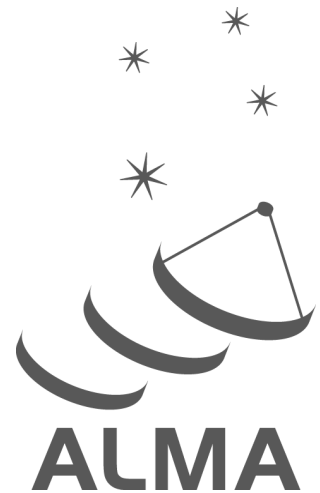


ALMA Science Archive Manual



www.almascience.org

User Support:

For further information or to comment on this document, please contact your regional Helpdesk through the ALMA User Portal at www.almascience.org . Helpdesk tickets will be redirected automatically to the nearest ALMA Regional Center at ESO, NAOJ or NRAO.

Revision History:

Version	Date	Editors
<hr/>		
4.1.0.x	March 2016	Felix Stoehr, Eric Murphy, Mark Lacy, Stéphane Leon, Erik Muller, Kouichiro Nakanishi
5.1.0.x	March 2017	Brenda Matthews, Felix Stoehr, Mark Lacy, Kouichiro Nakanishi, Stéphane Leon
6.1.0.x	March 2018	Brenda Matthews, Felix Stoehr, John Hibbard, Kouichiro Nakanishi, Kuo-Song Wang, Stéphane Leon
7.1.0.x	March 2019	Felix Stoehr, Brenda Matthews, Kuo-Song Wang, Adele Plunkett, Rie Miura
8.1.0.x	March 2020	Felix Stoehr, Brenda Matthews, Kuo-Song Wang, Adele Plunkett, Patricio Sanhueza

Contributors

In publications, please refer to this document as:

F. Stoehr et al. 2020, ALMA Science Archive Manual, ALMA Doc. 8.15 v1.0



Contents

1	Introduction	1
1.1	Quickstart	1
1.2	User support	1
2	Science Portal	3
2.1	Archive Data and Documents	3
2.2	Project access delegation	3
3	Science Archive	5
3.1	Content and policies	5
3.2	Querying the ALMA Science Archive	5
3.3	Browsing the results	7
3.4	Selecting the files for download	9
3.5	Downloading the data	11
3.6	Data packaging	11
3.6.1	Directory structure	11
3.6.2	Cycle 0	12
3.6.3	Cycle 1 and later	13
4	Programmatic access	15
4.1	astroquery	15
4.2	VO	15
5	Scientific exploitation of the data	17
6	Acknowledgement statement	19
7	Frequently asked questions	21

This page was intentionally left almost blank

Chapter 1

Introduction

1.1 Quickstart

Searching for and downloading ALMA data is done in a few easy steps:

1. Place search constraints into the Query Form (top left corner) at <https://almascience.org/asax>
2. Use the checkboxes to select observations of interest and hit "Request download" (top right)
3. Use the checkboxes to select files of interest and hit "Download selected"
4. Use one of the proposed download methods to save the selected files onto your computer

The functionalities offered by the ALMA Science Archive (ASA) for each of these four steps are detailed in sections 3.2, 3.3, 3.4 and 3.5, respectively.

1.2 User support

In each of the three partner regions an ALMA Regional Centre (ARC) has been established¹. Each ARC provides scientific user support in the core areas of proposal preparation (Phase 1), observation program preparation (Phase 2) and data analysis to users in their respective communities. The ARCs also host a full copy of the ALMA Archive and deliver data packages to users. The ARCs are the ALMA interfaces to the users.

Please do not hesitate to contact the ARCs or ARC nodes through the ALMA Helpdesk² for any question or comment. In particular we strongly encourage requests for face-to-face visits with the experts at one of the ARCs or ARC nodes for archival research purposes. All ARCs and ARC nodes or a cooperative department of the ARC do have computing facilities available.

¹<http://almascience.org/about-alma/alma-organization>

²<http://help.almascience.org>

Chapter 2

Science Portal

Each ARC maintains a full copy of the ALMA archive and hosts one of the three ALMA Science Portals (SP). From the toplevel link <https://www.almascience.org>, you will be automatically redirected to the SP of your closest ARC and your ALMA credentials are valid at each of them following the "single-sign-on" principle.

To obtain ALMA data, click on the "Data" entry in the top menu of the Science Portal. Here you will find links to the four types of ALMA data:

- Archive: ALMA data taken for PI programs
- Large Programs: Data products returned by the PIs for ALMA Large Programs
- Calibrator Catalogue: Information on calibration sources, including flux histories
- Science Verification: ALMA data taken for Science Verification

In addition, there is the top-menu item "Processing" which amongst other items contains information on the data processing performed by the Observatory and indications of the computing resources required to re-reduce data at your home institution should you wish to do so.

2.1 Archive Data and Documents

The Archive link directly takes you to the Science Archive Query Interface. The sub-menu "Archive documentation" shows "What's new", "Did you now?" and "Known issues" and also provides links to this documents, to the latest document containing a description of ALMA products as well as to a page listing the contents of the Cycle 0 deliveries¹.

Useful documents available from the Documents and Tools menu of the ALMA Science Portal include the ALMA Science Pipeline User's Guide, CASA documentation and descriptions of the ALMA quality assurance for data products ("QA2") for all cycles and the delivered data products under "ALMA QA2 Data Products" (see also Section 3.6).

2.2 Project access delegation

A PI can grant access rights to Co-Is or any other registered ALMA user directly from the ALMA Science Portal. To do this, the PI should authenticate using the "Log in" button in the top right corner of the Science Portal, click on their username and select "Profile" from the menu. Then select the "Project Delegation" tab and select the project to which you wish to grant another user proprietary access, and click on the "Add delegee" button at the bottom left of the page. Users who are granted access via the project delegation mechanism can log in to any of the three ARCs with their own account and can download the data from there. Data access is always

¹<http://almascience.org/alma-data/archive/delivery-list>

granted to entire projects rather than individual deliveries from a project. Delegating access to a project does not allow the delegee to grant project access to others.

Chapter 3

Science Archive

3.1 Content and policies

The ALMA Science Archive (ASA) contains all public and proprietary data of PI-proposal observations.

ALMA is producing currently roughly 500TB worth of raw-data and reduced data products per year. These data are stored at the Joint ALMA Observatory in Santiago, Chile and are copied from there to the three ARCs. Each ARC holds a complete copy of the ASA. (See chapter 14 of the Technical Handbook¹ for details.)

ALMA data from Regular and Large Programs are protected by a proprietary period of 12 months, after which they become public. The proprietary period of a DDT project data is 6 months. All public data can be accessed anonymously. Observational metadata (e.g., source names and positions, spectral setups, etc.) are not protected by a proprietary period and are available directly from the ASA as soon as the data have passed the first step of quality control, QA0.

Data from calibration observations carried out during PI observations have no proprietary period. As these observations are contained in the data files of the proprietary data, it is currently not possible to extract them in an automatic fashion. Should you be interested in calibrator data contained in proprietary observations, please write a ticket to the ALMA Helpdesk, and we will extract the data for you on a best-effort basis.

Please refer to the ALMA User's policies² for details.

3.2 Querying the ALMA Science Archive

The new ALMA Science Archive Query interface can be directly reached at <https://almascience.org/asax>.

The main input form appears when you hover over the search field in the top left corner. Sources can be queried for either by their name or by a search near a position on the sky specified in Right Ascension and Declination or Galactic coordinates. The name resolver uses Sesame³ to perform a semi-case-insensitive search for a source name. A search radius in degrees can be added to the end of the name search separated by a comma (e.g., "Arp220, 1"), where the default radius is 10 arcmin. Sesame (via NED, Simbad and Vizier) will parse names commonly found throughout the literature and return search coordinates. Sources can also be queried by their ALMA name, which is the name of the source as entered by the PI in the Observing Tool (OT). For this case, partial matches through wildcards (?, *) and boolean OR expressions ("|") can be used.

The query is executed instantaneously without you having to click on a search button or to hit return.

Various other constraints can be placed on a query to the ALMA Science Archive. As Figure 3.2 shows, they have been grouped into the following five categories: Position, Energy, Project, Publication, and Options.

¹<https://almascience.eso.org/documents-and-tools/latest/alma-technical-handbook>

²<https://almascience.eso.org/documents-and-tools/latest/alma-user-policies>

³<http://cds.u-strasbg.fr/cgi-bin/Sesame>

Autocompletion is enabled in all text fields except for the Source Name (Resolver) field, allowing for easier search. In addition to the single source search, a target list can be uploaded. Note that despite the suggestions from the autocompletion, the full set of operations (including wildcards and OR) is available.

Each query constraint comes with a tooltip help that provides a description, the units in case of queries for physical parameters, and examples. Clicking on an example will insert it directly into the query constraint field.

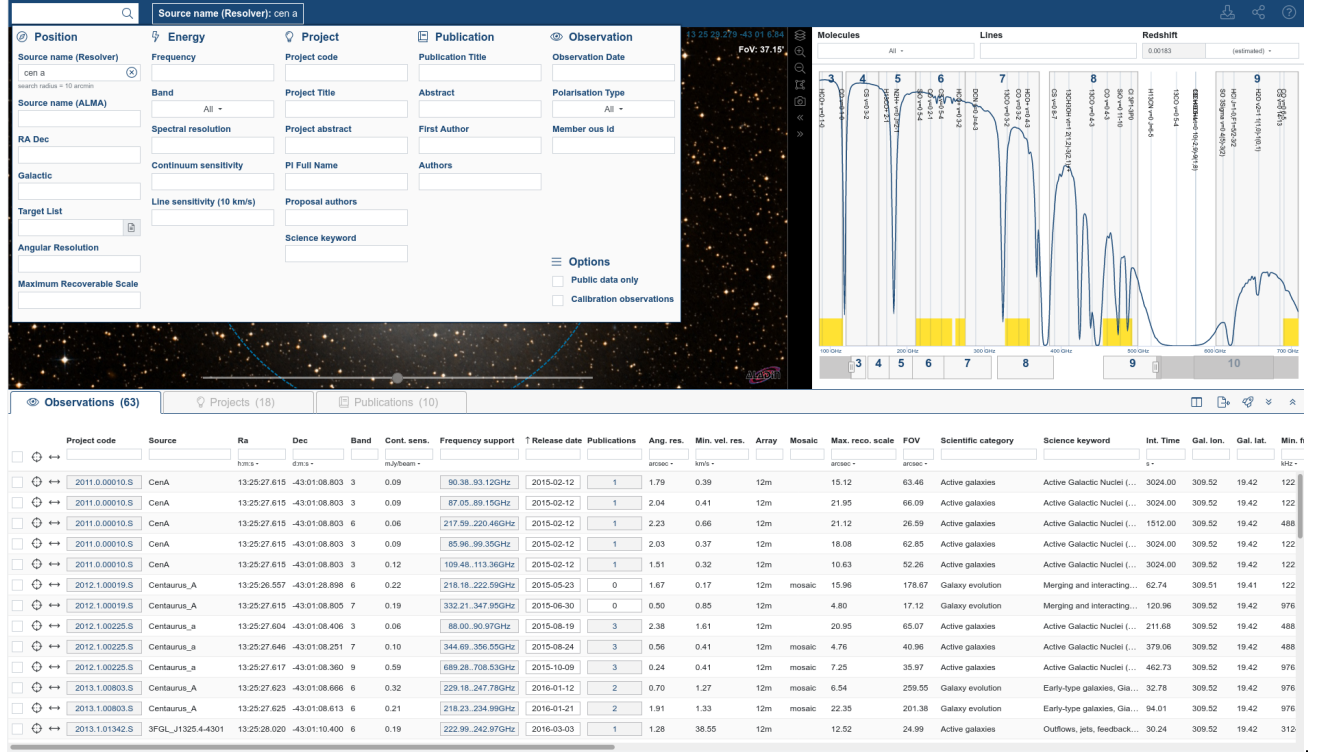


Figure 3.1: The archive query interface and an example tooltip, which shows the formats acceptable for entry and provides clickable examples.

Query constraints in addition to those mentioned above include:

- angular resolution (")
- maximum recoverable scale (")
- continuum sensitivity (mJy/beam)
- line sensitivity (10km/s) (mJy/beam)
- sky frequency (GHz)
- total bandwidth (GHz)
- spectral resolution (kHz)
- ALMA receiver band number
- observation date
- integration time (s)
- polarization type
- the titles and abstracts of the project as well as of ALMA publications

During the observing process, ALMA observes calibrator sources in addition to science observations requested by the PI. These are mostly planets and quasars. Uncheck the "Calibration observations" checkbox to include

those observations in the search. It is also possible to restrict the search to data that are already public and out of the proprietary period.

The query fields support a large number of operators. A full list of the capabilities can be accessed from the "?" link in the top right corner of the interface.

We also provide search on the scientific keywords that PIs specify when writing the proposals. These permit search for data per science topic. In addition we create scientific categories from those keywords to easier select entire scientific areas. Note that the categories we provide are different from the categories in the OT: the latter do change from cycle to cycle.

The maximum recoverable scale and the angular resolution computations use the 5%ile and 95%ile of the projected baseline lengths which is an improvement over previous calculations and also matches the way those quantities are computed in the OT and in AQUA.

The query constraints that have been placed are shown in the top row and can be easily removed, inverted or toggled off there.

3.3 Browsing the results

The result of the query is immediately presented in the result table as well as in the sky view and the frequency view. All three viewers are linked.

The sky view shows the spatial coverage of the observations and is based on AladinLite⁴. It will perform auto-zooming after a query to show the query results in an optimal way. If the sources are spread very much over the sky, instead of showing thousands of individual footprints a grid is displayed where the brightness of a grid-cell is proportional to the density of the observations in that cell.

If a coordinate or Sesame search is executed, the interface shows the currently used search radius as well as the centre location in blue. Zooming and panning with the mouse is supported. In addition, observations can be highlighted or selected by hovering or clicking on a footprint. The corresponding result-table row will scroll into sight.

The spectral coverage of the observations displayed in the result table is shown in the frequency view window. The yellow bars indicate the combined spectral coverage of the observations. A typical atmospheric transmission curve at the ALMA site helps guide the eye.

When proposing, PIs can enter the redshifts/velocities into the OT. We compute the median value of redshifts of the observations in the result table where they are available. This redshift is then used to display typical ALMA lines on top of the transmission curve. The estimated redshift can be overridden manually.

Zooming in, either with the mouse or by moving the bounds of the zoom window, will show more and more lines always following their relevance for ALMA which was determined by the number of times a line was part of an ALMA proposal. The list of lines to display can be modified by specifying directly a specific set of molecules in the Line filter box or by using one of the pre-selected categories from Splatalogue⁵.

The result table shows the values of the query. The results are available in three tabs where each row corresponds to one observation, one ALMA project or one Publication, respectively. In all three views, powerful subfiltering capabilities are available. The subfilters only act on the results of the given tab - as opposed to the query constraints that are reached at the top left of the interface. Like for the query constraints, subfilter constraints can be removed, inverted or toggled easily.

You can click on the headers of the columns to sort the table by the values of the particular column.

A red background of the checkbox in the left-most column indicates that the data are still in their proprietary period. In that case, the release date column provides an active link that you can use to add a reminder/event to your calendar (e.g. Google, Live, Yahoo!, iCal) to notify you when the data reaches the end of its proprietary period.

⁴<http://aladin.u-strasbg.fr/AladinLite>

⁵<https://www.cv.nrao.edu/php/splat/>

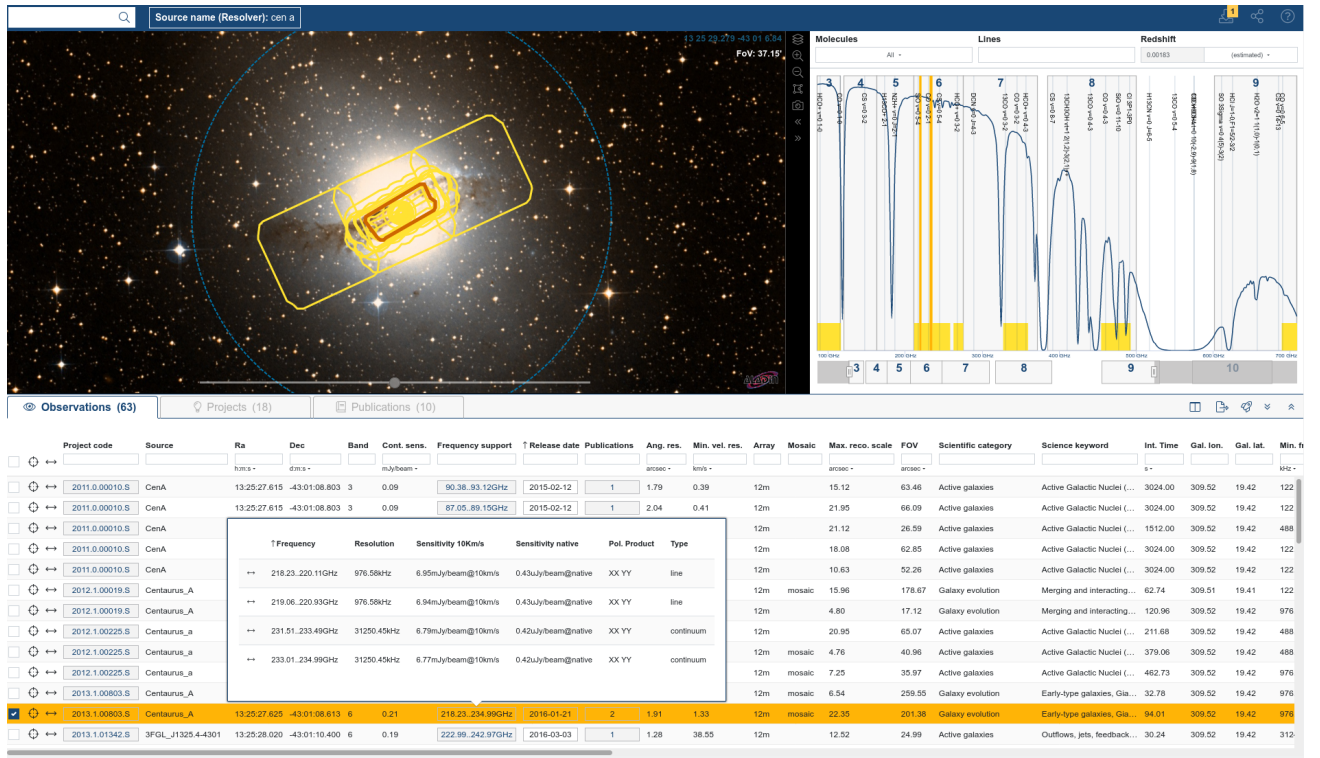


Figure 3.2: The results of an archive query. Each row entry has a footprint shown over a (selectable) sky background. Hovering over the observation row will highlight the footprint of that observation. The spectral coverage of the observations is shown in the right-hand panel.

Data without an available checkbox in the left column is either in the process of acquiring more data or in the process of being reduced and quality controlled. In rare cases, users may find and report problems with the data that has been delivered to them. Also in this case no checkbox for download is presented on the interface and the value in the Release date column is set to "In QA3" until the problem has been resolved.

The icons on the left-hand side, next to the checkboxes that can be used to select result rows for download, recenter the spatial and spectral viewers to the bounds of that particular row.

Several cells in the result table are active which is indicated with a slightly darker colour. These provide additional information. For example hovering over the project code link shows the full abstract of the proposal. It also displays the acknowledgement using that project code which then can be copied and pasted into a publication. Indeed, it is ALMA policy that this statement including the correct data-link must be placed into all publications making use of ALMA data (see also section 6).

One ALMA observation and thus one row in the Results table in "observation" display can have several Spectral Windows (spw) that were observed simultaneously. Although the query constraints on frequencies (frequency, bandwidth, spectral resolution) indeed do query against the values of the individual spectral windows in the ASA database, when displaying, only the observations for which a matching spw have been found are shown. The list of spectral windows is available by hovering over the "frequency support" cells.

All columns of the result table are visible and accessible by horizontal scrolling. The column order can be changed by dragging and dropping or by using the column-picker which is accessible at the top right of the result-table widget. Of particular interest might be the columns "Array" and "Mosaic" which give the ALMA array (12m, 7m, TP) and the pointing pattern, respectively.

The result table can also be exported to VOTable, TSV or CSV format or sent through the SAMP protocol to an application like Aladin or Topcat. Note that due to the security model of modern web-browsers this feature might not work everywhere.

In addition to being able to place constraints on the titles, authors or abstracts of publications resulting from ALMA data we offer the publication count constraint. Setting this to 0 and selecting "public data only" shows public but yet unpublished ALMA data.

The Results Bookmark/share link in the top right of the page provides you with a link that will instantly re-execute the search of the results that are currently being viewed. This is helpful when you want to send the results of a query to a colleague or just when you want to repeat the same query at a later period: As mentioned in section 3.1, the metadata of the observations are public as soon as those data have been observed. Therefore, at any given time the results do indeed match the current state of the ALMA observations.

3.4 Selecting the files for download

You can select data available for download with the checkbox in the left-most column of the Results Table. To select all available data for download, simply check the box at the top left of the table. Hitting the "Download request" button (top right corner) starts the ALMA Request Handler (RH), which will open in a separate tab or window (see Figure 3.4). If at least some of the selected data are still in their proprietary period and if you are not already authenticated, the RH will ask you to authenticate. Otherwise the Request Handler will immediately display the list of tar files that can be downloaded. A green check mark indicates the files for which you have access rights.

ALMA Request Handler

Login

Anonymous User: Request #2147094892194 ✓

Request Title: [Click to edit](#)

Download Selected

☒ readme
 ☒ product
 ☒ auxiliary
 ☐ raw
 ☐ raw (semipass)
 ☐ external

Project / OUSet / Executionblock	File	Size	Accessible
Request 2147094892194			
Project 2017.A.00021.T			
Science Goal OUS uid://A001/X12a3/X23a			
Group OUS uid://A001/X12a3/X23b			
Member OUS uid://A001/X12a3/X23c			
SB GRB17120_a_03_TM1			
<input checked="" type="checkbox"/> readme	member.uid_A001_X12a3_X23c.README.txt	130B	✓
<input checked="" type="checkbox"/> product	2017.A.00021.T.uid_A001_X12a3_X23c_001_of_001.tar	176.4MB	✓
<input checked="" type="checkbox"/> auxiliary	2017.A.00021.T.uid_A001_X12a3_X23c_auxiliary.tar	186.9MB	✓
<input type="checkbox"/> raw	2017.A.00021.T.uid_A002_Xc8ed16_Xbde.asdm.sdm	17.2GB	✓
<input type="checkbox"/> raw	2017.A.00021.T.uid_A002_Xc8ed16_Xc5b0.asdm.sdm	13.4GB	✓
<input type="checkbox"/> raw	2017.A.00021.T.uid_A002_Xc8ed16_Xcb17.asdm.sdm	14.7GB	✓
		Total: 45.7GB	

Figure 3.3: The results of a download request. This query selects only one project. Semipass data are not selected in the top boxes, so data having this status are not selected for download. You may opt only for raw data or only for products.

If authenticated, you will receive an email notification with the content of your request. This email also contains a link to the request in the RH. You might also decide to authenticate despite the fact that the data you would like to download are all out of the proprietary period in case you want to come back to the request. Indeed, a list of all (authenticated) requests can be obtained by clicking on the "My requests" link in the top right corner of the RH for authenticated users.

At this stage, you have the option of downloading the available data products as well as the associated raw data via tar files separately via the selection boxes shown in Figure 3.4.

For data from Cycle 5 onwards (i.e. 2017.*, 2018.*, ...), the data products themselves (FITS files) and the auxiliary products (log-files, scripts, quality reports) are available in separate tar files. Users who want to recreate the calibrated Measurement Sets (MS) (see also chapter 5) from the raw data therefore do not have to download the FITS products anymore but can just download the auxiliary tar file as well as the raw data. On the other hand, users who are only interested in the FITS products do not have to download the files provided in the auxiliary tar package anymore.

ALMA Request Handler

Anonymous User: Request #2147094892194 ✓
Request Title: [Click to edit](#)

Download Selected

✓ readme

✓ product

✓ auxiliary

raw

raw (semipass)

external

Project / OUSet / Executionblock	File	Size	Accessible
Request 2147094892194			
Project 2017.A.00021.T			
Science Goal OUS uid://A001/X12a3/X23a			
Group OUS uid://A001/X12a3/X23b			
Member OUS uid://A001/X12a3/X23c			
SB GRB17120_a_03_TM1			
readme	member.uid_A001_X12a3_X23c.README.txt	130B	✓
product	2017.A.00021.T.uid_A001_X12a3_X23c.001_of_001.tar	176.4MB	✓
auxiliary	2017.A.00021.T.uid_A001_X12a3_X23c.auxiliary.tar	186.9MB	✓
calibration	uid_A002_Xc8ed16_Xbded.calibration.plots.tgz	38.0MB	✓
calibration	uid_A002_Xc8ed16_Xbded.calibration.tgz	18.8MB	✓
calibration	uid_A002_Xc8ed16_Xc5b0.calibration.plots.tgz	35.0MB	✓
calibration	uid_A002_Xc8ed16_Xc5b0.calibration.tgz	13.6MB	✓
calibration	uid_A002_Xc8ed16_Xcb17.calibration.plots.tgz	38.6MB	✓
calibration	uid_A002_Xc8ed16_Xcb17.calibration.tgz	16.4MB	✓
log	member.uid_A001_X12a3_X23c.log.tgz	30.0KB	✓
log	member.uid_A001_X12a3_X23c.uid_A002_Xc8ed16_Xbded.log.tgz	78.0KB	✓
log	member.uid_A001_X12a3_X23c.uid_A002_Xc8ed16_Xc5b0.log.tgz	74.0KB	✓
log	member.uid_A001_X12a3_X23c.uid_A002_Xc8ed16_Xcb17.log.tgz	81.0KB	✓
qa	member.uid_A001_X12a3_X23c.concat.ms.Dterm_sky_imag_spw0-1.png	243.0KB	✓
qa	member.uid_A001_X12a3_X23c.concat.ms.Dterm_sky_imag_spw2-3.png	262.0KB	✓
qa	member.uid_A001_X12a3_X23c.concat.ms.Dterm_sky_real_spw0-1.png	196.0KB	✓

Figure 3.4: As above but with the expanded file-contents of the auxiliary tar file.

In addition, for data from Cycle5 and onwards, you can also select individual data-product files for download. In this case a small grey arrow is displayed in front of the 'products' and 'auxiliary' product tar files which allows you to expand the tar file content and select the files needed. In addition, each file is a link and thus the file can also be saved with a right-click and "Save As ..".

ALMA data do have quite a deep data hierarchy⁶. In short: observations that can be observed and calibrated together and independently are placed in a Member OUS (Observation Unit Set). Observations that have to be combined from several Member OUSs are placed into the same Group OUS. This is done until all observations of a Science Goal (corresponding to one Science Goal OUS) in the OT are distributed.

This hierarchy means that when you select specific targets in the Results Table, the data that will be displayed in the resulting Request hierarchy will include the data for all targets that were observed in the same parent Member OUS, as well as all targets observed in associated Member OUSs (ones with the same parent Group OUS). You can display the list of all science targets observed as part of each MOUS by expanding the SB tree which appears directly below the MOUS file icon. So for instance, if you select the 12-m data for a specific protoplanetary disk in the Results Table, and that object was observed as part of a survey of such disks that included multiple 12-m configurations, as well as 7-m and TP observations, the Request results will include the 4 MOUS for each configuration component, and the products listing will include all protoplanetary disks observed in each of those MOUS.

Currently two types of files are available for download:

- files containing the data products (FITS files, quality assessment results, scripts, log files, calibration and flagging tables).
- files containing the ALMA raw data (ALMA Science Data Model, ASDM)

The type of the file is indicated in the data hierarchy column. In case you only want to have a look at the products that have been created for the quality assessment, or if you only want to read the README file, only the product file(s) have to be downloaded.

Data marked with "semipass" are data which may be useful, but are known to have issues which lead them to be excluded from the creation of the data-products (e.g., missing calibrators, too little science data). The name of the Scheduling Block that was used to acquire the data is shown to make it easier to understand which data

⁶see e.g. Section 8.1. of <http://almascience.org/documents-and-tools/cycle5/alma-technical-handbook>

are contained in the products. We also offer the possibility to expand the entire list of source names contained in the products.

With the checkboxes on the top the different categories of products can be selected quickly.

As mentioned before, there are three individual archives at the ARCs and thus three individual RHs available. You can download data from any of the three ARCs. Note, however, that data requests are not transferrable. If you want to download data from a different ARC you need to create a new request.

3.5 Downloading the data

ALMA data can be potentially very large. Although a single file can also be downloaded by right-clicking on the link of the file and selecting "Save As .." from the browser menu, to assure efficient downloading and to make good use of the available bandwidth for ALMA downloads, you should download ALMA data in several parallel download streams.

After selecting the desired products/data for download and clicking the Download Selected button in the top-left portion of the window, you are prompted with four different download methods:

- Download Script: A bash download script is created that can be directly run from the command line on Linux and Mac.
- Download Manager: ALMA's download manager is launched as a browser applet, allowing for parallel downloading that can be paused and resumed.
- Web Start Download Manager: ALMA's download manager is launched as a desktop application via Java Web Start. Accordingly, downloads will not stop if you close your browser.
- File List: A new page is opened containing a list of URLs for downloading the data that may be useful for third-party download managers such as *DownThemAll*.

The recommended method for downloading data is using the Download Script. This allows one to easily download to computers that do not have web browsers available (e.g. processing environments). Save the script to your Linux or Mac, add execute rights, run it. After a successful download, the script will allow you to unpack the tar files directly.

```
chmod a+x downloadRequest12345.sh
./downloadRequest12345.sh
```

If you prefer to unpack the files later or on a different computer, then execute the command:

```
for i in *.tar; do tar -xvf $i; done
```

Note, when downloading via the file list, please do not attempt to chop the download into pieces to make it faster! This places a significant load on the servers and may result in your downloads being throttled. For example, in the Firefox plugin DownThemAll, make sure that the property "Max. Number of Segments Per Downloads" is set to 1. Likewise, it is easy to setup a download with more than 5 files in parallel, the archive limit. Please refrain from doing this.

3.6 Data packaging

3.6.1 Directory structure

The data are streamed to the user as tar files which allows for the delivery of the ALMA data together with the complete directory structure following the ALMA data hierarchy.

The ALMA Archive currently has tarfiles for either products or "raw" data files (where raw is defined here as

post-correlation, but pre-calibration) in ASDM format⁷. A typical package will unpack to a directory structure with the first three levels like:

```
2012.1.01234.S/science_goal.uid__A001_X12345_X123/group.uid__A002_X6789ab_X6789
/member.uid__A002_Xcdef1_X234/
```

(Note that Cycle 0 data do not have the UIDs for the OUSes appended to the directory names. Instead the delivery date is appended to the member OUS directory name). In the member... directory there are the following subdirectories:

```
calibrated calibration log product qa raw script
```

and a README file.

To unpackage the tarfiles rapidly you can set up a loop in bash:

```
for i in *.tar; do tar -xvf $i; done
```

or source the following C shell script if you are using tcsh or csh:

```
#!/bin/csh
foreach i (*.tar)
    tar -xvf $i
end
```

The contents of the directories are as follows:

- The **calibration** directory contains calibration and flag tables.
- The **log** directory contains the CASA logs from running the reduction scripts/pipeline.
- The **products** directory contains the FITS images.
- The **qa** directory contains the QA2 reports (manual calibration) or the pipeline weblog (pipeline calibration).
- The **script** directory contains the reduction scripts.
- The contents of the **raw** and **calibrated** directories depend on whether the data were from Cycle 0 or from later Cycles (see below).

More details about the packaged QA2 products can be found in the data product guide⁸.

3.6.2 Cycle 0

Cycle 0 was unique in that calibrated MS were archived together with the data products. The data for each Member OUS is combined into a set of tar files, the contents of which are linked from the Science Portal⁹. In particular, it is possible from these contents to see which tar files contain the image FITS files for a given delivery.

Two MS were stored. The first is a MS with the system temperature (T_{sys}) and water vapour radiometer (WVR) corrections applied. This is stored in the "raw" directory. The second is a fully calibrated MS, stored in the "calibrated" directory. Note that the raw ASDMs are not in the package, but are available from the ARCs upon request.

⁷These raw files are colloquially referred to as ASDMs, but ASDM strictly refers to the data model, not the file. The correct designation is Execution Block (EB).

⁸e.g., <http://almascience.org/documents-and-tools/cycle3/ALMAQA2Products3.0.pdf>

⁹<http://almascience.org/alma-data/archive/delivery-list>

3.6.3 Cycle 1 and later

In Cycle 1 and later, calibrated MS are no longer included in the package. If only the package tar file is downloaded (the "product" file in the request handler listing), then the raw directory is not present. If tar files of the raw data are downloaded, they will populate a raw directory when they are untarred. For the creation of the calibrated MS see Chapter 5.

Chapter 4

Programmatic access

4.1 astroquery

The AQ has been developed so that it can be accessed programatically, i.e. from programs or scripts. Adam Ginsburg has developed the external python tool

`astroquery`

. Astroquery is a python package which allows querying and retrieving data from many large observatories including ALMA. Authentication support for the download of proprietary data is provided as well as the extraction of FITS files only for Cycle 0 data (see also section 3.6.2).

As a simple example, querying the ALMA Archive for the source M83 and retrieving all the corresponding data is done by

```
from astroquery.alma import Alma
import numpy

# Querying the ALMA Science Archive for source M83
result = Alma.query_object('M83')

# Extracting a list of Member ObsUnitSets
member_ous = numpy.unique(result['Member ous id'])

# Creating an ALMA astroquery instance and downloading all data the identified Member OUS
myAlma = Alma()
myAlma.cache_location = '/big/external/drive/'
myAlma.retrieve_data_from_uid(member_ous[0])
```

At this stage, astroquery is the preferred way to query the ASA for a long list of sources. A full set of examples on the usage of the package for ALMA data is available on the astroquery ALMA page (<https://astroquery.readthedocs.org/en/latest/alma/alma.html>).

4.2 VO

ALMA is strongly committed to providing VO services for all ALMA data. We are offering Table Access Protocol (TAP) as well as Simple Imaging Protocol (SIAPv2) services for ALMA data at <https://almascience.org/tap> and <https://almascience.org/sia2>, respectively.

External providers offer VO services the archived FITS products¹. Note that those products cover only a small

¹<http://jvo.nao.ac.jp/portal/alma/archive.do>, <http://www4.chivo.cl>

fraction of the total number of sources and available spectral range. As when viewing the products in the archive, these FITS products are generated for quality assessment purposes and in most cases do not represent the best, or even science-ready, products.

Chapter 5

Scientific exploitation of the data

While in a few cases the FITS data-products that are provided by ALMA can be sufficient to do science directly, more often than not users should create their own images from the calibrated Measurement Set data (MS). As mentioned above, for Cycle 0 the calibrated MS are directly delivered with the data packages. For data later than Cycle 1, MS can be recreated from the raw data by running the command

```
casa -c *scriptForPI.py
```

within the latest version of CASA. For older data (pre Cycle 5) or data that were manually reduced, use the version of CASA that was used to create the data (see the README file or the QA2 report in the qa directory). The *scriptForPI.py* file can be found in the *script* directory. The calibrated MS will be created in the "calibrated" directory.

Please see the Processing section¹ of the ALMA Science Portal for details. From there, you can follow the links to understand the contents of the data products; download and install CASA; download and install the ALMA pipeline; and the hardware requirements for CASA and the pipeline. CASA also provides guides related to pipeline imaging reprocessing².

Moreover the ALMA Regional Centres also provide calibrated MS for PIs and archival researchers. The calibrated MS can be requested through the ALMA Helpdesk. In addition, for Cycle 5 and later, users can generate calibrated MS themselves through the service offered at <https://archive-new.nrao.edu>.

All users of the archive are welcome to request assistance with calibration and analysis through the ALMA Helpdesk, and face-to-face visits to the ARCs are advisable when additional assistance is needed.

¹<http://almascience.org/processing>

²https://casaguides.nrao.edu/index.php/ALMA_Imaging_Pipeline_Reprocessing

Chapter 6

Acknowledgement statement

All publications making use of ALMA data must include the following unaltered statement in the acknowledgement section:

"This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.01234.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ."

Please substitute the place-holder project code 2011.0.01234.S with the project code of the data used. Should data from several projects have been used, please use the following notation: ADS/JAO.ALMA#2011.0.01234.S, ADS/JAO.ALMA#2011.0.01235.S, ADS/JAO.ALMA#2011.0.01236.S.

In addition, publications from NA authors must include the standard NRAO acknowledgement:

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

Readers of the publication can enter either the full data tag or the project code into the "Project code" field of the ASA Query Interface (3.2) in order to obtain a list of the corresponding observations. We also use the data tag to link the publications to the data in the archive which will allow the user to query for publications using ALMA data in the ASA and also to show all publications related to the displayed data-rows.

If referring to sources detected for first time in ALMA fields you should follow one of the options of the naming convention

ALMA JHHMMSS.s+DDMMSS
ALMA JHHMMSS.ss+DDMMSS.s
ALMA JHHMMSS.sss+DDMMSS.ss
ALMA JHHMMSS.ssss+DDMMSS.sss

(approved by IAU "Clearing House" of Commission 5 Working Group on Designations), where J indicates J2000 coordinates.

The coordinates should be truncated according to the precision in the position of the source. Typically this should be approximately 1/10th of the size of the synthesized beam used in the discovery observation, e.g. for a 1" beam, declination should be given to a precision of 0.1" of arc and RA to 0.01s of time (ALMA JHHMMSS.ss+/-DDMMSS.s), for a 0.1" beam declination to 0.01" of arc and RA to 0.001s of time (ALMA JHHMMSS.sss+/-DDMMSS.ss) etc.

Chapter 7

Frequently asked questions

Please refer to the FAQ/Knowledgebase in the ALMA Science Portal for the answers to frequently asked questions: <https://help.almascience.org/index.php?Knowledgebase/List/Index/4/archive--data-retrieval>

Please do not hesitate to contact us through the ALMA Helpdesk in case you have questions, comments or suggestions.



The Atacama Large Millimeter/submillimeter Array (ALMA), an international astronomy facility, is a partnership of the European Organization for Astronomical Research in the Southern Hemisphere (ESO), the U.S. National Science Foundation (NSF) and the National Institutes of Natural Sciences (NINS) of Japan in cooperation with the Republic of Chile. ALMA is funded by ESO on behalf of its Member States, by NSF in cooperation with the National Research Council of Canada (NRC) and the National Science Council of Taiwan (NSC) and by NINS in cooperation with the Academia Sinica (AS) in Taiwan and the Korea Astronomy and Space Science Institute (KASI).

ALMA construction and operations are led by ESO on behalf of its Member States; by the National Radio Astronomy Observatory (NRAO), managed by Associated Universities, Inc. (AUI), on behalf of North America; and by the National Astronomical Observatory of Japan (NAOJ) on behalf of East Asia. The Joint ALMA Observatory (JAO) provides the unified leadership and management of the construction, commissioning and operation of ALMA.

