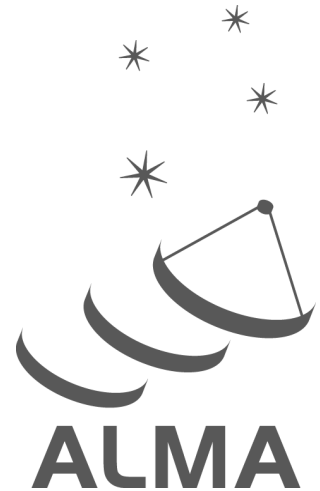


ALMA Science Archive Manual



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
1.0.x	March 2016	Felix Stoehr, Eric Murphy, Mark Lacy, Stéphane Leon, Erik Muller, Kouchiro Nakanishi

Contributors

In publications, please refer to this document as:

F. Stoehr et al. 2016, ALMA Science Archive Manual, ALMA Doc. 4.15 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 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	6
3.4	Selecting the files for download	7
3.5	Downloading the data	8
3.6	Data packaging	8
3.6.1	Directory structure	8
3.6.2	Cycle 0	9
3.6.3	Cycle 1 and later	10
4	Programmatic access	11
4.1	astroquery	11
4.2	VO	11
5	Acknowledgement statement	13
6	Frequently asked questions	15

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 at <http://almascience.org/aq> and hit "Search"
2. Use the checkboxes to select observations of interest and hit "Submit download request"
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 user.

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 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 <http://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 left-hand menu of the Science Portal. Here you will find links to the three types of ALMA data:

- Archive: ALMA data taken for PI programs
- Calibrator Catalogue: Information on calibration sources, including flux histories
- Science Verification: ALMA data taken for Science Verification

In addition, there is a link to a "Data Processing" page which 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 takes you to a page with a button that brings up the ALMA Query interface (AQ). On this page there is also a link to a page listing the contents of the Cycle 0 deliveries¹. Useful documents available from the Documents and Tools link include the Science Pipeline Manual and QuickStart guide, CASA documentation and descriptions of the ALMA quality assurance for data products ("QA2") and the delivered data products under "ALMA QA2 Data Products" (see also Section 3.6).

2.2 Project 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 delegation is to be applied, 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 granted to entire projects rather than individual deliveries from a project.

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

Chapter 3

Science Archive

3.1 Content and policies

The ALMA Science Archive (ASA) contains all public and proprietary data of PI-proposal observations. The metadata shown and offered for querying are those of the raw observations. For this reason, currently, one source of one project can lead to several rows in the ASA results table.

In full operations ALMA is expected to observe about 200TB worth of data every year. These data are stored at the Joint ALMA Observatory in Santiago, Chile and are copied from there to the three ARCs. Each ARC indeed holds a complete copy of the ASA. (See chapter 14 of the Technical Handbook¹ for details.)

ALMA data are protected by a proprietary period of 12 months, after which they become public. All public data can be accessed anonymously. Observational metadata 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 are public immediately. 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

Sources can be queried either by their name or by a search near a position on the sky specified in Right Ascension and Declination. 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 1°). Sesame (via NED, Simbad and Vizier) will parse names commonly found throughout the literature and return search coordinates. A green check mark indicates a successful search, otherwise, a red cross is returned. Sources can also be queried by their ALMA name, which is the name of the source as registered by the PI in the Observing Tool (OT). For this case, partial matches through wildcards (?, *) and boolean OR expressions ("|") can be used.

Various additional constraints can be placed on a query to the ALMA Science Archive. They have been grouped into Position, Energy, Time and Polarisation categories for those that are directly related to the properties of the photons and into Observation, Project and Options for constraints on the weather, the proposal and the display options. Autocompletion is enabled in all text fields except for the Source Name (Resolver) field, allowing for easier search. Note that despite the suggestions from the autocompletion, the full set of operations (including

¹<http://almascience.org/documents-and-tools/cycle4/alma-technical-handbook>

²<https://almascience.eso.org/documents-and-tools/cycle4/alma-user-policies>

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

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.

Query constraints in addition to those mentioned above include:

- angular resolution (")
- largest angular scale sensitivity (")
- precipitable water vapor (mm)
- sky frequency (GHz)
- total bandwidth (GHz)
- spectral resolution (kHz)
- ALMA receiver band number
- observation date
- integration time (s)
- Polarization type

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

By default the search results are displayed so that one row of the results table corresponds to one observation in the raw data. However, if you are only interested in a list of projects that match the search criteria, you can select the "project" option. Furthermore, a search can be constrained to only return public data, i.e. data that you have access to without being a PI or having been delegated access rights to data.

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 "Science Observations only" checkbox to include those observations in the search.

If they are known before the query, the additional constraints can be entered directly into the query form as described. However, for quicker exploration, additional constraints (filtering) can be applied also directly on the results table.

3.3 Browsing the results

After executing a search, you can browse the results in the Results Table tab. By default, six columns are shown that include the Project code, Source name, RA, Dec, Band, Integration time, Release date, Velocity resolution, and Frequency support.

Clicking into the headers of the columns will sort the Results table according to the values in that column. Reordering of the columns can be done by dragging and dropping the column headers horizontally.

By clicking the "more columns" link in the top right portion of the table, a much larger number of column fields can be added to the table. Three buttons on the top allow to quickly show all rows, to reset the default column order, and to order the columns alphabetically. The browser's text search function (CTRL+F or Command+F) can be used, too.

Columns above the red bar in the column picker window are displayed in the Results Table. They can be dragged and dropped below or above the red bar or can be reordered. Also, the red bar itself can be dragged and dropped. Finally, a checkbox in front of the column names makes it easy to quickly select or deselect an individual column for display.

As mentioned, using the empty boxes in the top row of each column, each of these fields can be used to additionally filter the data that is being shown.

The Results Bookmark 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.

You can also use the Export Table function to download the Results Table in VOTable (XML), CSV, or TSV formats. In this case, the complete table is always returned.

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.

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.

One ALMA observation and thus one row in the Results table in "raw" 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. If you hover with the mouse over the Frequency Support column, a summary of all the spw that have been observed for the given observation is displayed.

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 an 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 5).

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 "Submit download request" button starts the ALMA Request Handler (RH), which will open in a separate tab or window. 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.

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 individual tar files separately.

ALMA data do have a quite deep data hierarchy⁴. In short: observations that can be observed and calibrated together and independently are placed in a Member OUS. 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 if you query and select data from a Member OUS, e.g. containing observations done with the 12 m array, you might also be interested in observations from the same Group OUS containing the same source but which potentially have been observed with the 7 m and/or TP array. Therefore, the RH expands all selected data to show the complete Group OUSs.

⁴ see e.g. Section 8.1. of <https://almascience.eso.org/documents-and-tools/cycle4/alma-technical-handbook>

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.

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, and then unpack the obtained tar files:

```
chmod a+x downloadRequest12345.sh
./downloadRequest12345.sh
for i in `ls *.tar`; do tar -xvf $i; done
```

Note, when downloading via the file list, please do not attempt to chop the download of a single 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 us to deliver the ALMA data together with the complete directory structure following the ALMA data hierarchy to you.

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 format⁵. 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 does 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 `ls *.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 guide⁶.

3.6.2 Cycle 0

Cycle 0 was unique in that calibrated measurement sets 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 <https://almascience.nrao.edu/alma-data/archive/delivery-list>. In particular, it is possible from these contents to see which tar files contain the image FITS files for a given delivery.

Two measurement sets were stored. The first is a measurement set (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.

⁵These 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).

⁶<https://almascience.eso.org/documents-and-tools/cycle3/ALMAQA2Products3.0.pdf>

3.6.3 Cycle 1 and later

In Cycle 1 and later, calibrated measurement sets 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. Measurement sets can be recreated from the raw data quickly by running

```
execfile('scriptForPI.py')
```

which is in the

```
script
```

directory from within CASA.

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 tool astroquery <https://astroquery.readthedocs.org/en/latest/>. 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. The necessary Table Access Protocol (TAP) infrastructure¹ is in place and already now serves as backend for the Archive Query web interface. Originally planned to serve only ALMA FITS products from the ALMA Pipeline, the strategy had to be revised recently to allow for a search through the complete holdings based on metadata. The preparative steps for this project are underway and VO services ObsTAP and SIAPv2 are currently planned for 2017. A VO service for FITS products is expected to follow.

¹using code from <https://github.com/openadc/tap>

External providers offer VO services for the FITS products that resulted from the QA2 process². Note that those products cover only a small fraction of the total number of sources and available spectral range.

²<http://jvo.nao.ac.jp/portal/alma/archive.do>, <http://www.chivo.cl>

Chapter 5

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), NSC 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 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 6

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.

