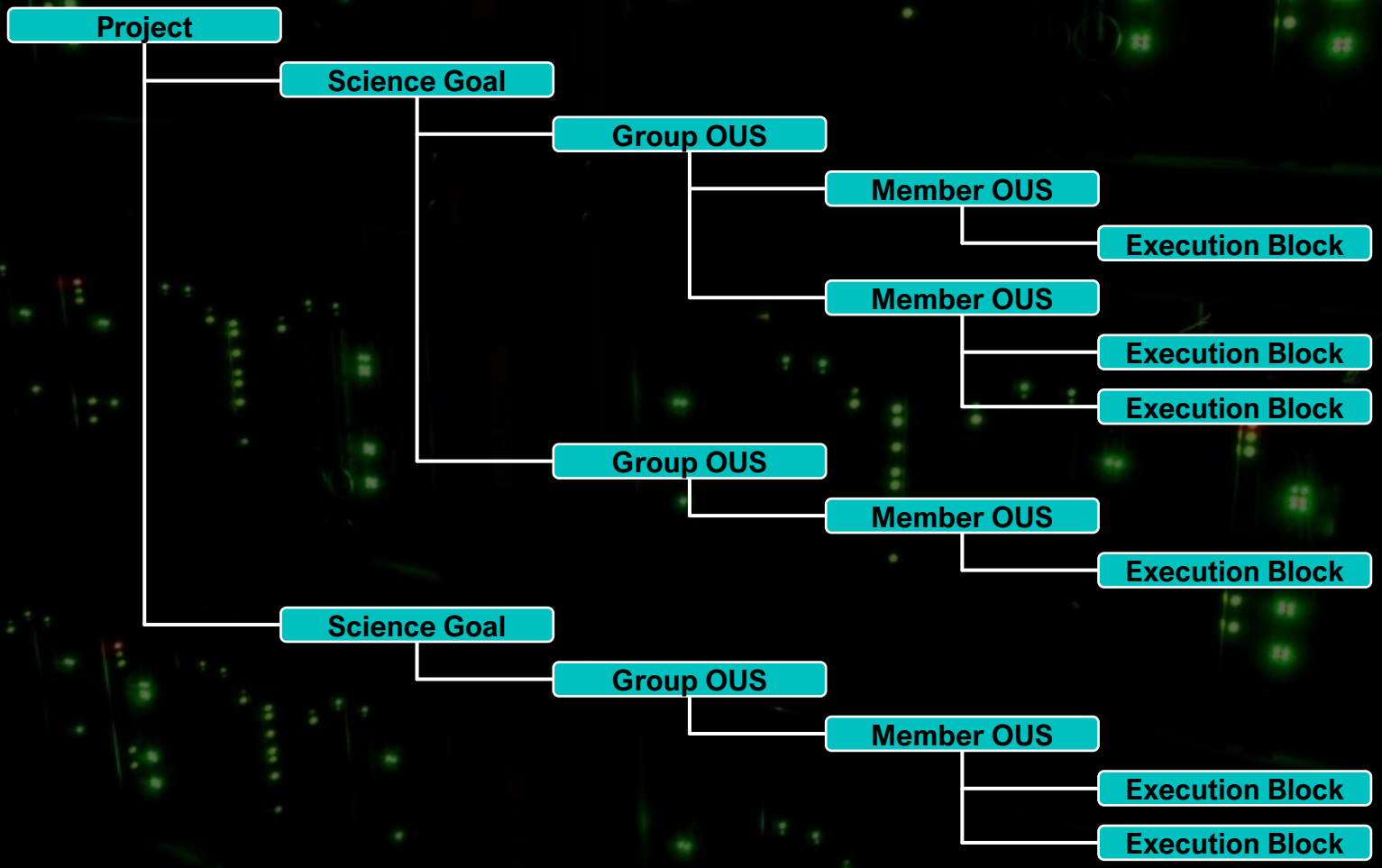
Archive content

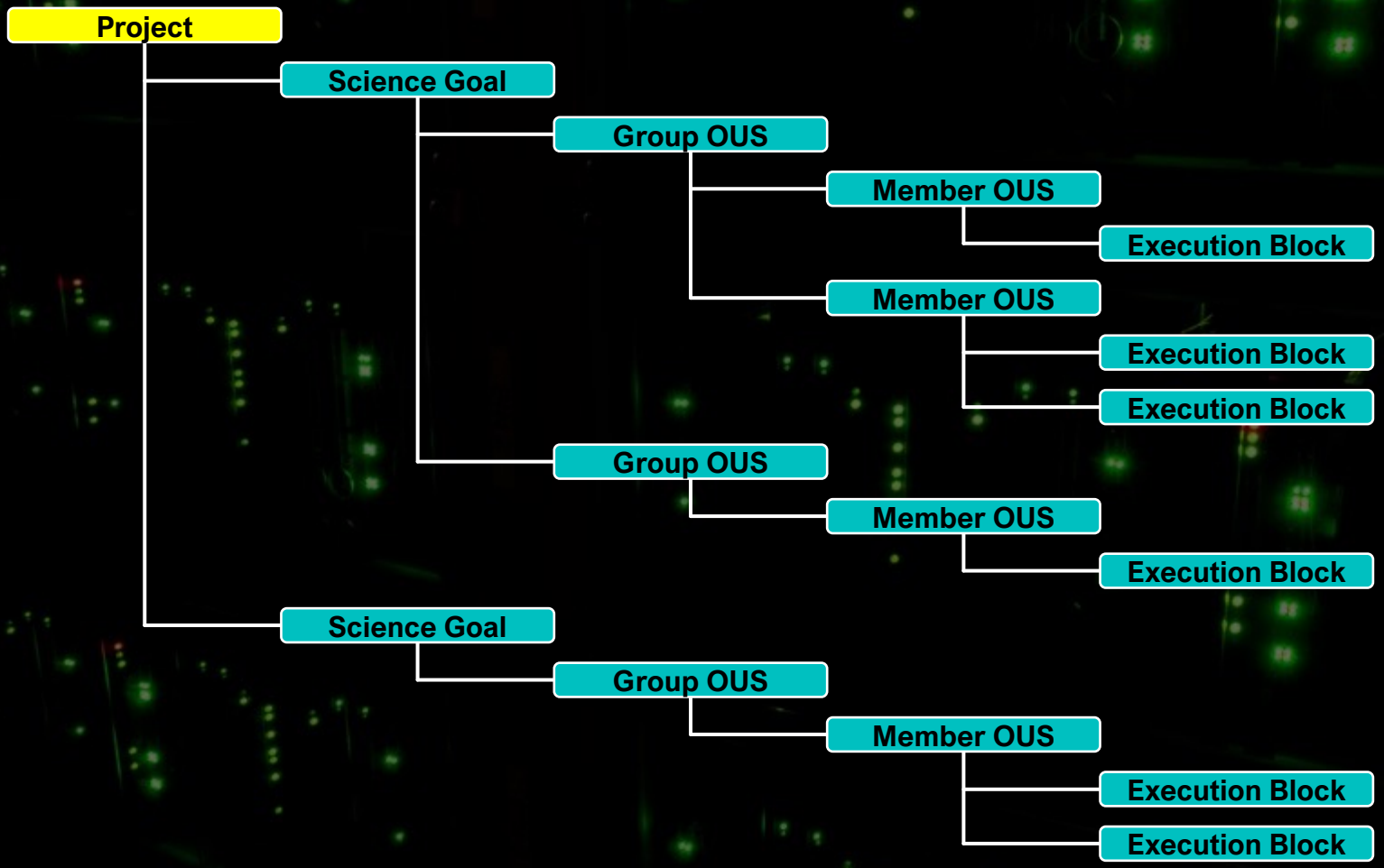
George Bendo and Marcella Massardi

ALMA Science Archive School
Italian ARC node headquarters
05-07 October 2022

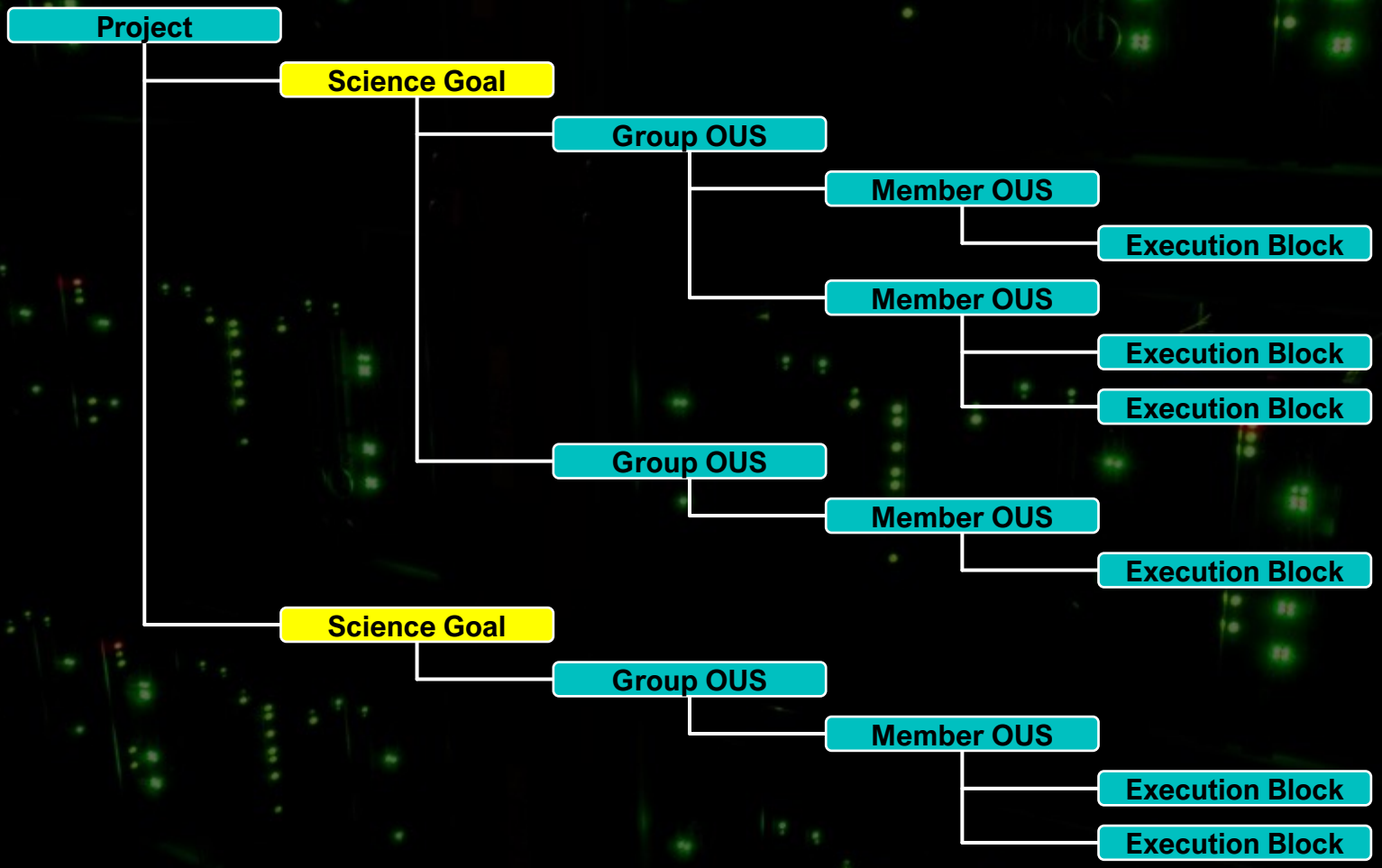
ALMA data in the archive (as well as downloaded ALMA data) are generally organized into the following structure.



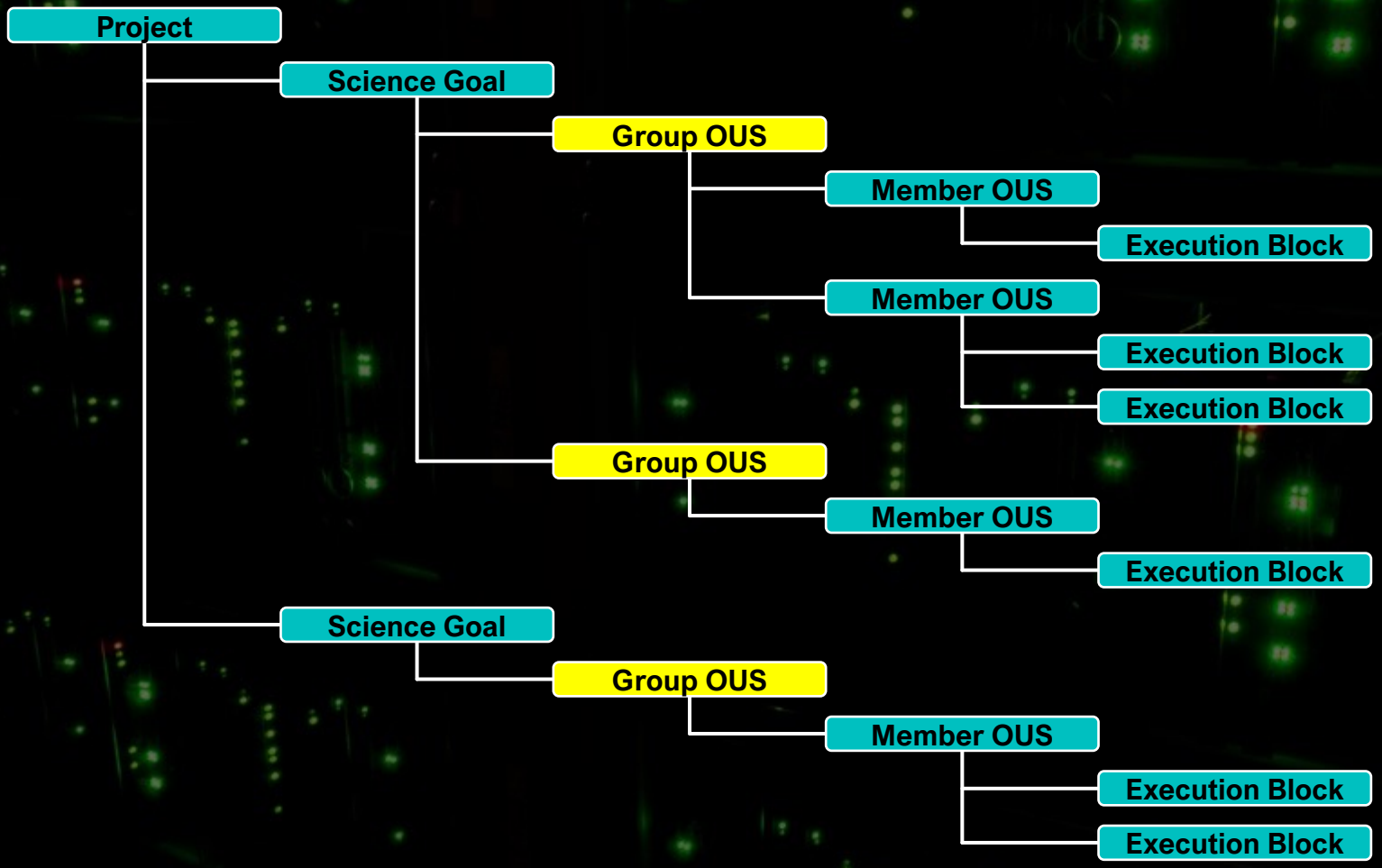
Project - All observations associated with a specific proposal.



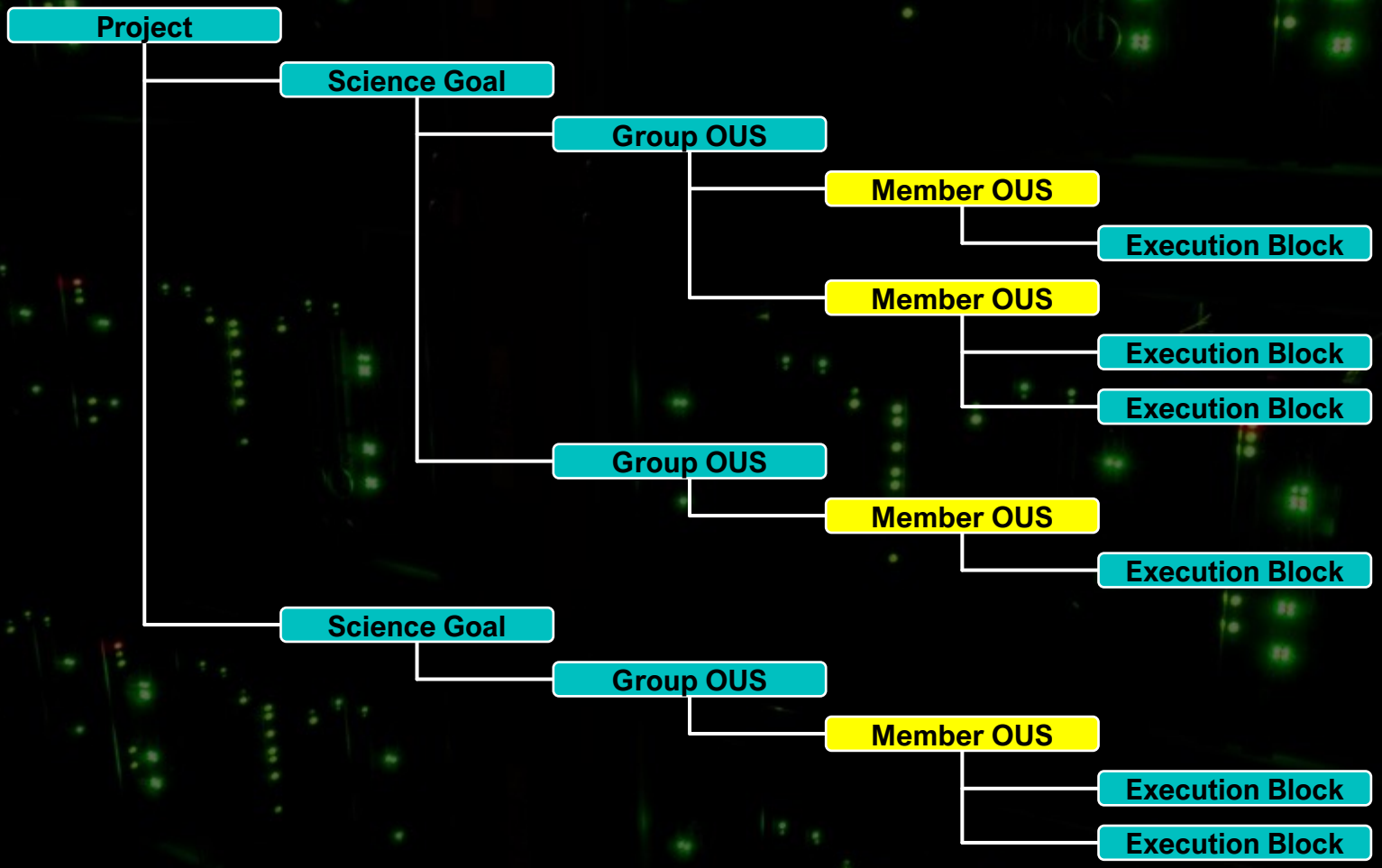
Science Goal OUS - All observations associated with a specific science goal in that proposal.



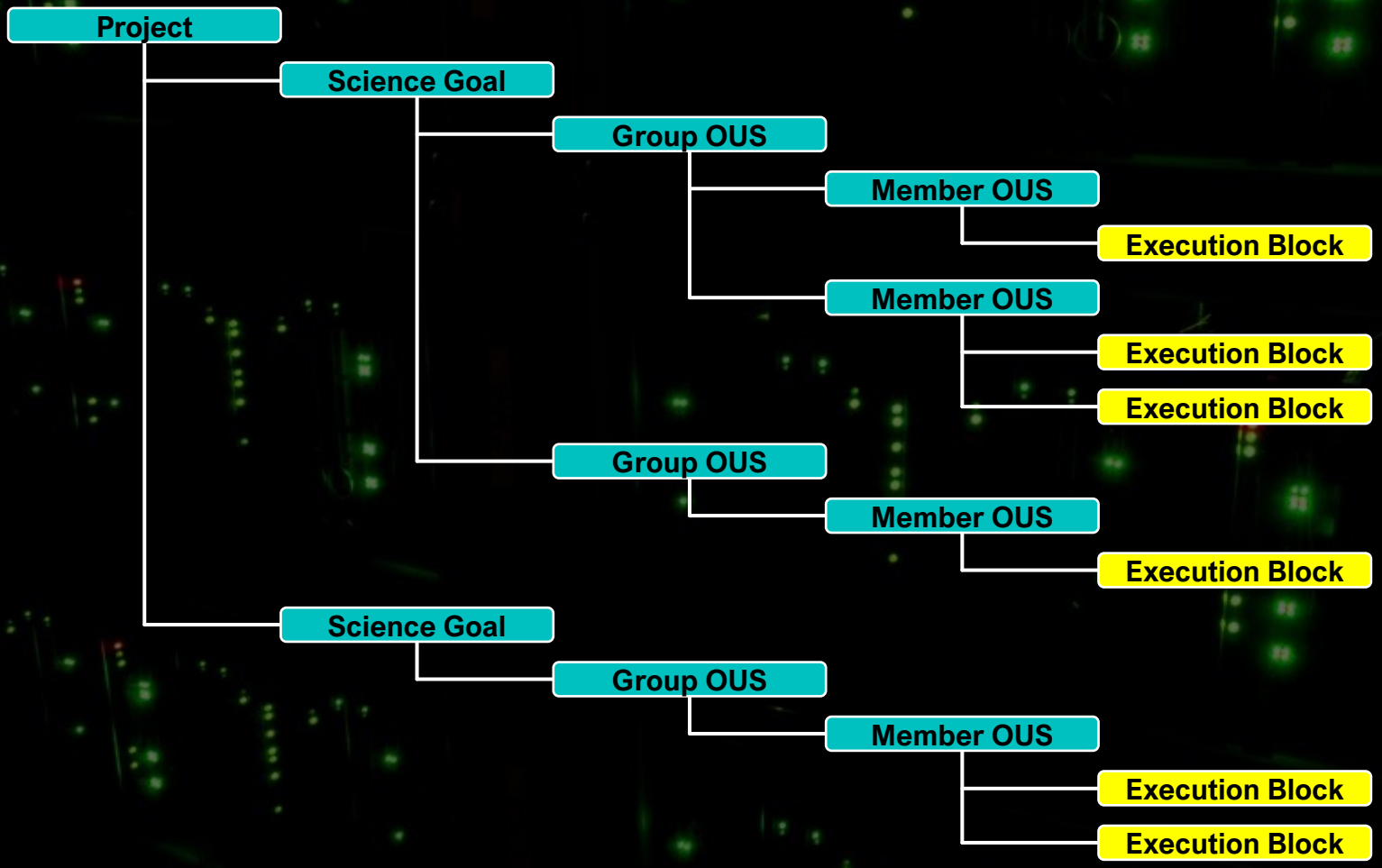
Group OUS – Associated observations within a Science Goal (e.g., observations of the same fields with the same spectral tunings but with different arrays or array configurations.



Member OUS – A specific set of observations of the same fields using the same tunings and array or array configuration.



Execution block – An individual “unit” of the observations needed for a Member OUS.



Each Member OUS (or SB) may have the following files available for download:

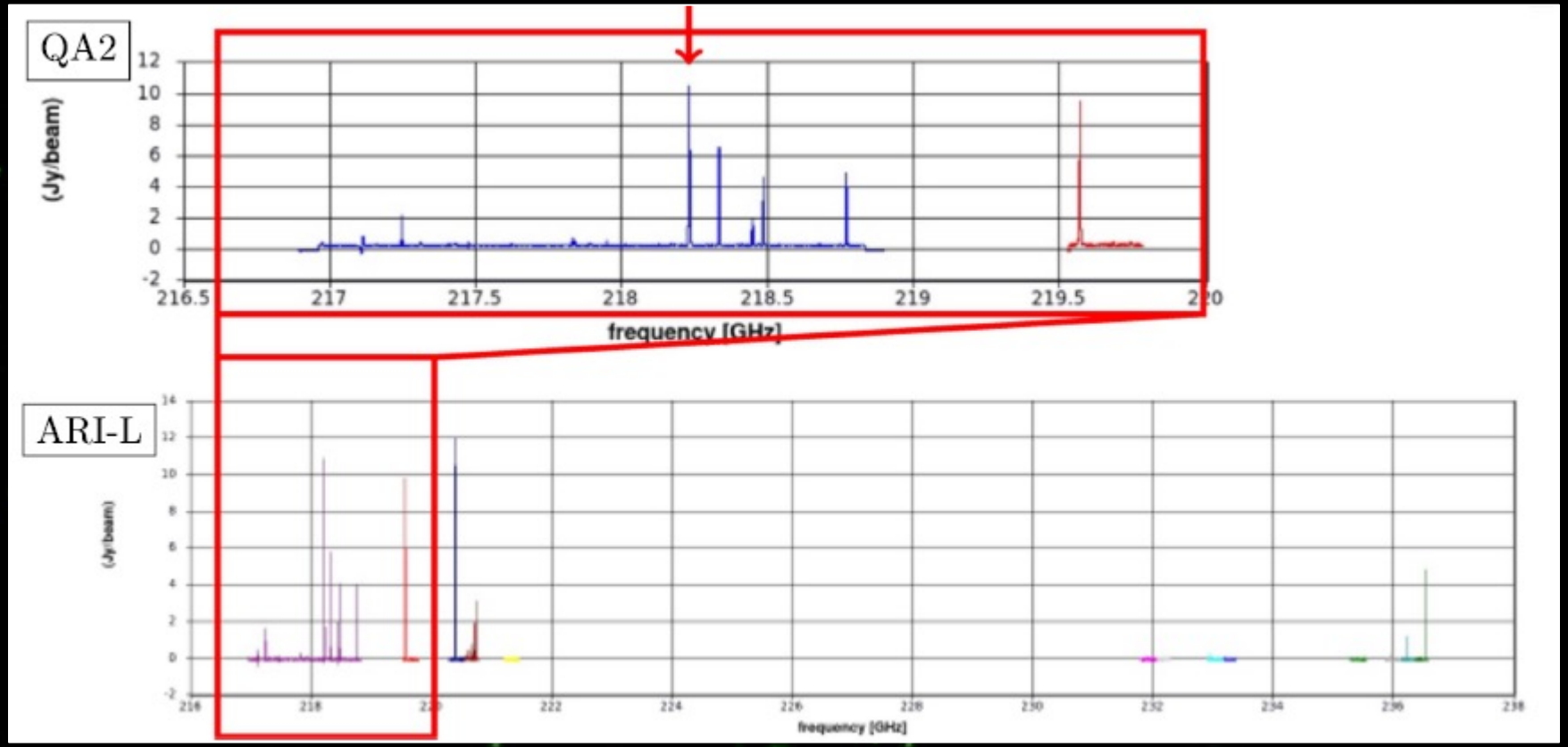
readme	A text file with very basic information
product	Final images and image cubes
auxiliary	A file containing logs, quality assurance information, scripts, and calibration data
raw	Raw visibility data
external	Enhanced data products (including enhanced images or visibility data) created after the data delivery

Currently, two forms of external data products are available from the archive:

Large Program data products – Enhanced data from Large Programs

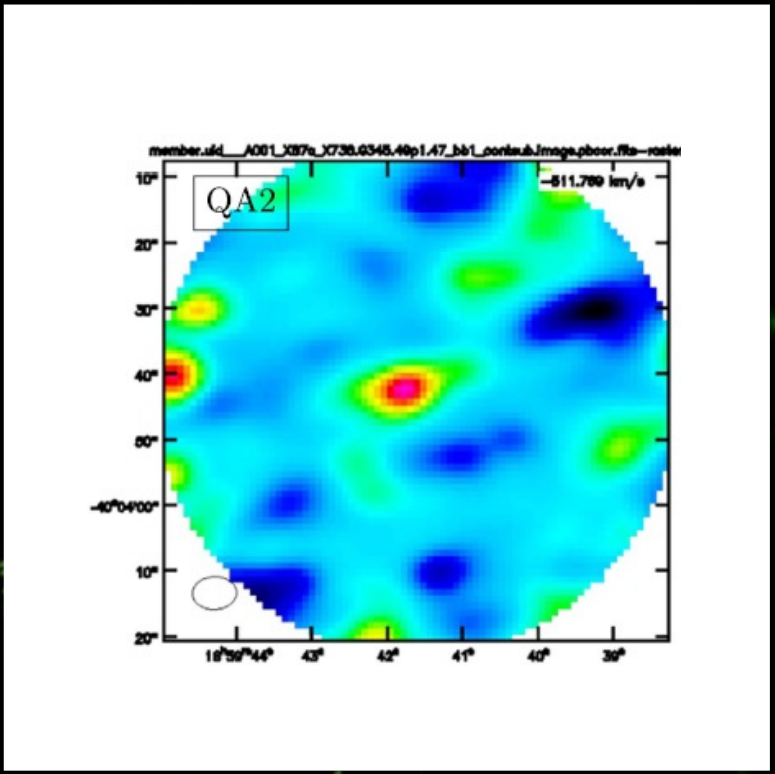
ARI-L data products – Data from Cycles 2-4 reprocessed using the CASA pipeline

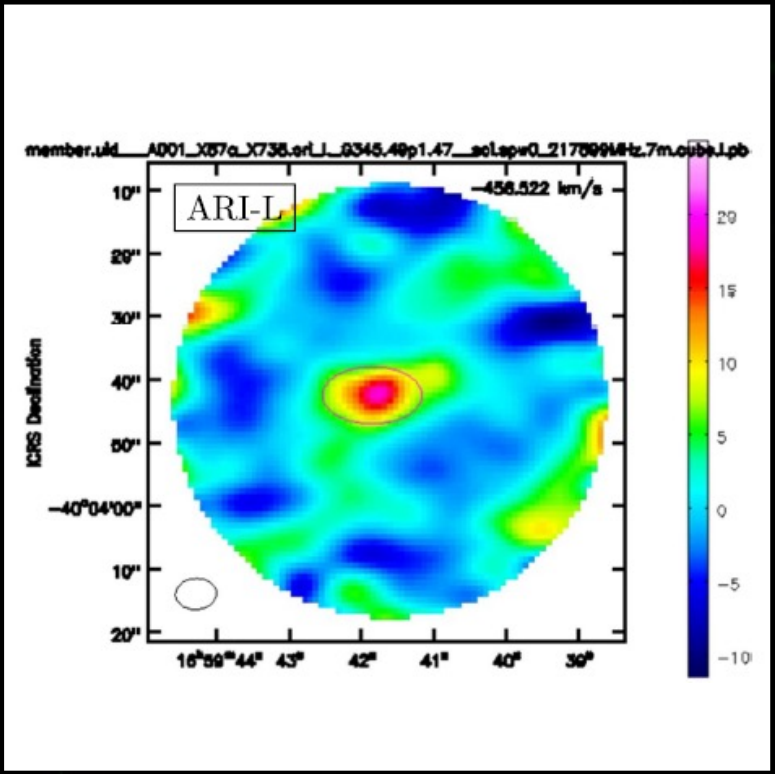
ARI-L (the Additional Representative Images for Legacy) is an ALMA Development Project (PI: Massardi) that restored the ALMA calibration and performed imaging with the ALMA Pipeline. This was applied to data from Cycles 2-4 that lacked pipeline-processed data with representative images comparable to those of later cycles.



Old (**QA2**) versus new (**ARI-L**) data.

ARI-L (the Additional Representative Images for Legacy) is an ALMA Development Project (PI: Massardi) that restored the ALMA calibration and performed imaging with the ALMA Pipeline. This was applied to data from Cycles 2-4 that lacked pipeline-processed data with representative images comparable to those of later cycles.





Old (**QA2**) versus new (**ARI-L**) data.

Details about ARI-L:

- The data are generally available for Cycles 2, 3, 4 (Project Codes 2013*, 2015*, 2016*).
- The image products use native spectral resolution and Briggs 0.5 weighting.
- The images are included in the archive previews and visualization.
- The data can be queried as the collection “ari_l”.
- The files can be downloaded as external data products.
- The data for each pipeline processable MOUS include (for each source and calibrator):
  - Aggregate continuum image.
  - Images of the mfs continuum for each spectral window.
  - Cubes for each spw.

ARI-L statistics:

- 18 people
- 3.5 years
- 2649 MOUSs processed
- 411 652 files ingested into the ASA
- 126 302 090 channels imaged and available from the ASA

Website: https://almascience.eso.org/alma-data/aril

Acknowledgment: Massardi et al., 2021, PASP, 133, 085001 (2021PASP..133h5001M)

Most data that are delivered to the archive have passed ALMA's quality assurance steps. However, some data may be available from the archive that are labelled as Semi-pass.

QA0 Semi-pass – The observations were not used to create the final images (because, for example, the observations are missing calibration observations, or the science target observations were not completed).

QA2 Semi-pass – The observations are complete but did not meet the science requirements in the proposal (e.g., the sensitivity goals or beam size requirements).

These data are still potentially usable for science, albeit with some limitations.

Data from different ALMA cycles were produced differently and therefore may have different characteristics.

- **Cycle 0 (2011)** data are in a different format from other data, and the directory structure is different from other cycles.
- Data from **Cycles 1-2 (2012, 2013)** are in an older format that is missing some keywords and where the visibility weighting may be different, which will cause problems if the data are combined with data from later cycles.
- Data from **Cycles 0-4 (2011-2016)** were generally processed manually before being uploaded to the archive, while data from **Cycles 5 and later (2017 and later)** were pipeline processed.

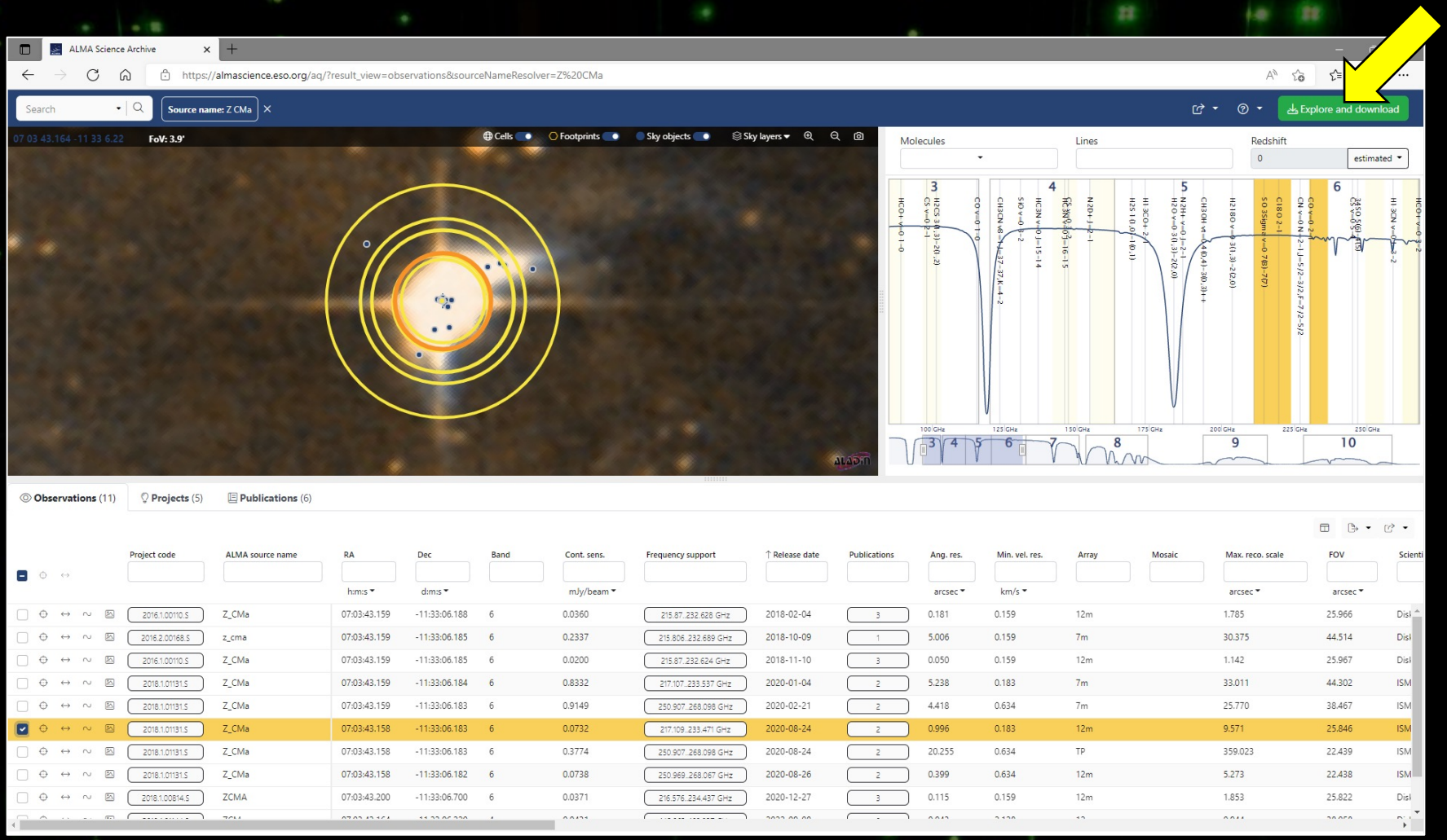
The images in the archive generally have the following suffixes:

`*mask.fits`	The mask used when an image was created
`*pb.fits` or `*.flux.fits`	The primary beam response for a field
`*pbcor.fits`	A primary-beam corrected image
`*sd.im.fits`	A single-dish image

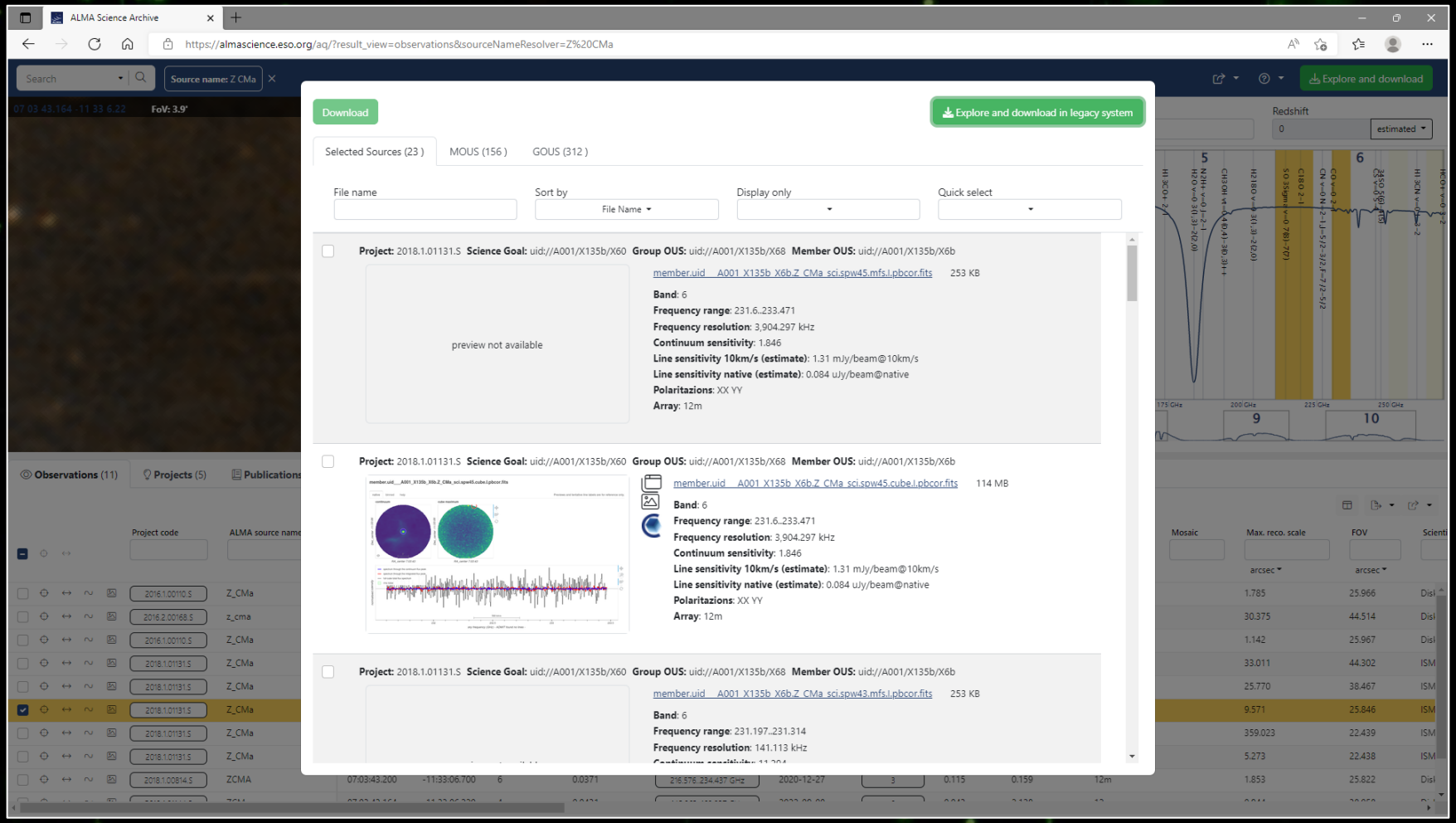
Additional parts of the filenames indicate the following:

`*mfs.A.*` or `*mfs.POLA*`	A polarization angle map
`*.mfs.P.*` or `*mfs.POLI*`	A linear polarization intensity map
`*spw##.cube.I.*`	A spectral image cube of a single spectral window
`*spw##.mfs.I.*`	A continuum image for a single spectral window
`*spw##_##_##_##.cont.I.*`	An aggregate bandwidth or continuum image
`*spw##_##_##_##.cont.I.alpha.*`	A spectral index image
`*spw##_##_##_##.cont.I.tt0.*`	An image conmtaining the zeroth Taylor term for a continuum image
`*spw##_##_##_##.cont.I.tt1.*`	An image conmtaining the first Taylor term for a continuum image
`*spw##_##_##_##.cont.IQUV.*`	An aggregate bandwidth or continuum full Stokes cube

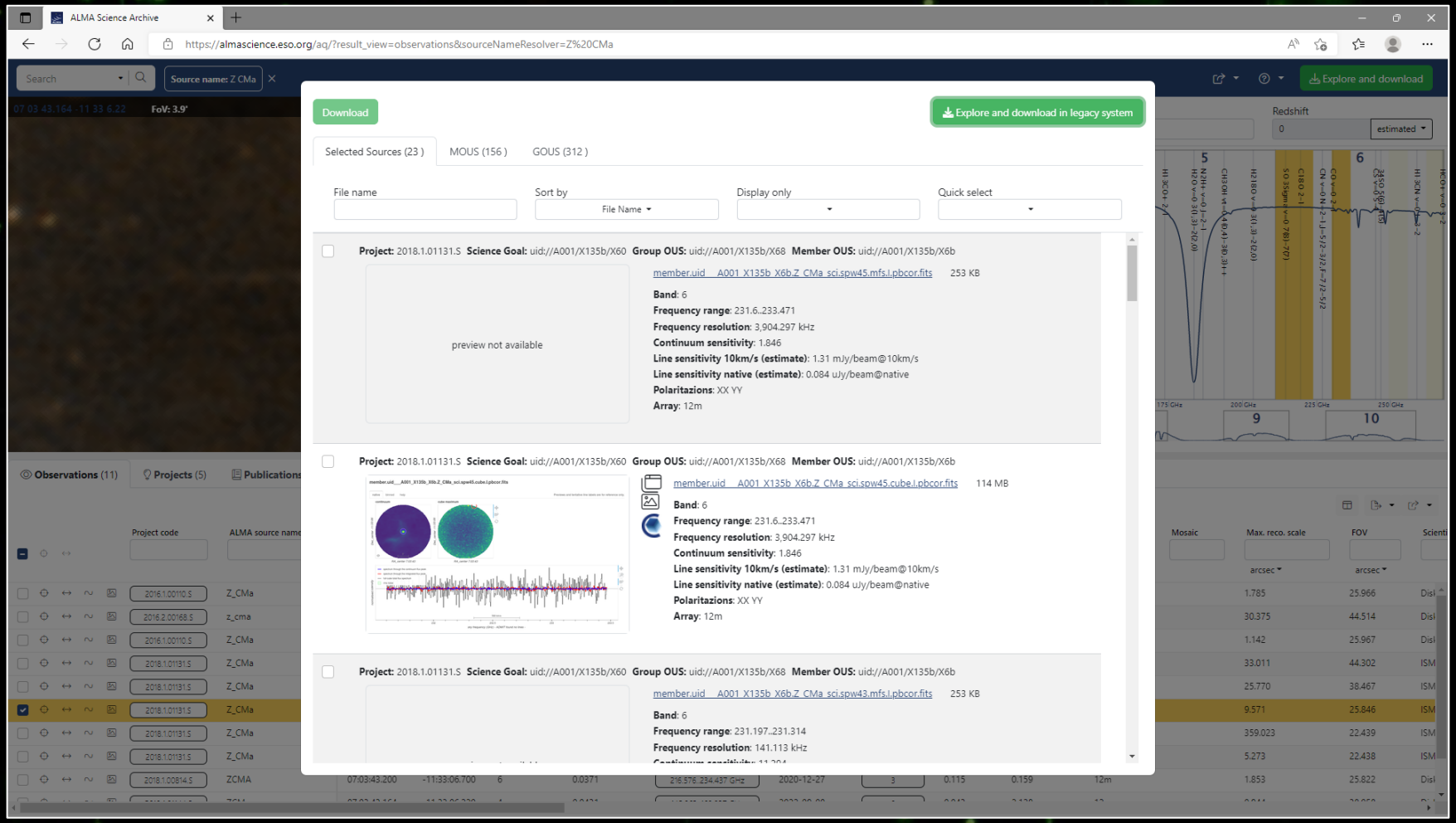
After selecting data, click on the download button at the top right of the screen to initiate a download.



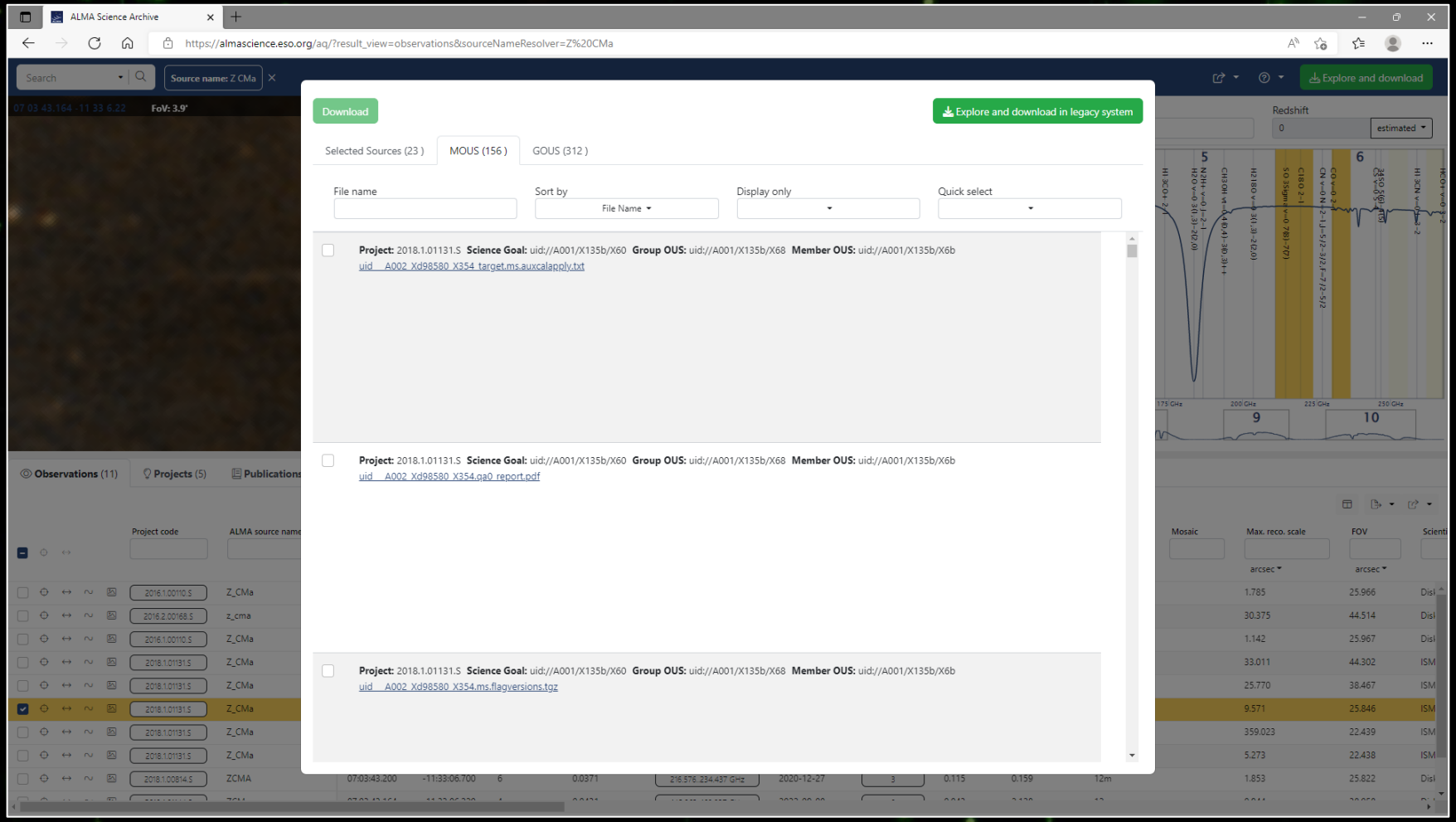
The updated version of the download interface will appear as an inset window containing solely files related to the selected dataset. (Note that the download interface has recently undergone significant changes.)



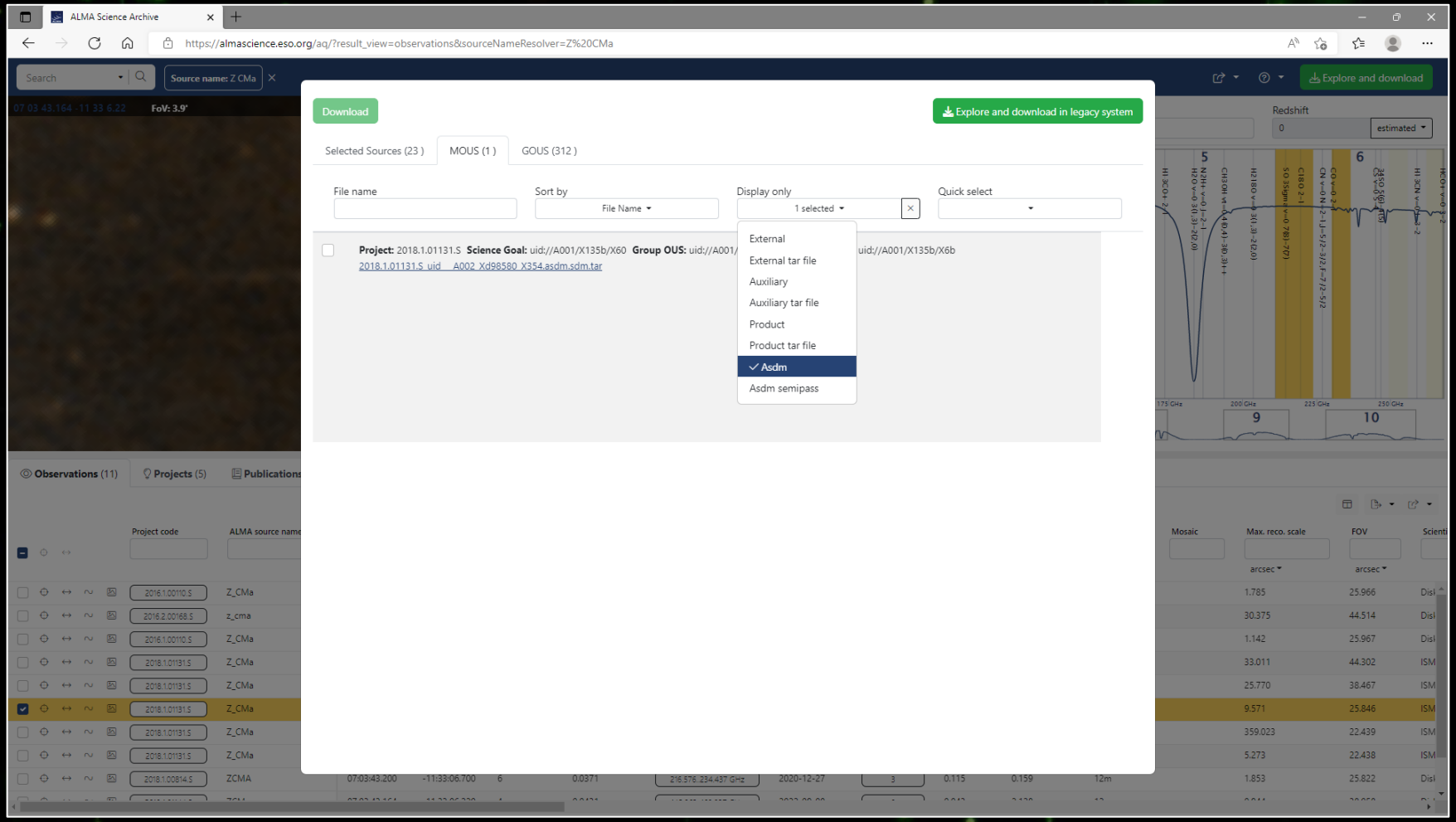
The default view will show the Selected Sources tab, which lists the images in the archive.



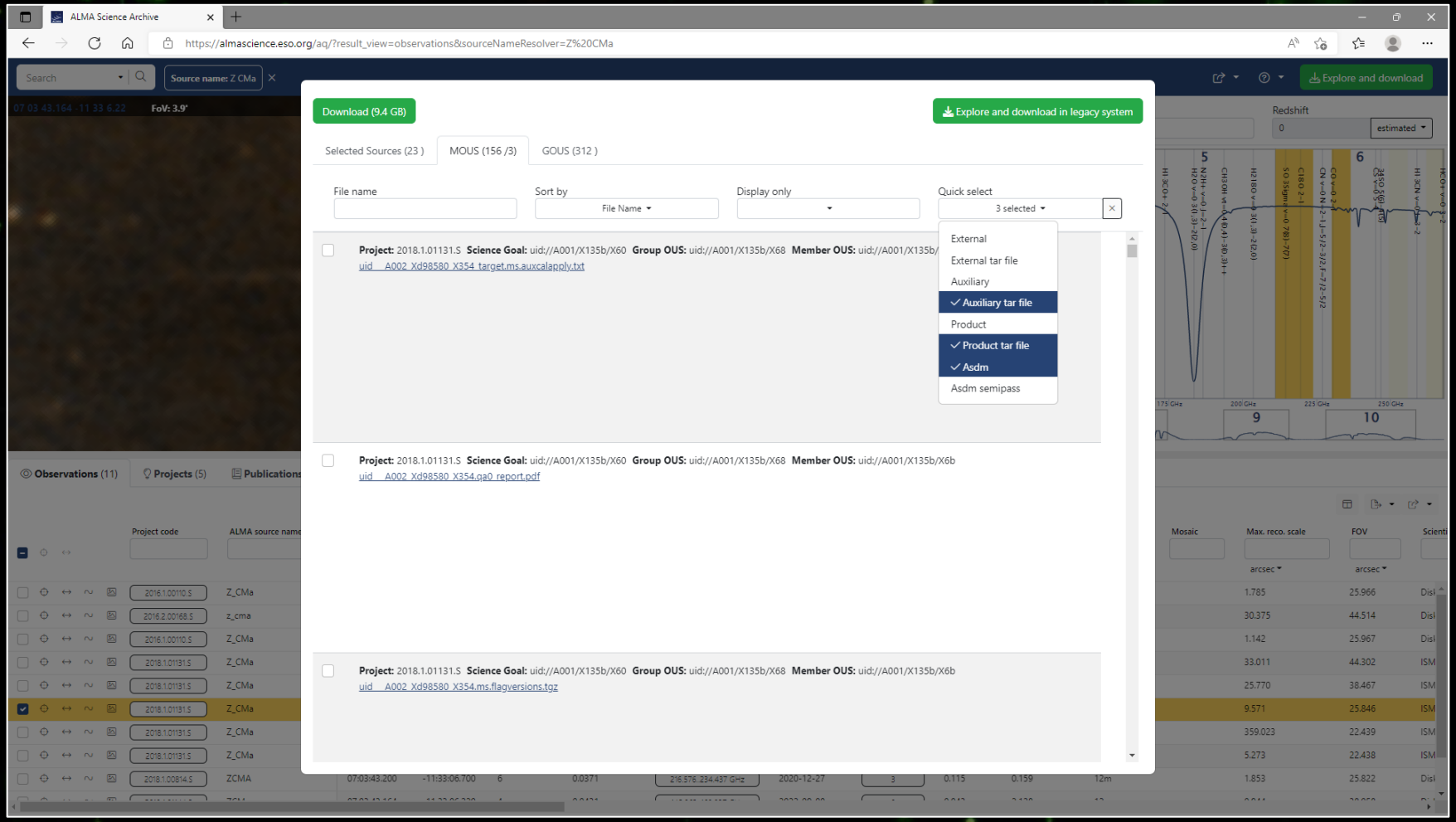
Clicking on the MOUS tab will show the data within the individual Member OUS (or Scheduling Block). Clicking on the GOUS tab will show all files within a Group OUS.



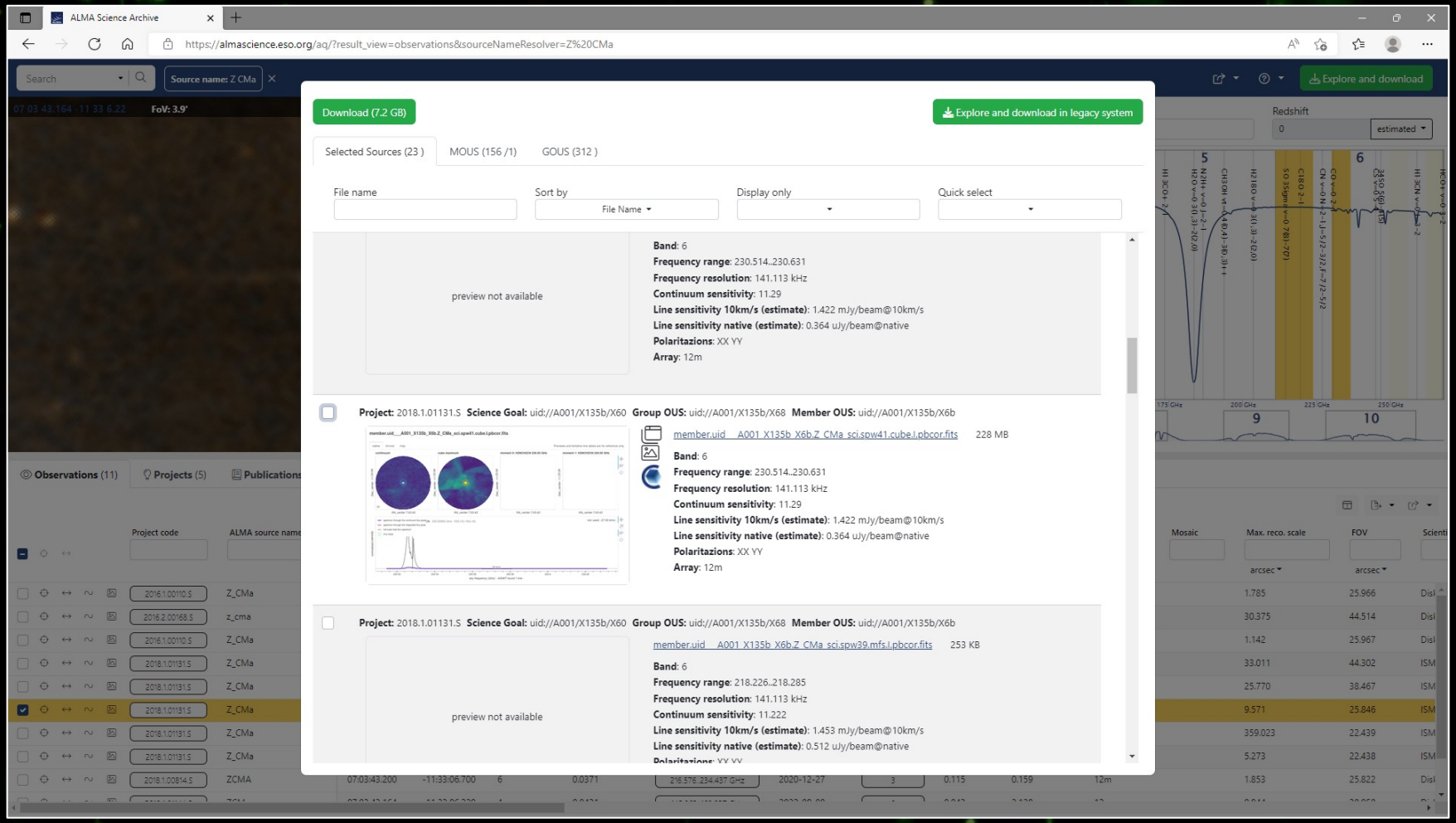
It is possible to filter the search results to show only specific file types, which is most useful when viewing the results from the MOUS and GOUS tabs.



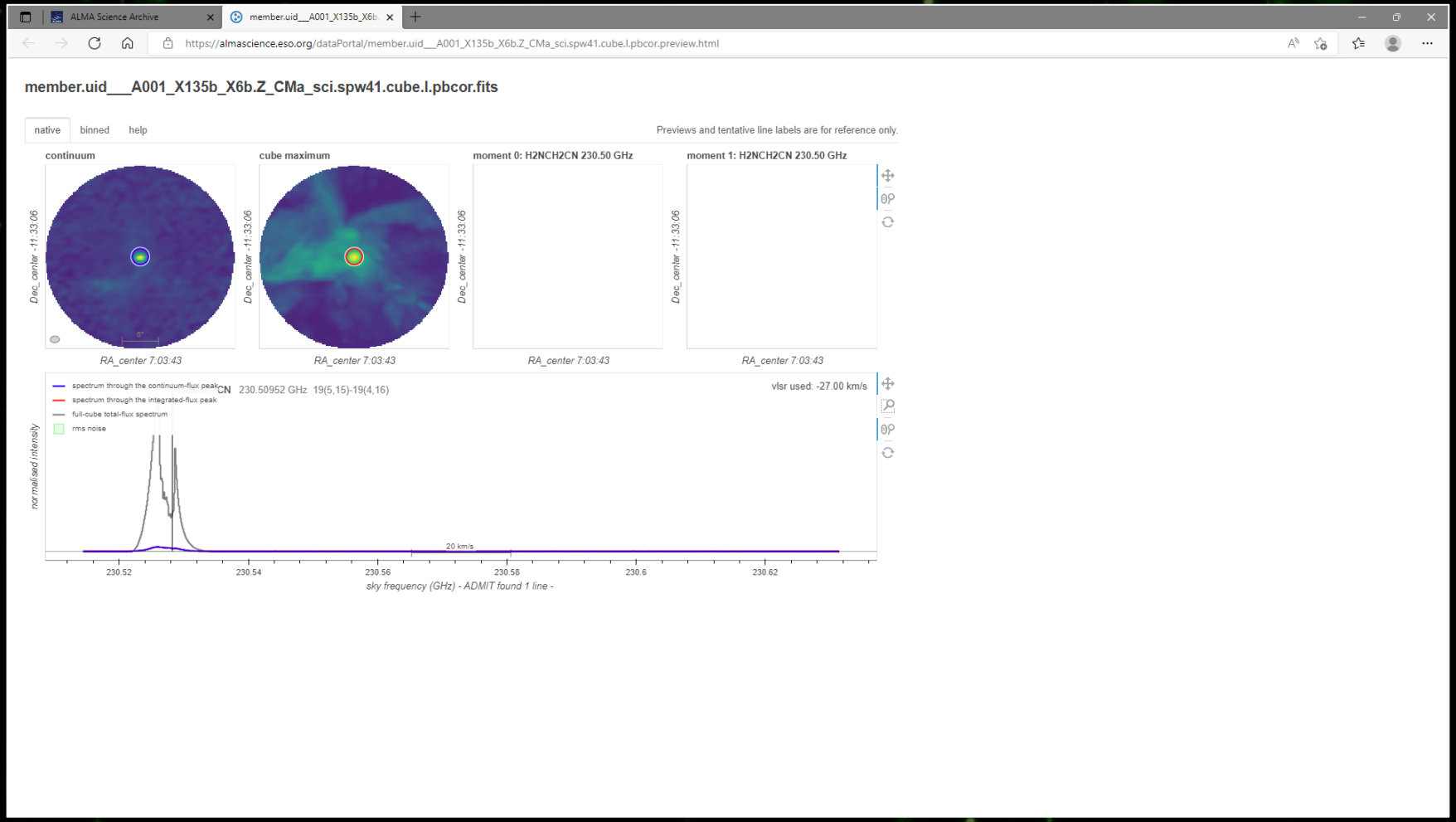
It is also possible to select one or more file types for download. This can be particularly useful for quickly selecting and downloading data for re-imaging or analysis.



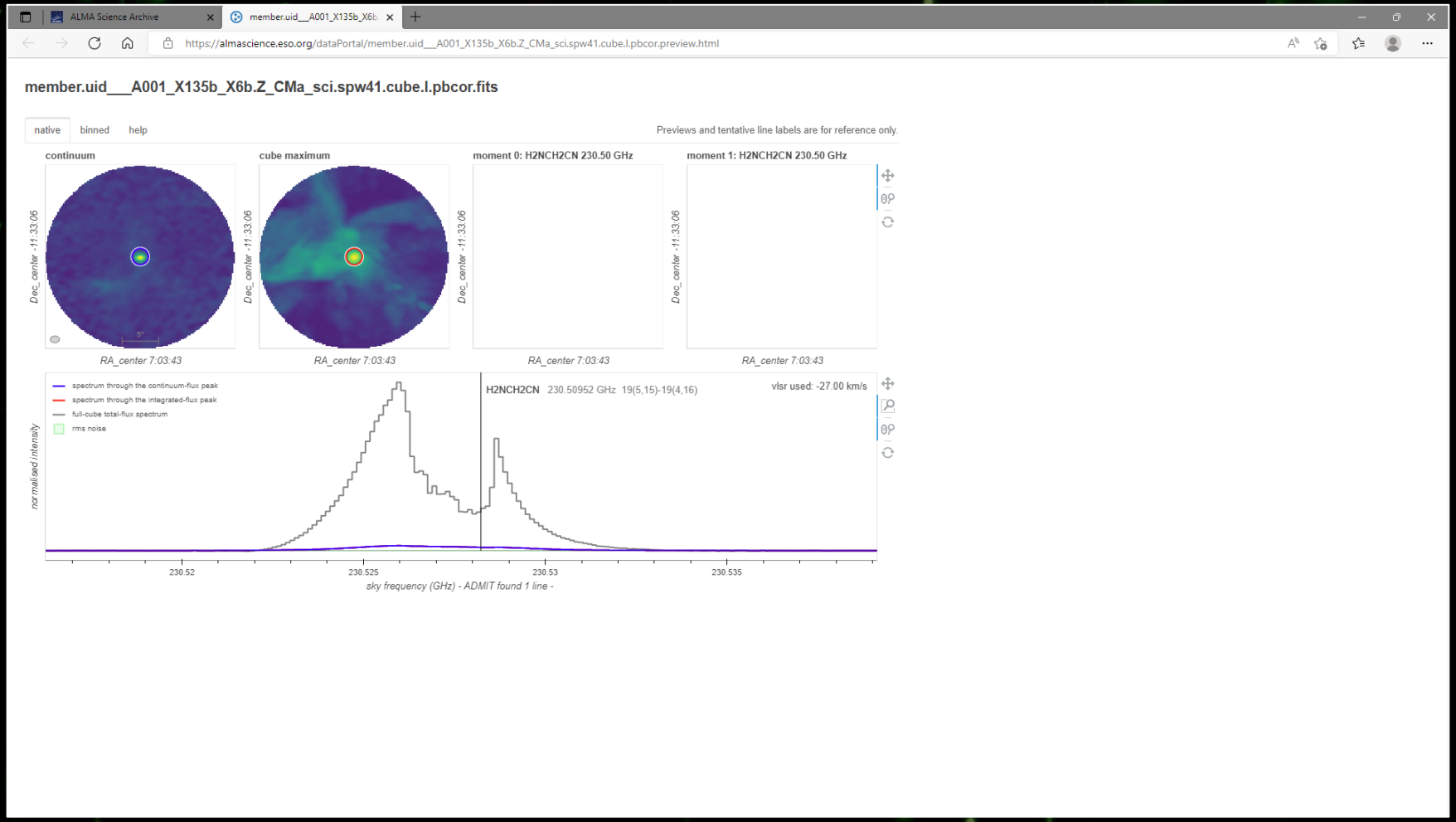
The new interface also provides a number of preview options, including displaying various interactive displays of images and extracted spectra as well as options to display images in CARTA.



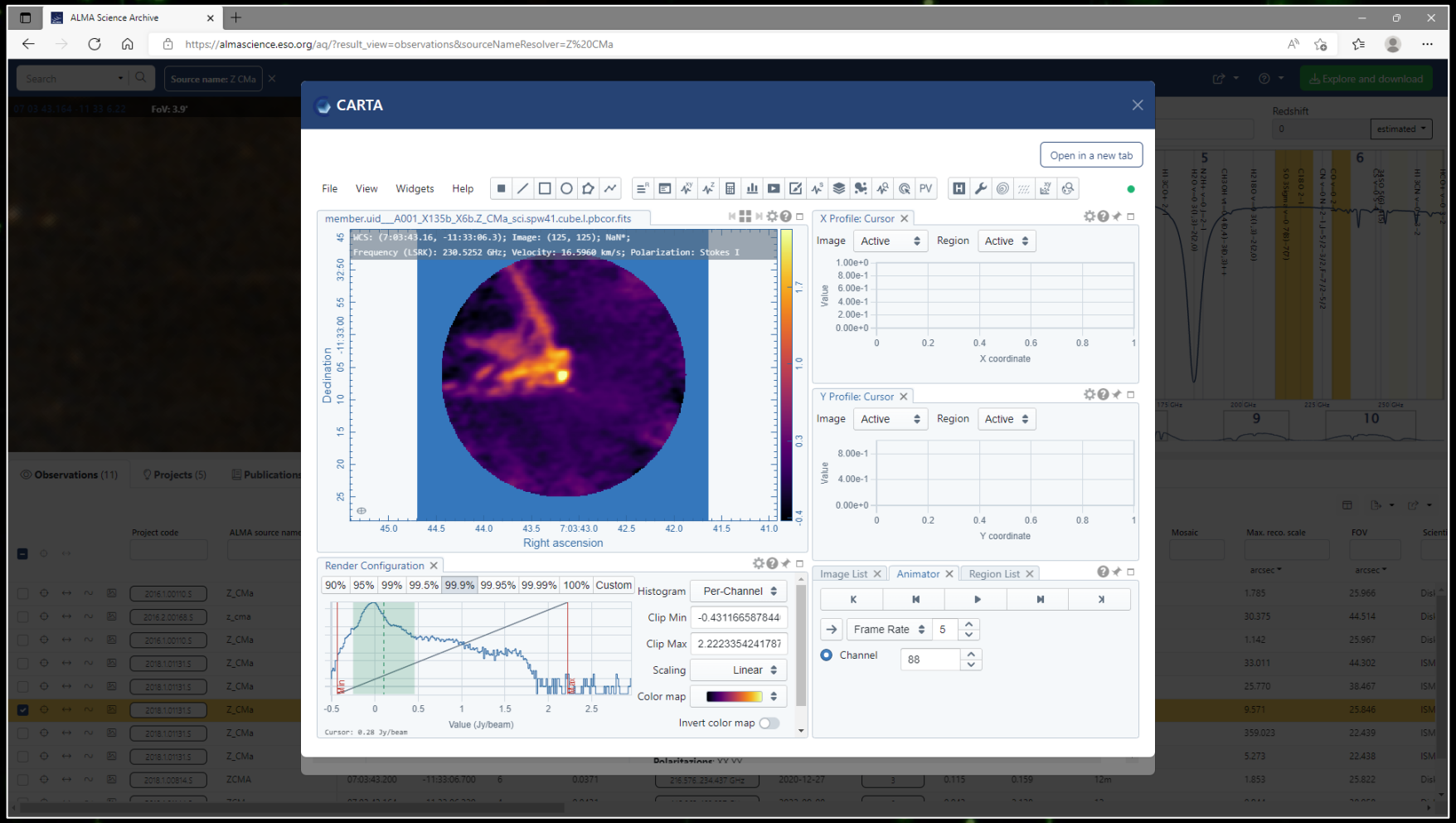
The new interface also provides a number of preview options, including displaying various interactive displays of images and extracted spectra as well as options to display images in CARTA.



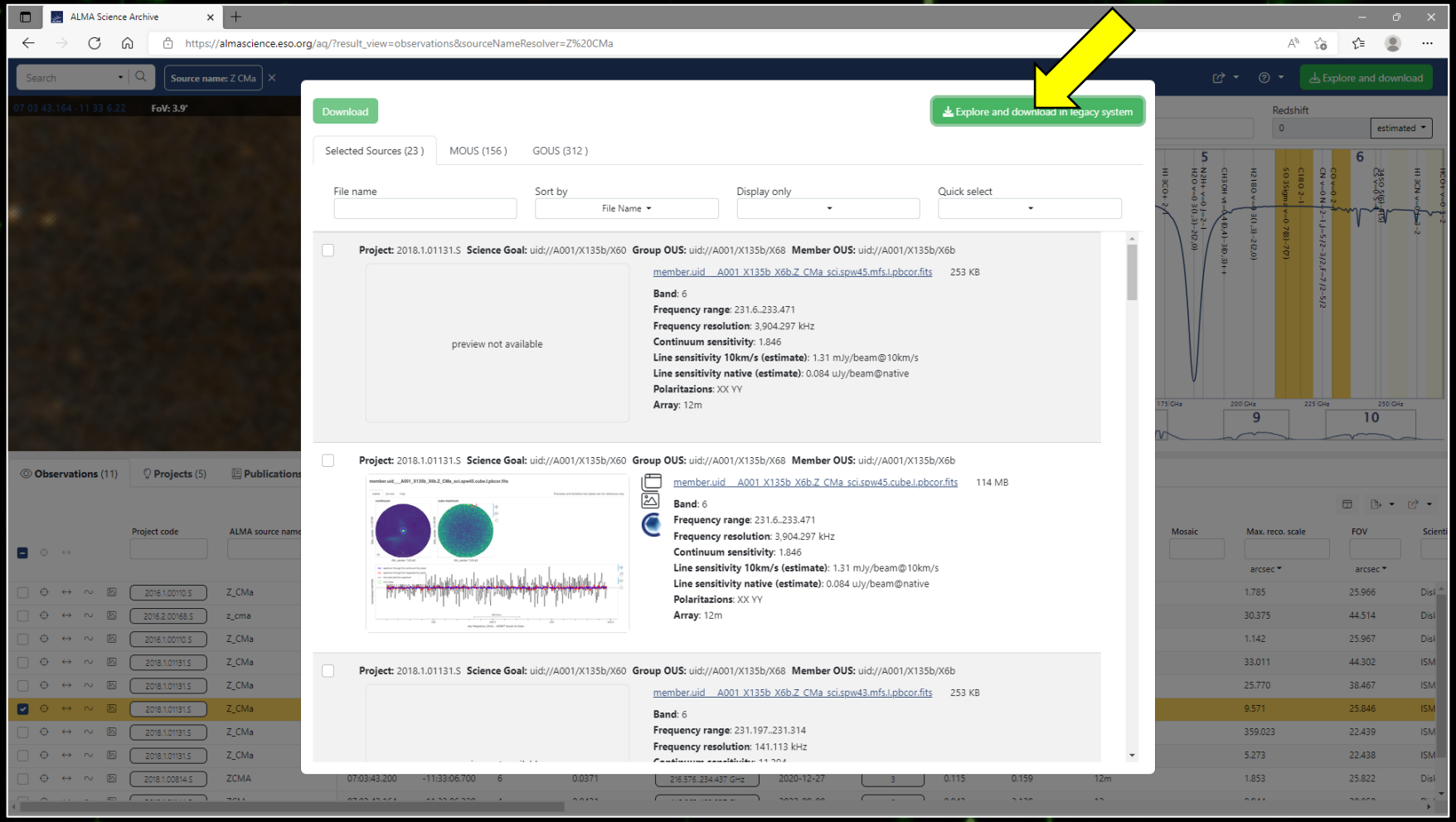
The new interface also provides a number of preview options, including displaying various interactive displays of images and extracted spectra as well as options to display images in CARTA.



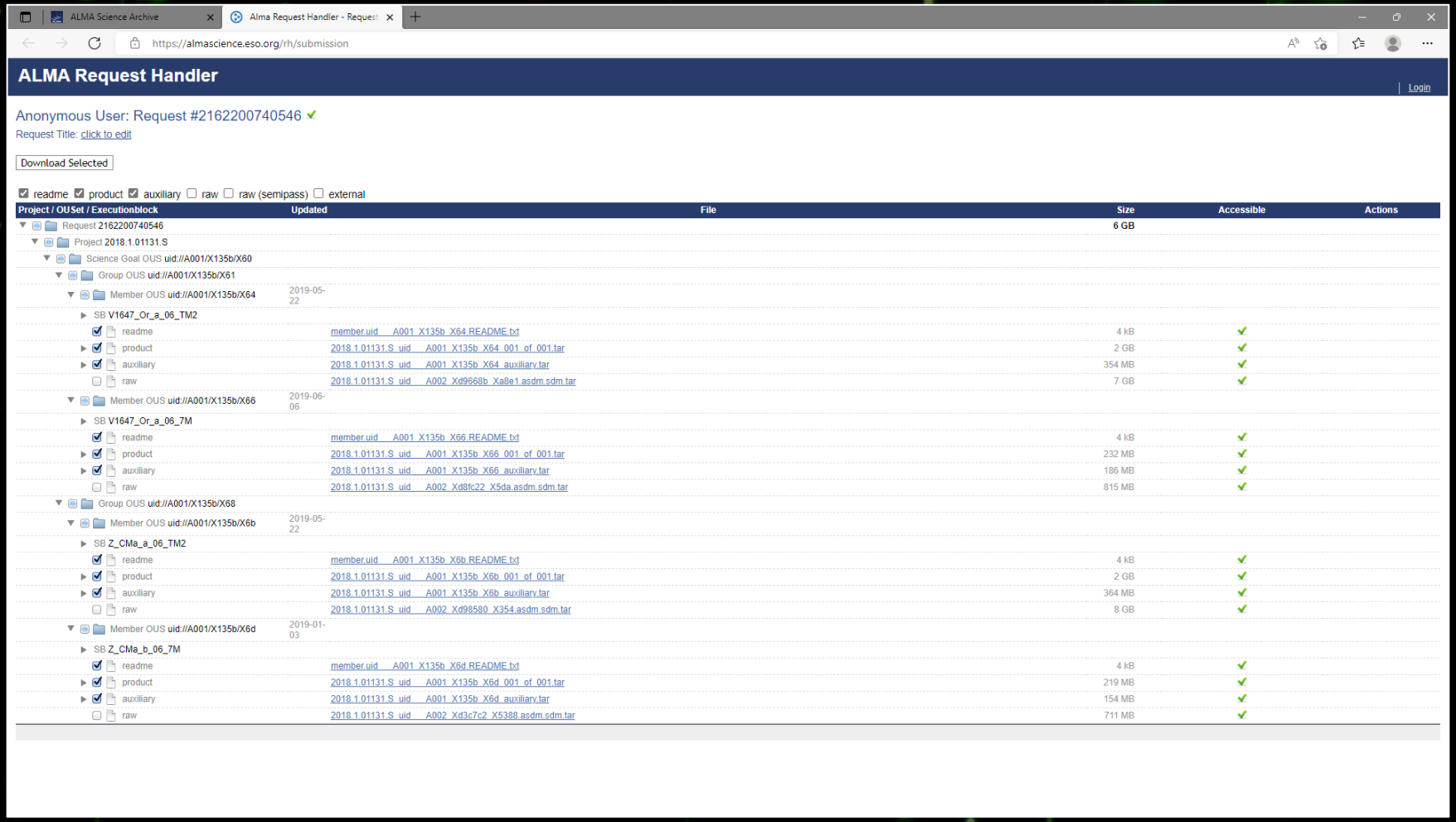
The new interface also provides a number of preview options, including displaying various interactive displays of images and extracted spectra as well as options to display images in CARTA.



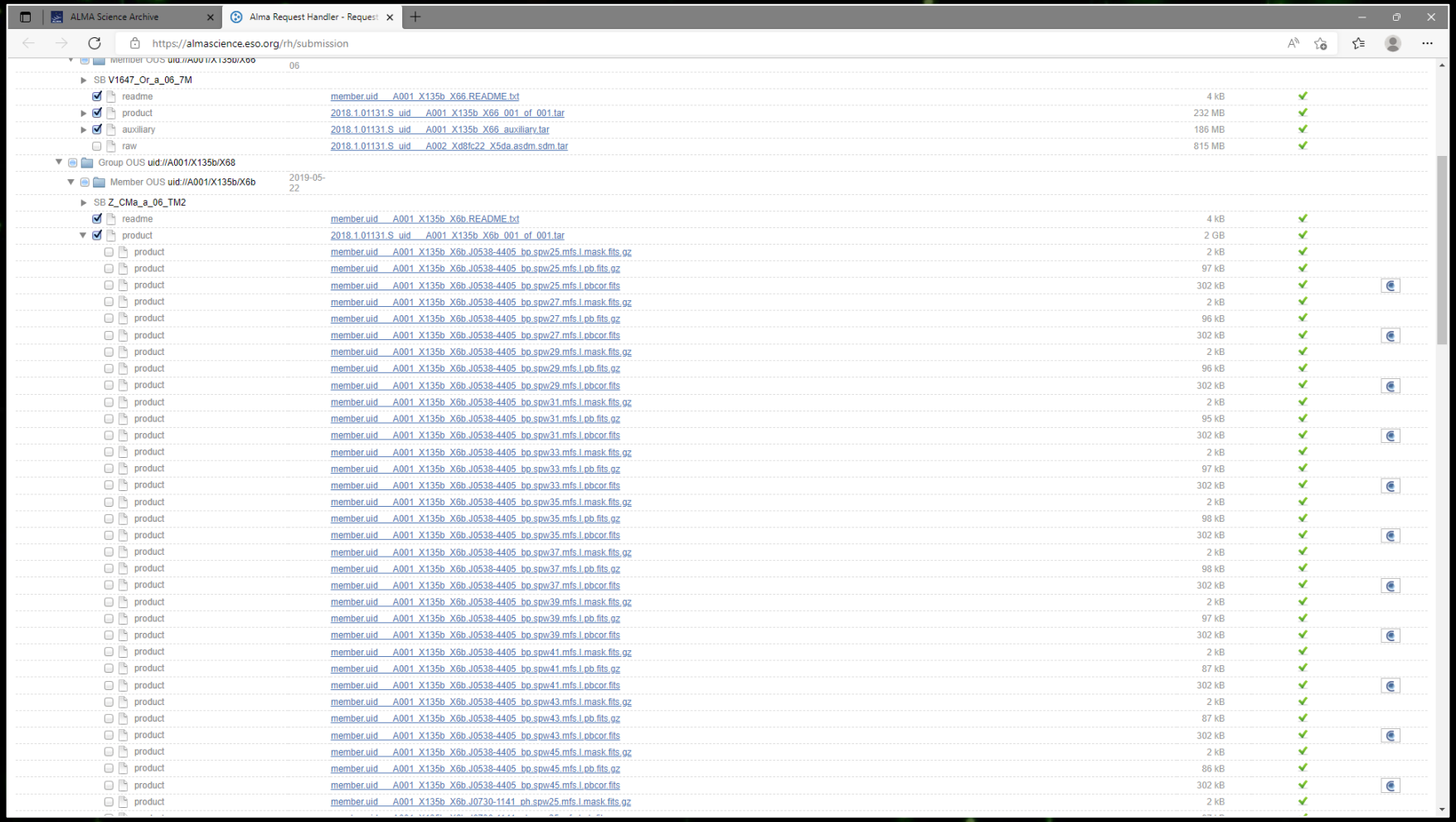
The old download interface is currently still accessible by clicking on the button at the top right of the current download interface.



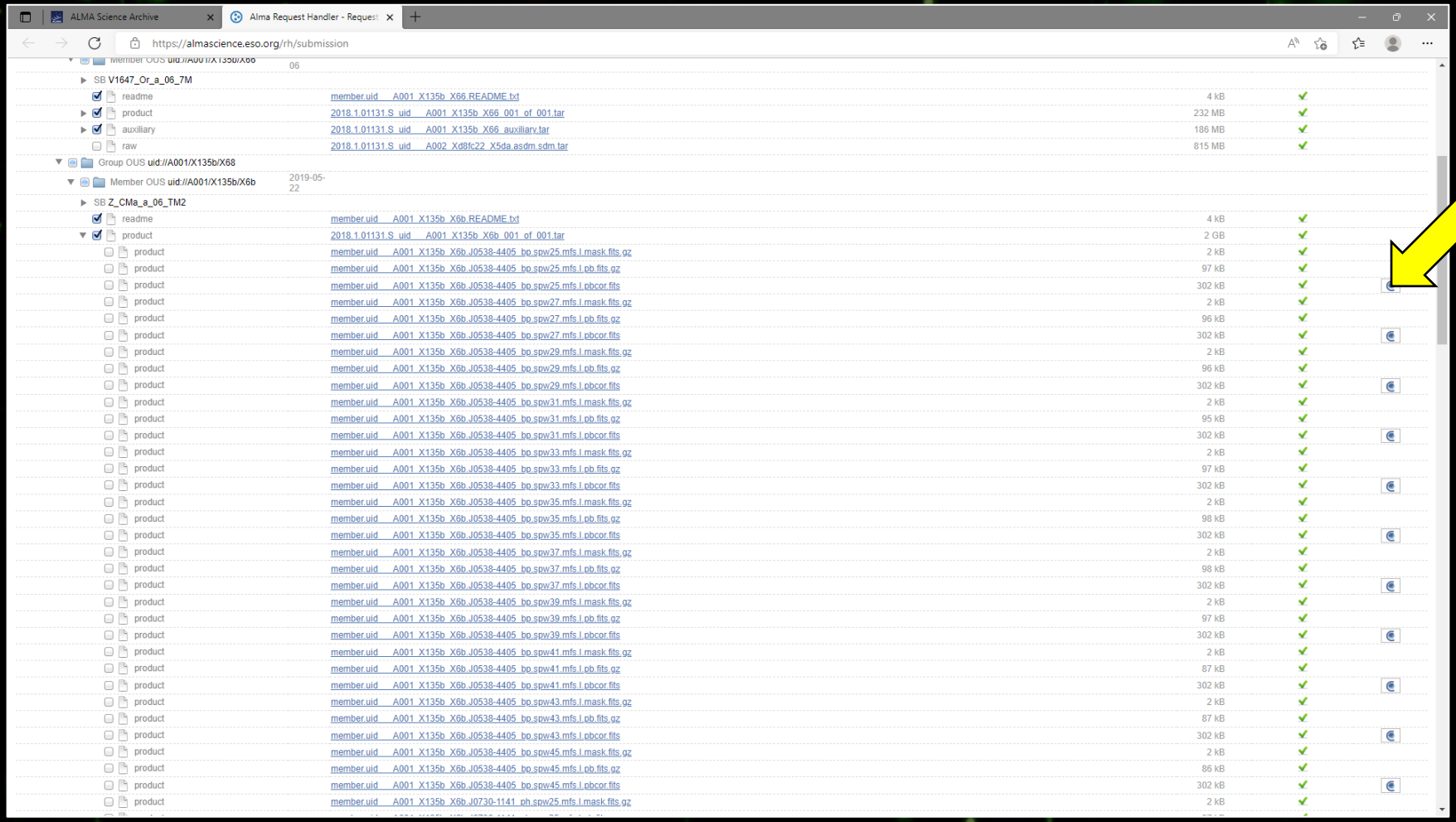
The old interface will show all of the data from the projects associated with the selected datasets. The list is hierarchical, which is good for understanding the data organization. However, for programs with multiple Scheduling Blocks, it can be hard to pick out a single dataset of interest.



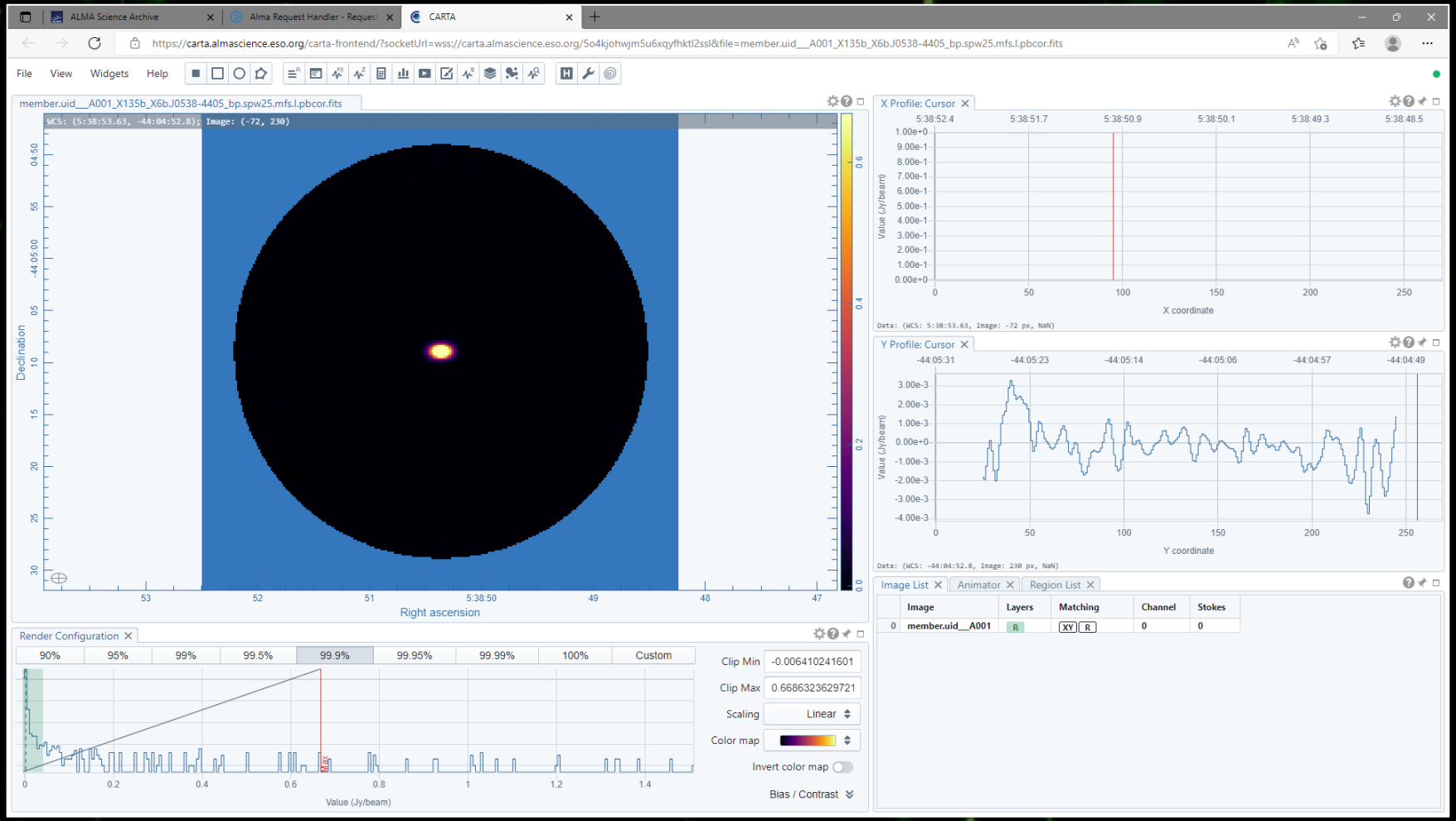
Clicking on the triangles pointing to the right will produce an expanded list that will show the contents of the individual tar files. (These lists can be collapsed again by clicking on the downwards-pointing arrows.)



It is also possible to preview individual images in the product tar file by clicking on the symbol with the C on the far right of the page. This will open a new page displaying the image using the CARTA interface.



It is also possible to preview individual images in the product tar file by clicking on the symbol with the C on the far right of the page. This will open a new page displaying the image using the CARTA interface.



Clicking on the download option in the current download interface will provide a download script.

The old download interface offers the options to download multiple files using a download script or a file list.

The individual results in either download interface page include links that can also be clicked on to download the individual files.

The download script can be executed in a Linux/Mac console to download data. The file must be made executable using `chmod` before doing this. When the file is executed, the data (in the form of tar files) will be downloaded to the current directory.

- If the script is interrupted, it is possible to restart the downloads from where they were stopped by restarting the script.
- When the download is complete, the data can be optionally be unpacked.

Data can also be downloaded in Python.

Here is an example using pyVO:

```
>>> datalink = pyvo.dal.adhoc.DatalinkResults.from_result_url(
  f"https://almascience.eso.org/datalink/sync?ID={'uid___A001_X135b_X6b'}")
>>> for dl in datalink:
  dl.cachedataset(filename=os.path.basename(dl['access_url']))
```

Here is an example using astroquery:

```
>>> mous=almaQuery['member_ous_uid'][0]
>>> linkList=get_data_info(mous,expand_tarfiles=True)
>>> downloadAlma=Alma()
>>> downloadAlma.login('username')
>>> downloadAlma.cache_location='/directory/'
>>> downloadAlma.retrieve_data_from_uid(mous)
```

One of the Jupyter notebooks at https://almascience.eso.org/alma-data/archive/archive-notebooks provides additional instructions and examples on using these tools.

When archival data are downloaded and unpacked, the files will be sorted into a directory structure that looks like the following:

```
2018.1.01131.S
    science_goal.uid___A001_X135b_X60
        group.uid___A001_X135b_X68
            member.uid___A001_X135b_X6b
                calibration
                log
                product
                qa
                raw
                script
```

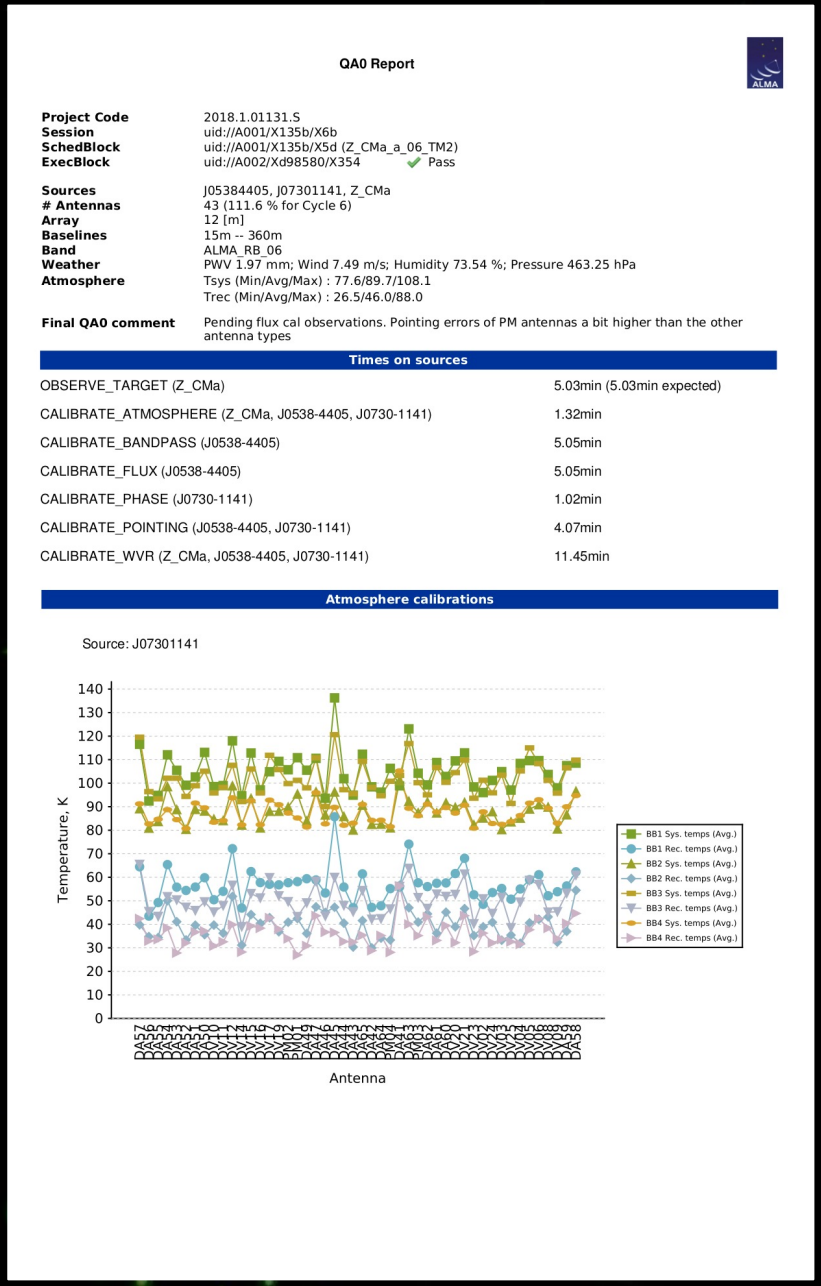
The directories contain the following files:

`calibration`	Calibration plots and tables
`log`	Log files
`product`	Fully processed images
`qa`	Quality assurance data
`raw`	Raw data (ASDM format)
`README`	A text file with information from calibration and imaging as well as general file information
`script`	Data processing scripts

The quality assurance directory contains information on the quality of the data. This may include:

- **QA0 report** (pdf file)
- **QA2 report** (pdf file)
- **QA2 png files** (labelled `qa2_part#.png`)
- **QA2 text files** (labelled `*textfile.txt`)
- **WebLogs** (within a .tgz file)

Other specialized files are created for single dish and polarization data.



QA0 Report

Project Code	2018.1.01131.S
Session	uid://A001/X135b/X6b
SchedBlock	uid://A001/X135b/X5d (Z_CMa_a_06_TM2)
ExecBlock	uid://A002/Xd98580/X354 ✔ Pass
Sources	J05384405, J07301141, Z_CMa
# Antennas	43 (111.6 % for Cycle 6)
Array	12 [m]
Baselines	15m -- 360m
Band	ALMA_RB_06
Weather	PWV 1.97 mm; Wind 7.49 m/s; Humidity 73.54 %; Pressure 463.25 hPa
Atmosphere	Tsys (Min/Avg/Max) : 77.6/89.7/108.1 Trec (Min/Avg/Max) : 26.5/46.0/88.0
Final QA0 comment	Pending flux cal observations. Pointing errors of PM antennas a bit higher than the other antenna types

Times on sources

OBSERVE_TARGET (Z_CMa)	5.03min (5.03min expected)
CALIBRATE_ATMOSPHERE (Z_CMa, J0538-4405, J0730-1141)	1.32min
CALIBRATE_BANDPASS (J0538-4405)	5.05min
CALIBRATE_FLUX (J0538-4405)	5.05min
CALIBRATE_PHASE (J0730-1141)	1.02min
CALIBRATE_POINTING (J0538-4405, J0730-1141)	4.07min
CALIBRATE_WVR (Z_CMa, J0538-4405, J0730-1141)	11.45min

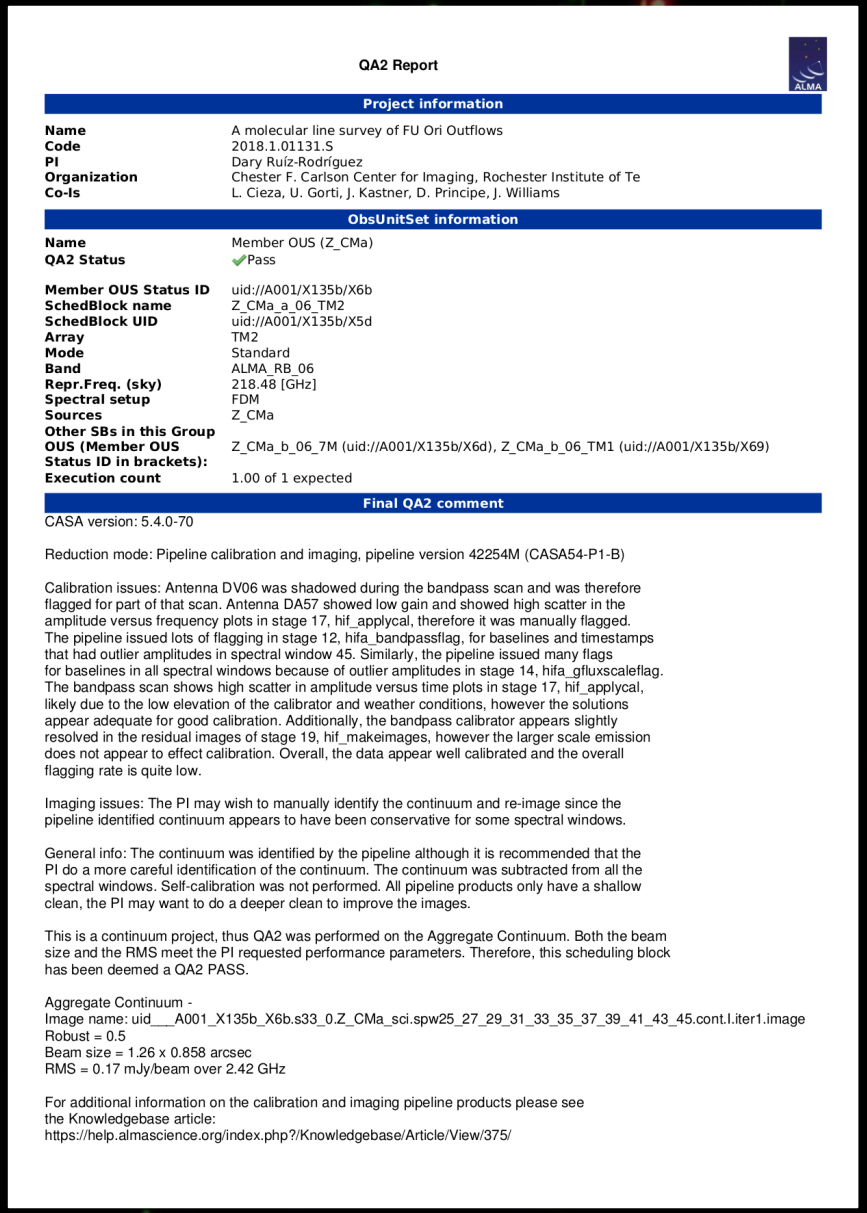
Atmosphere calibrations

Source: J07301141

The quality assurance directory contains information on the quality of the data. This may include:

- **QA0 report** (pdf file)
- **QA2 report** (pdf file)
- **QA2 png files** (labelled `qa2_part#.png`)
- **QA2 text files** (labelled `*textfile.txt`)
- **WebLogs** (within a .tgz file)

Other specialized files are created for single dish and polarization data.



QA2 Report

Project information

Name	A molecular line survey of FU Ori Outflows
Code	2018.1.01131.S
PI	Dary Ruíz-Rodríguez
Organization	Chester F. Carlson Center for Imaging, Rochester Institute of Te
Co-Is	L. Cieza, U. Gorti, J. Kastner, D. Principe, J. Williams

ObsUnitSet information

Name	Member OUS (Z_CMa)
QA2 Status	Pass
Member OUS Status ID	uid://A001/X135b/X6b
SchedBlock name	Z_CMa_a_06_TM2
SchedBlock UID	uid://A001/X135b/X5d
Array	TM2
Mode	Standard
Band	ALMA_RB_06
Repr.Freq. (sky)	218.48 [GHz]
Spectral setup	FDM
Sources	Z_CMa
Other SBs in this Group OUS (Member OUS Status ID in brackets):	Z_CMa_b_06_7M (uid://A001/X135b/X6d), Z_CMa_b_06_TM1 (uid://A001/X135b/X69)
Execution count	1.00 of 1 expected

Final QA2 comment

CASA version: 5.4.0-70

Reduction mode: Pipeline calibration and imaging, pipeline version 42254M (CASA54-P1-B)

Calibration issues: Antenna DV06 was shadowed during the bandpass scan and was therefore flagged for part of that scan. Antenna DA57 showed low gain and showed high scatter in the amplitude versus frequency plots in stage 17, hif_applycal, therefore it was manually flagged. The pipeline issued lots of flagging in stage 12, hifa_bandpassflag, for baselines and timestamps that had outlier amplitudes in spectral window 45. Similarly, the pipeline issued many flags for baselines in all spectral windows because of outlier amplitudes in stage 14, hifa_gfluxscaleflag. The bandpass scan shows high scatter in amplitude versus time plots in stage 17, hif_applycal, likely due to the low elevation of the calibrator and weather conditions, however the solutions appear adequate for good calibration. Additionally, the bandpass calibrator appears slightly resolved in the residual images of stage 19, hif_makeimages, however the larger scale emission does not appear to effect calibration. Overall, the data appear well calibrated and the overall flagging rate is quite low.

Imaging issues: The PI may wish to manually identify the continuum and re-image since the pipeline identified continuum appears to have been conservative for some spectral windows.

General info: The continuum was identified by the pipeline although it is recommended that the PI do a more careful identification of the continuum. The continuum was subtracted from all the spectral windows. Self-calibration was not performed. All pipeline products only have a shallow clean, the PI may want to do a deeper clean to improve the images.

This is a continuum project, thus QA2 was performed on the Aggregate Continuum. Both the beam size and the RMS meet the PI requested performance parameters. Therefore, this scheduling block has been deemed a QA2 PASS.

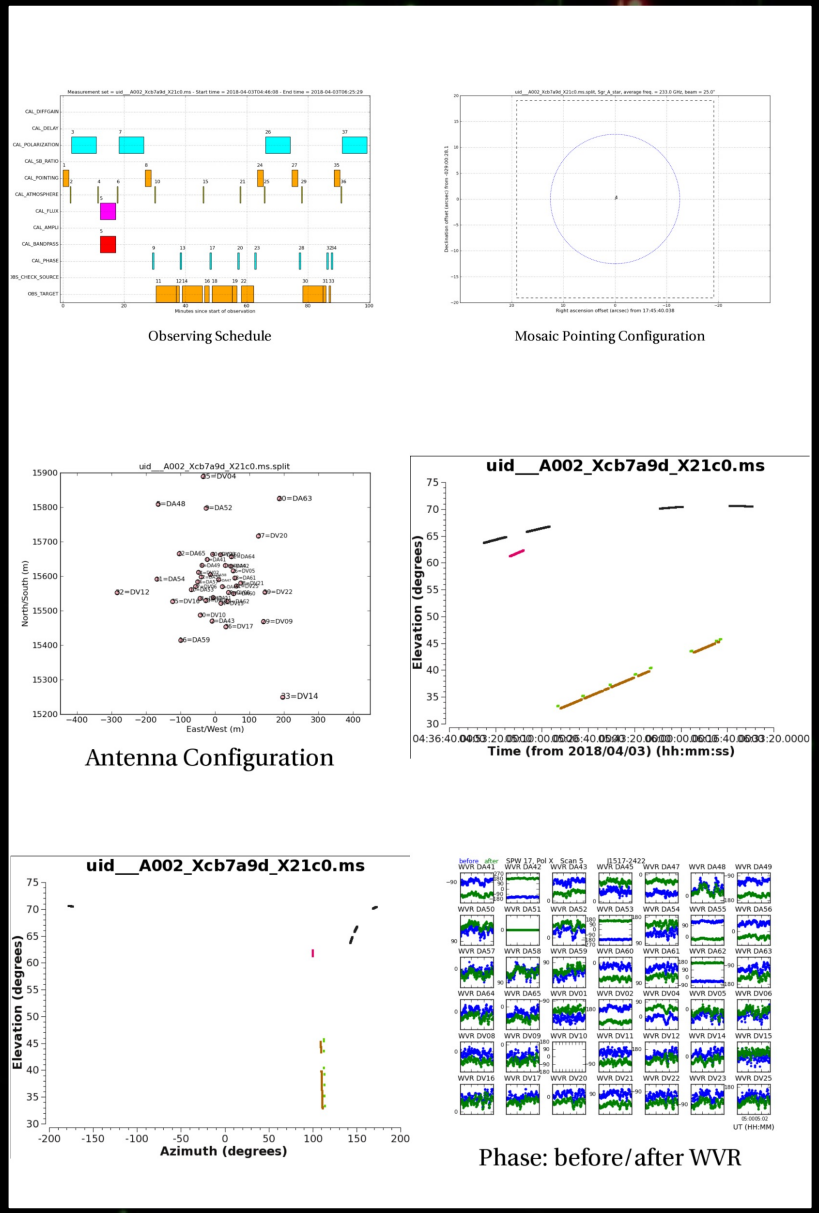
Aggregate Continuum -
Image name: uid___A001_X135b_X6b.s33_0.Z_CMa_sci.spw25_27_29_31_33_35_37_39_41_43_45.cont.I.iter1.image
Robust = 0.5
Beam size = 1.26 x 0.858 arcsec
RMS = 0.17 mJy/beam over 2.42 GHz

For additional information on the calibration and imaging pipeline products please see the Knowledgebase article:
https://help.almascience.org/index.php?/Knowledgebase/Article/View/375/

The quality assurance directory contains information on the quality of the data. This may include:

- **QA0 report** (pdf file)
- **QA2 report** (pdf file)
- **QA2 png files** (labelled `qa2_part#.png`)
- **QA2 text files** (labelled `*textfile.txt`)
- **WebLogs** (within a .tgz file)

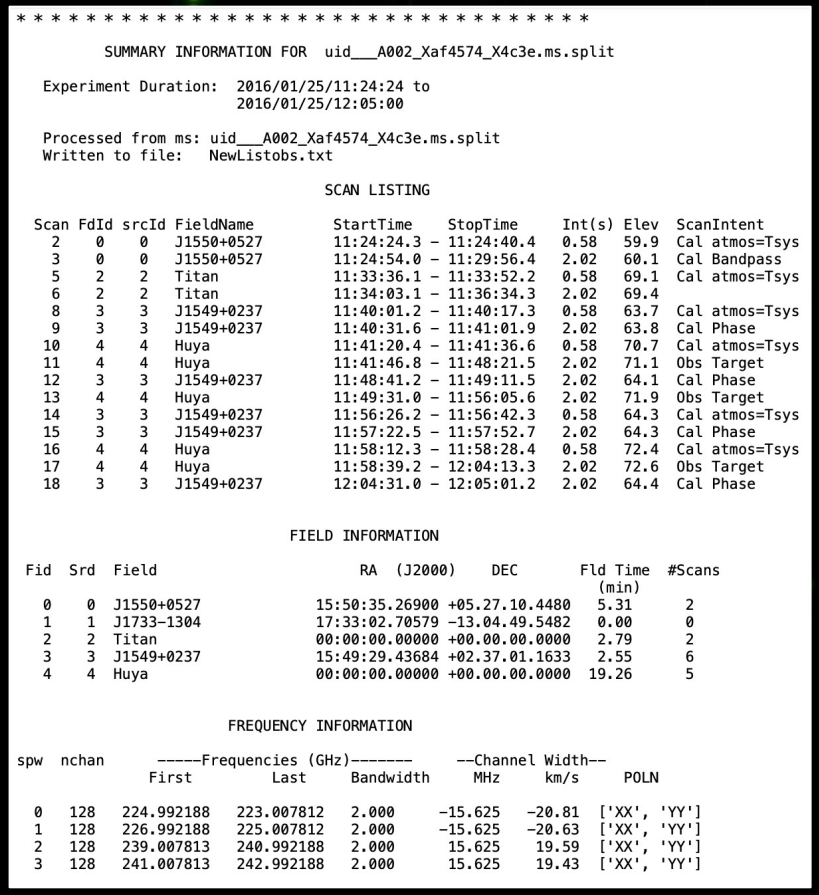
Other specialized files are created for single dish and polarization data.



The quality assurance directory contains information on the quality of the data. This may include:

- **QA0 report** (pdf file)
- **QA2 report** (pdf file)
- **QA2 png files** (labelled `qa2_part#.png`)
- **QA2 text files** (labelled `*textfile.txt`)
- **WebLogs** (within a .tgz file)

Other specialized files are created for single dish and polarization data.

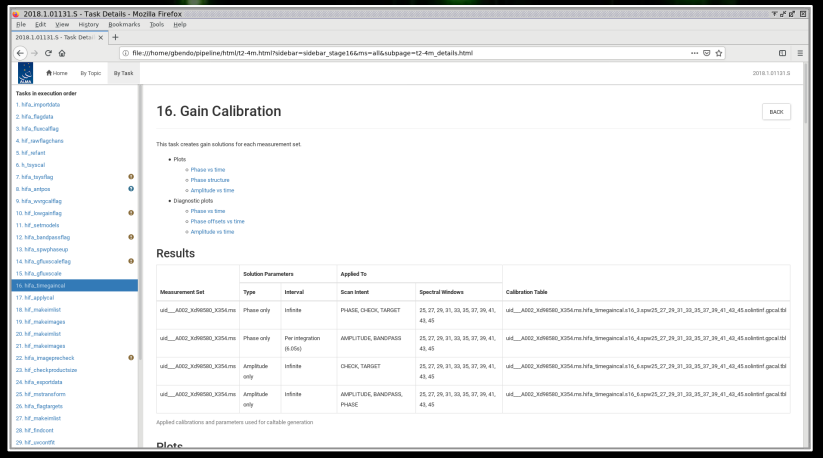


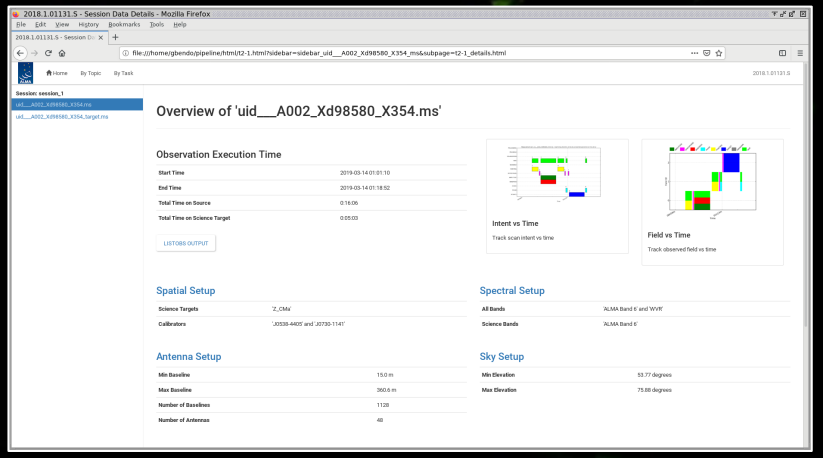
```
**********************************

        SUMMARY INFORMATION FOR  uid___A002_Xaf4574_X4c3e.ms.split

  Experiment Duration:  2016/01/25/11:24:24 to
                        2016/01/25/12:05:00

  Processed from ms: uid___A002_Xaf4574_X4c3e.ms.split
  Written to file:   NewListobs.txt

                                SCAN LISTING

 Scan FdId srcId FieldName         StartTime    StopTime     Int(s) Elev  ScanIntent
   2    0    0   J1550+0527        11:24:24.3 - 11:24:40.4   0.58   59.9  Cal atmos=Tsys
   3    0    0   J1550+0527        11:24:54.0 - 11:29:56.4   2.02   60.1  Cal Bandpass
   5    2    2   Titan             11:33:36.1 - 11:33:52.2   0.58   69.1  Cal atmos=Tsys
   6    2    2   Titan             11:34:03.1 - 11:36:34.3   2.02   69.4
   8    3    3   J1549+0237        11:40:01.2 - 11:40:17.3   0.58   63.7  Cal atmos=Tsys
   9    3    3   J1549+0237        11:40:31.6 - 11:41:01.9   2.02   63.8  Cal Phase
  10    4    4   Huya              11:41:20.4 - 11:41:36.6   0.58   70.7  Cal atmos=Tsys
  11    4    4   Huya              11:41:46.8 - 11:48:21.5   2.02   71.1  Obs Target
  12    3    3   J1549+0237        11:48:41.2 - 11:49:11.5   2.02   64.1  Cal Phase
  13    4    4   Huya              11:49:31.0 - 11:56:05.6   2.02   71.9  Obs Target
  14    3    3   J1549+0237        11:56:26.2 - 11:56:42.3   0.58   64.3  Cal atmos=Tsys
  15    3    3   J1549+0237        11:57:22.5 - 11:57:52.7   2.02   64.3  Cal Phase
  16    4    4   Huya              11:58:12.3 - 11:58:28.4   0.58   72.4  Cal atmos=Tsys
  17    4    4   Huya              11:58:39.2 - 12:04:13.3   2.02   72.6  Obs Target
  18    3    3   J1549+0237        12:04:31.0 - 12:05:01.2   2.02   64.4  Cal Phase


                              FIELD INFORMATION

 Fid  Srd  Field                      RA  (J2000)    DEC        Fld Time  #Scans
                                                                  (min)
   0    0  J1550+0527           15:50:35.26900 +05.27.10.4480    5.31      2
   1    1  J1733-1304           17:33:02.70579 -13.04.49.5482    0.00      0
   2    2  Titan                00:00:00.00000 +00.00.00.0000    2.79      2
   3    3  J1549+0237           15:49:29.43684 +02.37.01.1633    2.55      6
   4    4  Huya                 00:00:00.00000 +00.00.00.0000   19.26      5


                            FREQUENCY INFORMATION

spw  nchan      -----Frequencies (GHz)-------      --Channel Width--
                  First          Last      Bandwidth      MHz      km/s      POLN

  0   128   224.992188   223.007812   2.000      -15.625   -20.81  ['XX', 'YY']
  1   128   226.992188   225.007812   2.000      -15.625   -20.63  ['XX', 'YY']
  2   128   239.007813   240.992188   2.000       15.625    19.59  ['XX', 'YY']
  3   128   241.007813   242.992188   2.000       15.625    19.43  ['XX', 'YY']
```

The quality assurance directory contains information on the quality of the data. This may include:

- **QA0 report** (pdf file)
- **QA2 report** (pdf file)
- **QA2 png files** (labelled `qa2_part#.png`)
- **QA2 text files** (labelled `*textfile.txt`)
- **WebLogs** (within a .tgz file)

Other specialized files are created for single dish and polarization data.





To produce calibrated visibility data, which can be used to create new images, do the following in a terminal:

1. Go to the `script` directory.
2. Start CASA. For pipeline-calibrated data, start CASA in pipeline mode using the `--pipeline` option in the terminal. (When starting CASA from the app on a Mac, quitting CASA in the terminal and then restarting it with this option works.)
3. Execute the script using the command `execfile(´scriptForPI.py´)`.

Restoring may take a while, especially for large dataset or on machines with lower specifications.

The calibrated visibility data will be placed in a new directory called `calibrated`. These data can be used to create new images.