ALMA Science Archive Manual
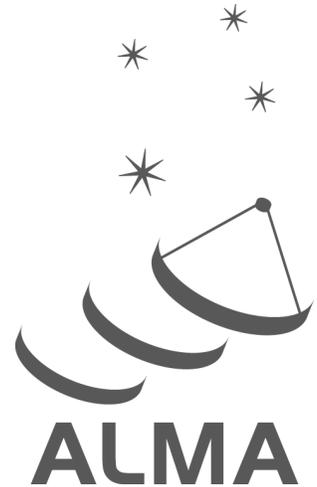
ALMA

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
4.1.0.x	March 2016	Felix Stoehr, Eric Murphy, Mark Lacy, Stéphane Leon, Erik Muller, Kouchiro Nakanishi
5.1.0.x	March 2017	Brenda Matthews, Felix Stoehr, Mark Lacy, Kouchiro Nakanishi, Stéphane Leon
6.1.0.x	March 2018	Brenda Matthews, Felix Stoehr, John Hibbard, Kouchiro 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
8.2.0.x	March 2021	Felix Stoehr, Adele Plunkett, Andy Lipnicky, Brenda Matthews, Kuo-Song Wang, Patricio Sanhueza
9.1.0.x	March 2022	Felix Stoehr, Adele Plunkett, Andy Lipnicky, 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

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/aq

2. Use the checkboxes to select observations of interest and hit "Explore and 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 3.6, respectively.

1.2 User support

In each of the three partner regions an ALMA Regional Centre (ARC) has been established[1]. 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 Helpdesk[2] for any question or comment. In particular we strongly enourage 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 have computing facilities available.

---

[1] https://almascience.org/about-alma/alma-organization
[2] https://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 five 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

- ARI-L: Additional Representative Imaging for Legacy

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 takes you directly to the Science Archive Query Interface. The sub-menu item "Archive documentation" includes links to the latest documents describing ALMA data products (including Cycle 0) and a table of recent updates, tips and tricks, and known issues.

Useful documents available from the Documentation menu of the ALMA Science Portal include the User's Guide for ALMA Pipeline, descriptions of the ALMA quality assurance for data products ("QA2") and CASA documentation (see also Section 3.7).

We also provide a link to Video Tutorials of the ASA explaining the functionality of the web interface and tools. Furthermore, we provide an extensive set of Jupyter Notebook tutorials. These tutorials explain how to query and download data from the ASA programmatically e.g. with Python[1].

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, the "Project delegation" tab is accessed using the "Next" button in the upper right. Once in the Project delegation tab, select the "Add delegee" button on the bottom left and select which project you wish to add a delegated user to. Finally, at the bottom select the

---

[1] https://almascience.org/alma-data/archive/archive-notebooks

access level you wish the user to have, type their name into the other field to find them, and click "Save". 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. 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 currently producing 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 Handbook[1] 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 Users' policies[2] for details.

3.2 Querying the ALMA Science Archive

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

The main input form appears when you hover over the search field in the top left corner. Sources can be queried either by name or by a position on the sky specified in Right Ascension and Declination or Galactic coordinates. The name resolver uses Sesame[3] to perform a semi-case-insensitive search for a source name. A search radius in arcmin 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 3.1 shows, they have been grouped into the following five categories: Position, Energy, Project, Publication, and Observation.

Autocompletion is enabled in all text fields except for the Source Name (Resolver) field, allowing for easier search.

---

[1] https://almascience.org/documents-and-tools/latest/alma-technical-handbook
[2] https://almascience.org/documents-and-tools/latest/alma-user-policies
[3] http://cds.u-strasbg.fr/cgi-bin/Sesame

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.
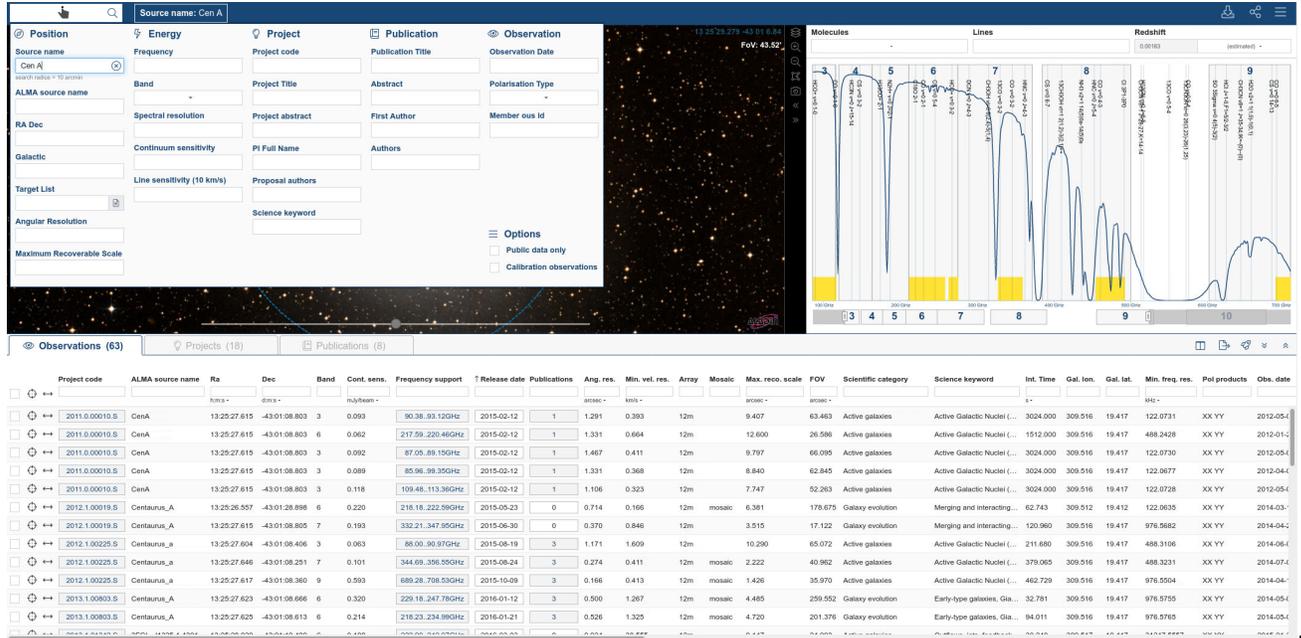

Figure 3.1: The archive query interface.

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)
- spectral resolution (kHz)
- ALMA receiver band number
- observation date
- polarization type
- the title and abstract 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 planets and quasars. Check 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 Help menu item of the three-bar-icon 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 80%ile of the projected baseline lengths which is an improvement over previous calculations and also matches the way those quantities are computed in the ALMA Observing Tool and during ALMA's quality assurance.

The ASA also features the search for object types. We query SIMBAD and NED and extract all sources from those catalogues that fall into any ALMA footprint. Those data are then stored in a database and can thus be offered fro querying. Per default we show all ALMA observations where one or more of the selected objects fall into the footprint. In addition, we also compute the best-matching source of the NED and SIMBAD sources with the observation as requested by the PI. For each source falling into the footprint we compute a score value based on the distance to the coordinates given by the PI, the closeness of the source name to the source name given by the PI using a string-distance measure as well as the strength of the emission of the source in case it is provided in SIMBAD or NED. The source with the largest score is then tagged 'best match' and it is possible on the interface to restrict the search and the display to those sources. Chances are high that the PIs then really have targeted those best-match sources.

WARNING: Use the object-type search with great care! The object type identification is by construction incomplete. There will be false-positives and false-negatives. Identifications may change with time as is true for all scientific measurements. Also several categories may apply to a single object but only the main type is used here.

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

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 AladinLite[4]. It will perform auto-zooming after a query to show the query results in an optimal way. If the sources are spread widely 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/velocites 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 targeted by an ALMA proposal. The list of lines to display can be modified by specifying directly a specific set of molecules in the "Lines" filter box or by using one of the pre-selected categories from Splatalogue[5].

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 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.

---

[4] https://aladin.u-strasbg.fr/AladinLite
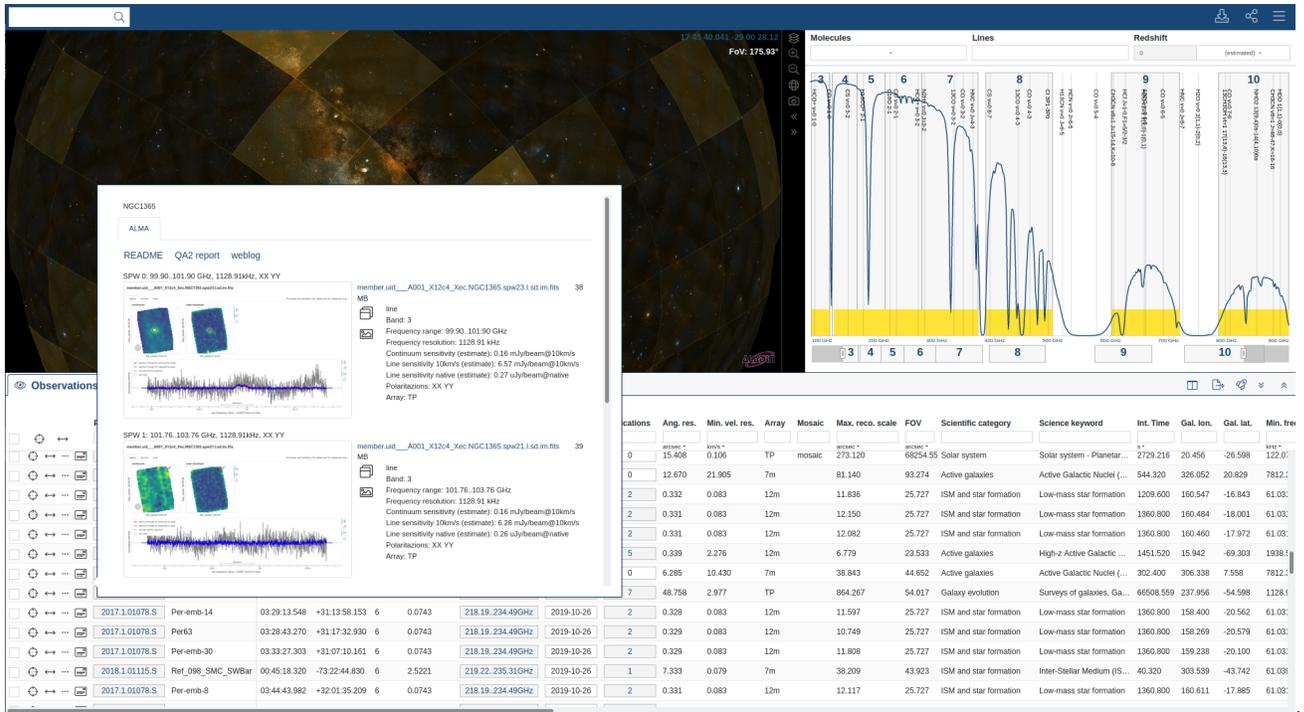[5] https://www.cv.nrao.edu/php/splat/

Figure 3.2: The results of an archive query here showing a preview tooltip. 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.

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 proprietary period ends.

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. In these cases, no checkbox for download is presented on the interface either, and the value in the "Release date" column is set to "In QA3" until the problem has been resolved.

The three icons on the left-hand side, next to the checkboxes for data selection, recenter the spatial and spectral viewers to the bounds of that particular row, respectively.

Hovering over the ellipse/tilde icon, a pop-up will open showing the the most similar proposals or project to the one under consideration. We use a state-of-the-art text-similarity method to determine the list of the most similar entities. That list does not only contain links to the proposals or projects; a link is also provided to open the list of similar projects or proposals in a new browser tab inside a new AQ. It is therefore possible to add more constraints or subfilters to the lists of similar items.

The icon next to the text-based-similarity search opens up a pop-up with previews. Previews, i.e. quick-look images allowing users to grasp - with a single glance - the content of the data files, are a highly-desired ingredient of modern astronomical archives. A particular challenge for ALMA is to provide previews that give the maximally-useful essence of large 3D data-cubes with up to 3800 spectral channels. After substantial experimenting, we opted for previews as shown in Fig 3.3. The previews are static and fully self-contained interactive html files. You can zoom and pan the images and spectra for which we make use of the bokeh.org [6] library. Data products can be separated into two parts, the continuum emission (the flux for a pixel that is constant through the entire data-cube) and the line emission (the remainder after the continuum was subtracted). Just collapsing, i.e. averaging the line emission, in order to create a two-dimensional image erases the structure of the emission that sometimes can be concentrated into only a few spectral channels. We find that the peak flux image (for
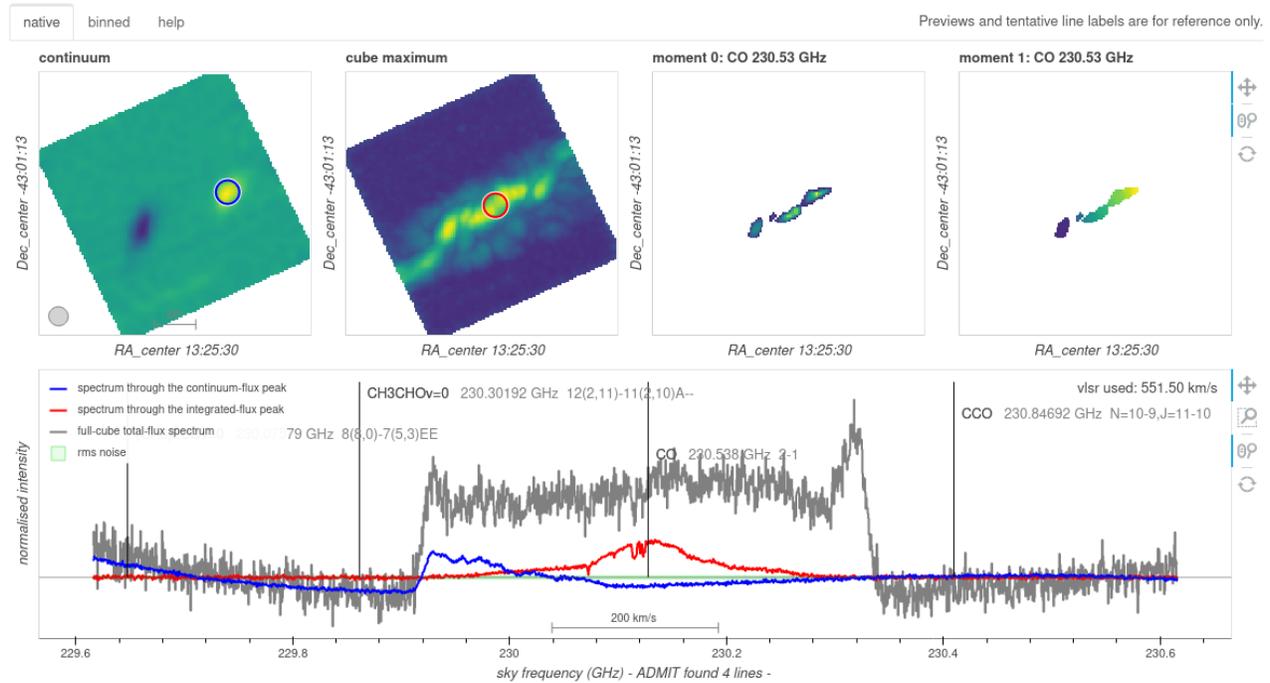
---

[6] https://bokeh.org

Figure 3.3: Previews are available for a large fraction of ALMA's FITS files. They show the continuum image, the peak-flux image (moment 8), the moment-0 and moment-1 of the strongest line in case ADMIT has detected any lines as well as several spectra through the cube together with the lines and tentative line labels.

each pixel the largest value along the spectral axis, sometimes also called "moment 8") gives the most useful representation of the line emission.

In addition to showing those images, we run the ALMA Data Mining Toolkit (ADMIT)[7] over the data-cubes. That software, originated from an ALMA development project, tries to find spectral lines in the data-cube and - if provided with a guess for the relative velocity or redshift of the observed source - to identify them and label them with the molecule and transition tentatively responsible for the line emission. We use the following algorithm to provide the velocity to the ADMIT software. If a velocity was given by the Principal Investigator (PI), in the ALMA Observing Tool (OT), that velocity is used. If no velocity was given, we look for other existing ALMA observations of the same source and use the velocity if a value is available. If not, we try to match the observed source with a source from SIMBAD and NED and use the velocity from those services. If none is available, we check to see if the PI has provided lines they were looking for in the OT and compute the velocities from that information. This process provides guesses for velocities for about 80% of all FITS files. Four spectra across the cube are shown in the spectral panel of the preview. In case strong lines have been detected, the moment-0 and moment-1 image of the strongest line are displayed. The same plots do also exist in a binned version on the preview which makes the detection of weak broad lines easier. Finally, a help tab provides detailed descriptions of the content.

WARNING: The previews are only for reference and the line-labels are only tentative! Under **no** circumstance the information in the previews can be used directly for science. ADMIT requires the input of the velocity of the source, but while PIs do put sometimes velocity values into the OT, there are many reasons why those values can be wrong. (e.g. a field of galaxies where the velocity is chosen so that all galaxies fall into the SPW, no velocity known at all, a lensed galaxy, ...). The line labels are only guesses.

Several cells in the result table provide additional information, and these active cells are indicated with a slightly darker colour. For example hovering over the project code link shows the full abstract of the proposal. It also

---
[7]https://admit.astro.umd.edu

displays the acknowledgement, using that project code, which must be copied and pasted into any publication making us of the ALMA data. 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 under the "Observations", display tab 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, only the observations for which a matching spw have been found are displayed. 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 selection tool 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 icon in the top right corner starts the ALMA Request Handler (RH), which will open in a separate tab or window (see Figure 3.5). 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 RH 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.
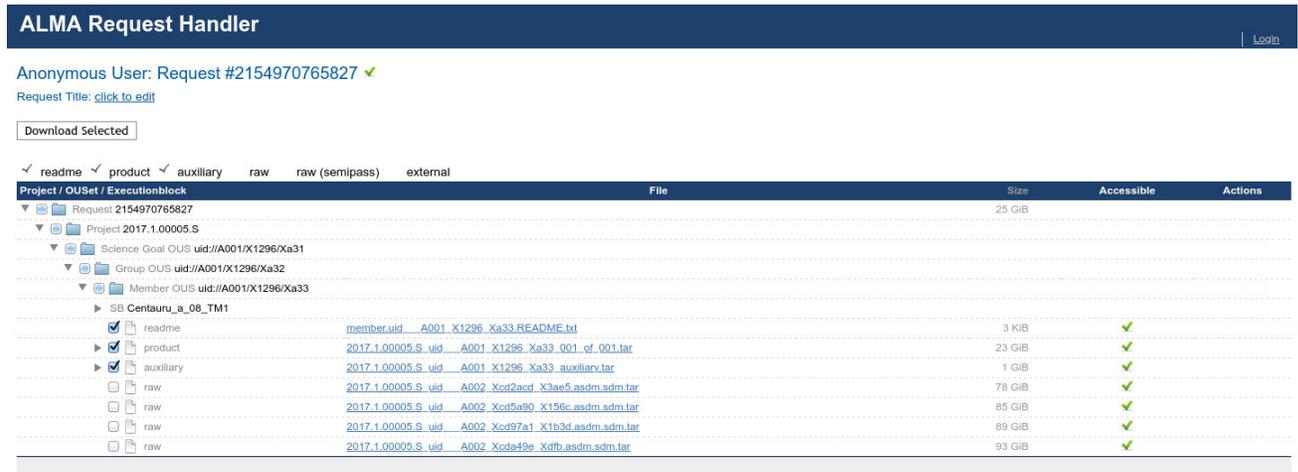

Figure 3.4: The results of a download request. This query selects only one project. Semipass data are not selected in the top boxes, so data having this status are not selected for download.

If authenticated, you will receive an email notification. This email also contains a link to the request in the RH.

Figure 3.5: As above but with the expanded file-contents of the auxiliary tar file.

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 the last 20 (authenticated) requests can be obtained by clicking on the "My requests" link in the top right corner of the RH for authenticated users.

In the RH, select which files you would like to download using the selection boxes as shown in Figure 3.5.

At this stage, you have the option of downloading the available data products as well as the associated raw data as tar files. Either by selecting the check-boxes (see Fig. 3.5) or, as all files on the RH page are links, directly through your browser with a right-click and 'Save as ...'.

For all cycles (except Cycle 0) 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.

In addition, 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.

ALMA data do have quite a deep data hierarchy[8]. Each Science Goal entered into the ALMA Observing Tool is organized into Observing Unit Sets (OUS) by ALMA according to the following rules: observations that can be observed and calibrated together and independently are placed in a Member OUS. If observations that have to be combined from several Member OUS, those Member OUS are placed into the same Group OUS.

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 Member OUS by expanding the SB

---

[8] see e.g. Section 8.1. of https://almascience.org/documents-and-tools/latest/alma-technical-handbook

tree which appears directly below the Member OUS file icon (see Fig. 4.1). 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 Handler results will include the 4 Member OUS for each configuration component, and the products listing will include all protoplanetary disks observed in each of those Member OUS.

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.

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 are 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 transferable. If you want to download data from a different ARC you need to create a new request.

3.5   Visualizing the data

All of the main ALMA science FITS files (except Cycle 0) can be visualized directly in your browser through the Cube Analysis and Rendering Tool for Astronomy (CARTA[9]). CARTA utilizes client-server architecture, parallelization, GPU-accelerated rendering, and various latency-hiding techniques. The user interacts through the client interface running natively in their local web browser. Large images (TB-size ready) hosted from a remote server can be visualized within seconds. This makes it the ideal software to integrate with the ASA.

In the Request Handler view with the expanded products, CARTA icons are shown on the right-hand side of the interface. Once you click on any of them, CARTA will automatically open in a new web browser tab and display the image (please make sure to allow the ASA to open pop-ups in your web browser first).

For a spectral cube, clicking on the the z-profile icon at the top will bring up a subwindow showing the spectrum under the cursor-position along the frequency axis. The Animator tab in the lower right corner allows you to scroll through the channels of a data-cube.

While to open CARTA in the same browser session the RH will open will open a new browser tab for each file, all files are actually available in *any* of those tabs. By selecting additional files from the "File" menu of CARTA you can use the Image list tab, register several images together and blink between them, for example.

For detailed information about CARTA please refer to the user manual[10].

3.6   Downloading the data

ALMA data can be 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, click the Download Selected button in the top-left portion of the window. You will then prompted with two different download methods:

- Download Script: A bash download script is created that can be directly run from the command line on

---

[9]<https://cartavis.github.io>
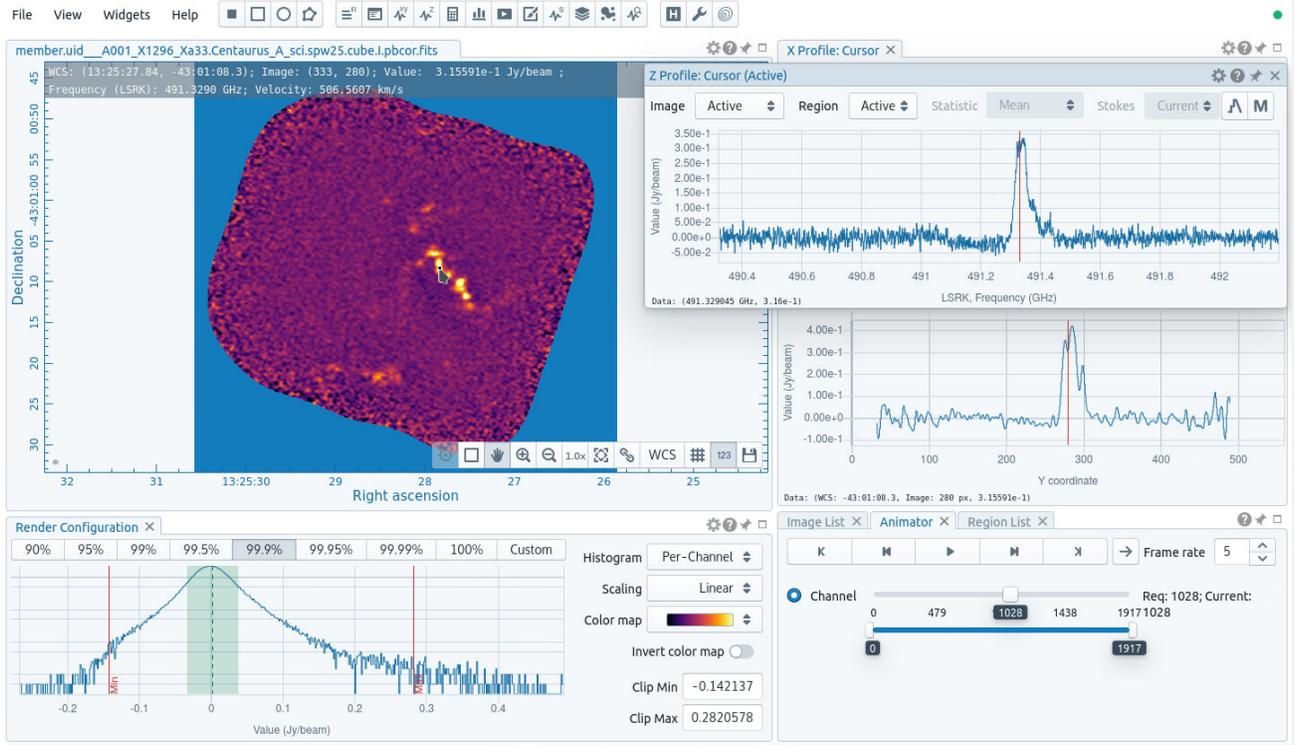[10]<https://carta.readthedocs.io/en/latest>

Figure 3.6: 1-click CARTA remote visualization which is available for all ALMA FITS files.

Linux and Mac.

- 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 u+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 be gentle on the maximum number of parallel streams you are using at any one given time. 5 is recommended.

3.7 Data packaging

3.7.1 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 format[11]. 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.

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 guide[12].

3.7.2 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 Portal[13]. 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.

3.7.3 Cycle 1 and later

In Cycle 1 and later, calibrated MS are no longer included in the package. If only the "product" tar file is downloaded, only the FITS products will be downloaded, as explained above in previous sections. 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 also the "auxiliary" data will have to be downloaded, see Chapter 5.

3.8 Sensitivity estimation

In order to estimate the continuum and line-sensitivities shown on the Archive Query interface, the metadata of the stored Execution Blocks (EB) that are QA0 Pass are taken and filled into the ALMA sensitivity calculator[14]. These metadata are the requested array (12m, 7m or TP), number of polarisations, integration time, source declination, antenna number, observing frequency, bandwidth and channel number as well as the information about the observing conditions. We use the measured Tsys values where they are available and the measurement

---

[11] These 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).

[12] e.g.,https://almascience.org/documents-and-tools/cycle7/alma-qa2-data-products-for-cycle-7

[13] https://almascience.org/alma-data/archive/archive-documentation/delivery-list

[14] https://almascience.org/proposing/sensitivity-calculator

of the precibitable water vapour (PWV) otherwise unless none is available. In the latter case the default PWV for the band is used.

3.9   Precipitable water vapour

Where available the median of all the precipitable water vapour (PWV) measurements is reported on the ALMA query interface. For 7m observations, no such measurements are available and thus, as for the Observing Tool and the ALMA Sensitivity Calculator, the median value of the default PWV values is used.

Chapter 4

Programmatic access

4.1 Jupyter Notebooks

The ASA has been developed so that it can be accessed programatically, i.e. from programs or scripts.

`astroquery`

Astroquery[1][2] 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 3.7.2).

We provide extensive Jupyter Notebook tutorials on how to use the query and download functionality of the ASA programatically: https://almascience.org/alma-data/archive/archive-notebooks. The notebooks detail the use of pyvo[3] and astroquery. All programmatic access to ALMA is based on VO protocols and standards.

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), Server-Side Operations for Data Acces (SODA), DataLink services for ALMA data at https://almascience.org/tap, https://almascience.org/sia2, https://almascience.org/soda and https://almascience.org/datalink, respectively.

External providers offer VO services of the archived FITS products[4]. Note that those products cover only a small fraction of the total number of sources and available spectral range.

---

[1] https://astroquery.readthedocs.io/en/latest/
[2] https://arxiv.org/abs/1901.04520
[3] https://pyvo.readthedocs.io/en/latest
[4] https://jvo.nao.ac.jp/portal/alma/archive.do, https://www.chivo.cl

ALMA Science Archive

Jupyter Notebooks

This page contains Jupyter Notebooks to programatically access the ALMA Science Archive. The notebooks interact through Virtual Observatory standards with ALMA's ObsCore Table Access Protocol (TAP) service.

Queries in TAP are written in the SQL-like Astronomical Data Query Language (ADQL). ADQL queries include spatial queries as well as operations on other properties/columns of the database. This also allows the user detailed control over the returned columns.

In these Jupyter notebook we will exemplify some of the most common queries. For this we will be using the astropy affiliated PyVO client, which is interoperable with other valid TAP services from other observatories.

Table of Contents

0. Installation

Install the required software with

```
pip3 install jupyterlab matplotlib pandas sklearn
pip3 install --upgrade --pre astroquery
```

download the Jupyter Notebooks

```
wget https://almascience.org/alma-data/archive/archive-notebooks/alma-science-archive-notebooks.tgz -O - | tar -xz
cd alma-science-archive-notebooks
```

Figure 4.1: First of the Jupyter Notebook tutorials explaining programmatic access of ALMA.

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, the calibrated 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 the *script* directory. The calibrated MS will be created in the "calibrated" directory.

Please see the Processing section[1] 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 and the hardware requirements for CASA and the pipeline. CASA also provides guides related to pipeline imaging reprocessing[2].

Moreover the ALMA Regional Centres also provide calibrated MS for PIs and archival researchers. At all three ARCs services are in place to generate calibrated MS on demand[3].

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] https://almascience.org/processing
[2] https://casaguides.nrao.edu/index.php/ALMA_Imaging_Pipeline_Reprocessing
[3] https://help.almascience.org/kb/articles/how-do-i-obtain-a-file-of-calibrated-visibilities-measurement-set-for-alma-data

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/kb/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.