ALMA science archive

Elisabetta Liuzzo
INAF- Istituto di Radioastronomia
Italian node of European ALMA Regional Centre
Credits to Marcella Massardi

UniTo – April 2018
Reasons to use archived data

- Check if data are already available for a target
- Check the feasibility of a project looking for similar targets
- Retrieving information on a large sample of objects (e.g. statistics of populations, stacking, ...)
- Retrieving information on a single object but with different configuration (e.g. multifrequency studies) or in different epochs (e.g. variability studies)
- Extracting unpublished information from existing data (e.g. finding additional spectral lines, targets in the same region/time of other observations, )
- For ALMA in particular avoid the stress of competition and oversubscription
ALMA data on the Science Portal
https://almascience.eso.org/alma-data

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive</td>
<td>Holds all public and proprietary data, providing a form-based query tool to the data and allows for anonymous or authenticated download</td>
</tr>
<tr>
<td>Calibrator Catalog</td>
<td>A web-based user interface to the ESO ALMA calibrator catalogue</td>
</tr>
<tr>
<td>Science Verification</td>
<td>Data from observations performed with ALMA, demonstrating the capabilities of ALMA</td>
</tr>
<tr>
<td>Publication</td>
<td></td>
</tr>
<tr>
<td>Acknowledgement</td>
<td></td>
</tr>
</tbody>
</table>
**Calibrator Catalogue**

A web-based user interface to the calibrator database is provided through the

The ALMA Calibrator Source Catalogue is a database of astronomical measurements of calibrator sources, mostly bright quasars in the mm and sub-mm regime. It contains over 11000 ALMA measurements of over 1800 sources (1 July 2015). The most important properties are flux density with uncertainty at a given frequency and angular structure information or limits, and polarization measurements for some calibrators. External information had been provided via 'seed' catalogues, such as from VLA (Calibrator Manual), SMA\(^1\), ATCA, CRATES, and coordinates from VLB1, and we acknowledge the institutions and individuals who contributed this effort. Note that the absolute flux density scale does not include the model uncertainty of the solar system object used as primary amplitude calibrator, and in some cases it can exceed the stated flux density error in the catalogue. Also, the observation and reduction procedures changed in mid-2012, so the measurements before this date may state flux density errors that underestimate the true uncertainty.

A main use of the database is to allow the selection of bandpass and phase calibrators for science observations. It also contains a set of 30 quasars, grid sources evenly distributed over the sky, which are monitored regularly enough to provide amplitude calibration in addition to solar system objects. Calibrator sources are selected either manually during phase 2 preparation or by the astronomer on duty, or automatically via on-line queries during the execution of the observation. In each case suitable criteria regarding visibility, flux density and proximity to the science targets have to be specified.

Calibrator observations, outside of science observations, are currently carried out in a special observing mode and have to be fully integrated into science operations. Equally, data reduction and database ingestion has to be automated via the ALMA Pipeline. Access for users is provided via a web-based user interface through the ALMA Regional Centers, and VO integration of special queries is planned for a future Cycle. The intention is to provide a more complex, public search tool for calibrator sources, which can also be accessed through the Observing Tool and included into the Scheduling Blocks.

The Query Form allows to specify various search criteria regarding position, frequency, flux density and time of the calibrator measurements. Equatorial coordinates in J2000 (ICRS) can be specified for the position search, whereas a search by equatorial coordinates plus declination is needed. For the frequency search, one can select only the range of frequencies. The flux density can be specified in Jy or as a relative value with respect to a calibrator.
ALMA Calibrators
https://almascience.eso.org/alma-data/calibrator-catalogue

It is possible to download (cvs format) only the most recent and/or the data from all the epochs available for selected calibrators or in regions.
ALMA Calibrators

https://almascience.eso.org/alma-data/calibrator-catalogue
Science Verification Data

For general information on the Science Verification process as well as the status and future plans of Science Verification projects, please use the link below:

Science Verification Information

Currently Available Science Verification Data:

We now have several datasets available to demonstrate the early capabilities of ALMA. In some cases these projects were observed before 16 antennas were available and while many of the subsystems were still being tested, so they should not be construed to represent the quality of the data that can be expected from the system as it is today. They are provided here as a means for the user to become acquainted with the ALMA data structure, observing strategies and reduction techniques. Given that the data have been taken during the construction phase, there may be more idiosyncrasies present than will be expected during full operations, so we ask the user to please review carefully the CASA guides provided with the datasets that represent unique observing modes or strategies, as indicated below.

Note that only data with prepared CASA guides are kept up to date with the current CASA release. For the other data, please consult the following knowledge base article: "If my data were calibrated and imaged in CASA 3.3 and I want to redo it, are there resources to help?"

For reference the list of Science Verification targets that was provided with the Cycle 0 Call for Proposals is given in Table 2 which indicates which observations have been completed or are in progress. We do not expect to observe the other sources on that list.

1. **TW Hya: Band 7, high spectral resolution.** Many thanks to the following people for suggesting this source for ALMA Science Verification: Meredith Hughes, Stuart Corder, Chunhua Qi, Karin Oberg, Michiel Hogerheide, Andrea Isella, Dmitry Semenov.

   Additional data on TW Hya is available (without a separate CASA guide) here: Band 3, Band 6.

2. **NGC3256: Band 3, low spectral resolution.** Many thanks to the following people for suggesting this source for ALMA Science Verification: Kazushi Sakamoto, Alson Peck, Satoshi Matsushita, Martin Zwaan.

3. **Antennae galaxies: Band 7, high spectral resolution.** Many thanks to the following people for suggesting this source for ALMA Science Verification: Christine Wilson, Junko Ueda, Francois Boulangé, Nicole Nesvadba, Cinthya Herrera.

   Additional data on Antennae is available (without a separate CASA guide) here: Band 6

4. **M100 Band 3, low spectral resolution.** Many thanks to the following people for suggesting this source for ALMA Science Verification: Prabon Grosbol and Catherine Vlahakis.

   Updated ACA (7-m Array and TP Array) data on M100, complete with a CASA Guide demonstrating the combination of 12-m, 7-m, and TP Array data, are now available. The combined images are also available. The CASA guide is written for CASA version 4.3. An updated version of the previously released 12-m Array data products are provided for CASA 4.3.

5. **SgrA* Band 6, recombination lines.** Many thanks to the following people for suggesting this source for ALMA Science Verification: Andreas Eckart, Stephane Leon, Steve Longmore,
At proposal stage the PI makes some choices and requests needed to reach his purposes. The PI splits the project in

Science Goals

Minimum proposed observational unit including targets in the same sky region that roughly share the same calibration aimed at reaching a requested sensitivity in a given angular (resolution and LAS) and spectral setup.

e.g. Different bands on the same target are in different SG

Different configurations of the array to reach a certain angular scale coverage are in the same SG (ACA+main array)

Each SG is converted into a Scheduling Block:

Minimum observational unit including targets in the same sky region and their calibrators to be observed with the same instrumental setup.

They are the minimum set of instructions to perform an observation.

To allow flexibility of scheduling they last ~30min

Different configurations cannot stay in the same SB.

It is possible that an SB has to be repeated to reach the requested sensitivity.

Then at observative stage we define

Execution Block

Each repetition of the observation of a SB.

This is the minimum observable unit and the minimum data reduction unit (as they include all the calibrators for an observative session), but might not be enough to reach the PI requests.

It is the minimum archive unit!!!
Science goal:
Sources in the same sky region that share the same calibration, spectral setup and PI requests

OUS= Observing Unit Set
Smallest unit for data processing

Sensitivity goals are defined on SG basis

A Group can contain several configurations to be combined in data processing (e.g. several arrays), each of them is a Member. Now for ALMA there is 1 Group/SG

A Member can contain multiple executions of Scheduling Blocks. It is the minimum scheduling entity. Pipeline operates at this level

The Scheduling Block is the smallest entity used for observing

Each repetition of a SB is an Execution Block
Data Quality Assessment

The goal of ALMA Quality Assurance (QA) is to deliver to the PI a reliable final data product that has reached the desired control parameters outlined in the SG, that is calibrated to the desired accuracy and free of calibration or imaging artifacts i.e. ALMA performs **science-goal-oriented service data analysis**

ALMA QA happens on 4 levels:

**QA0:** near-real time verification of weather and hardware issues carried out on each execution block immediately after the observation.

**QA1:** verification of longer-term observatory health issues like absolute pointing and flux calibration.

**QA2:** offline calibration and imaging (using CASA) of a completely observed MOUS.
- Performed by expert analysts with the help of a semi-automatic procedures, based on common practice.
- Calibration can be “Manual” or based on the “Pipeline”
- Imaging so far is always manual (partially depends upon the analyst “taste”)
- It is limited to verify the achievement of the PI requests for each MOUS (do not even consider other data possibilities)
- Results are archived and given to the PI.

**QA3:** (optional) PIs may request rereduction, problem fixes, possibly reobservation
Project travels & archive mirrors
Data format

**ALMA Science Data Model (ASDM)** Final archived product from each observation
Each has an unique hexadecimal name (eg uid://A002/X2fed6/X3f).
Each contains the meta-data (headers, descriptions of the observation setup, etc), and the binary data (the raw data).

The first step of any data processing is importing the ASDM in the format suitable for the software used.

**Measurement Set (MS)** Data format used in CASA
Constituted by several tables referring each other and collecting most (not all!) the information in the ASDM.
What is in the archive?

For each project the main deliverables are
Raw Data, Calibration Scripts and Tables

Users need to run the proper version of CASA to generate the Calibrated Data. The resulting calibrated data is considered science-ready.

As a consequence of the process only data that passed QA2 (at least in part) are in the archive

Some Imaging Products are delivered too, as result of QA2 processing
(in Early Science provided on a best effort basis, not necessarily science-ready)

a) for Line Observations:
- continuum-subtracted (where needed) image cubes at the requested resolution
- a continuum image for all line-free channels (where possible)

b) for Continuum Observations:
- a continuum image combining all SPWs

Images in the archive are provided as starting point on the way to obtain the final images and a valuable basis for archive researchers
(i.e. they are not considered science-ready!!)
What is in the archive?

Images in the archive might cover only a fraction of the spectrum available.

Spectrum observed (and available in the raw data)

Imaged data for QA2

Different data and PI requests on different sources generate different products. In the archived images but raw data contain the full spectral windows.
Differences among the cycles

Early science cycles might differ in product formats, pipeline availability, CASA version to run calibration scripts.

Cycle 0 caveats:
- Packaged differently from other cycles (check “Delivery lists” contents to download only images)
- Archive download does not refer to data tree structure
- Calibrated with CASA 3.2-3.4 (updates since then)
- No pipeline available
- Many Knowledgebase articles available to deal with them
  → Rule of Thumb: download them and reduce them from scratch, if possible with a new CASA version. Rely on experts for it

Cycle 1 onwards:
- Allow download of “ligth” (<1GB) products (QA2 images and README) and/or “heavy” (>10GB) raw data
- Check in the downloaded README files and script the CASA version used
  → Rule of Thumb: download the product first, check that they are what you need and their quality, then download the raw data. Reproduce the products running the scripts. Change the images modifying the imaging scripts. Rely on experts for changes in the calibration
Tha Science ALMA Archive
https://almascience.eso.org/alma-data/archive

Archive
Archive Query

Documentation
We provide a comprehensive ALMA Science Archive Manual.

Data delegation
PIs can grant access to one of their projects to a registered ALMA user by logging into the Science Portal, going to the user profile page in the top right corner and then adding delegates in the "Project delegation" tab.

Cycle 0 content
Please go here to see the content of the Cycle 0 deliveries.
Data delegation

PIs can grant access to one of their projects to a registered ALMA user by logging into the Science Portal, going to the user profile page in the top right corner and then adding delegates in the "Project delegation" tab.

Cycle 0 content

Please go here to see the content of the Cycle 0 deliveries.
<table>
<thead>
<tr>
<th>Position</th>
<th>Energy</th>
<th>Time</th>
<th>Polarisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source name</td>
<td>Frequency</td>
<td>Observation date</td>
<td>Polarisation type</td>
</tr>
<tr>
<td>(Resolver)</td>
<td>Bandwidth</td>
<td>Integration time</td>
<td></td>
</tr>
<tr>
<td>Source name</td>
<td>Spectral resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ALMA)</td>
<td>Band</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA Dec</td>
<td>Target list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactic</td>
<td>Angular resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target list</td>
<td>Largest angular scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angular resolution</td>
<td>Field of view</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>Project</td>
<td>Publication</td>
<td>Options</td>
</tr>
<tr>
<td>Line sensitivity</td>
<td>Project code</td>
<td>Bibcode</td>
<td>View:</td>
</tr>
<tr>
<td>(10 km/s)</td>
<td>Project title</td>
<td>Title</td>
<td>- observation</td>
</tr>
<tr>
<td>Continuum sensitivity</td>
<td>PI name</td>
<td>First author</td>
<td>- project</td>
</tr>
<tr>
<td>Water vapour</td>
<td>Proposal authors</td>
<td></td>
<td>- publication</td>
</tr>
<tr>
<td></td>
<td>Project abstract</td>
<td></td>
<td>- public data only</td>
</tr>
<tr>
<td></td>
<td>Publication count</td>
<td></td>
<td>- science observations only</td>
</tr>
<tr>
<td></td>
<td>Science keyword</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The ALMA archive: help
The query runs on the raw data so returns one entry per target per Execution Block. It is possible that for a project several rows are displayed for the same source. Projects that contain many sources, many Sbs or mosaics might return many lines. Columns values are only indicative. Data structure can be more complex than what shown.
The ALMA archive: download manager

3) Select the data you want

**ALMA Request Handler**

Anonymous User: Request #1896309891

Request Title: Click to edit

- Download Selected

<table>
<thead>
<tr>
<th>Project / OUSet / ExecutionBlock</th>
<th>File</th>
<th>Size</th>
<th>Accessible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request 1896309891</td>
<td>2013.1.01151.S readme.txt</td>
<td>759.8MB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2013.1.01151.S uid__A001_X144_X210_001_of_001.tar</td>
<td>2.3GB</td>
<td></td>
</tr>
</tbody>
</table>

Remember that a Member OUS is the smaller data processing unit

Download products only for quick view of images

- Product data are typically <1GB
- Raw data for whole projects are typically >10GB
- Processing might increase folder size by factors 2-8
Don’t struggle on it!!!
You can ask an account on our ARC cluster to deal with ALMA data!

Just sent an email to help-desk@alma.inaf.it indicating the reason of your request
And visit our webpage http://www.alma.inaf.it/index.php/The_ARC_cluster
The ALMA archive: download manager

4) Choose the download method

Choose one of the following download methods:

- **Download Script**
  
  The downloads are scripted for you. You just need to execute the script from the command line. [Help]

- **Download Manager**
  
  ALMA’s download manager is launched as a browser applet. This is a simpler, more user-friendly way to download files in parallel, allowing you to pause and resume.

- **Web Start Download Manager**
  
  ALMA’s download manager is launched as a desktop application via Java Web Start. It will not stop if you close your browser.

- **File List**
  
  View a text file containing a list of URLs. This is useful for using third-party download manager’s such as DownThemAll.
The ESO telbib

http://telbib.eso.org/
### The ESO telbib

**ESO Telescope Bibliography**

Refine Search

- **Year**
  - 2015 (40)
  - 2014 (97)
  - 2013 (65)
  - 2012 (20)

- **Journal**
  - ApJ (121)
  - A&A (54)
  - MNRAS (16)
  - Nature (11)
  - PASJ (6)

- **Instrument**
  - ALMA Bands (222)
  - LABC0A (14)
  - XSHOOTER (8)
  - FORS2 (5)
  - SHFI (5)

Results 1 - 25 of 222 found for (instrument:ALMA_Bands)

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Title</th>
<th>INSTRUMENTS</th>
<th>ACCESS TO DATA</th>
<th>Fulltext ADS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Sakai, Takeshi et al.</td>
<td>ALMA Observations of the IRDC Clump G34.43+00.24 MM3: DNC/HNC Ratio</td>
<td>ALMA_Bands</td>
<td>2011.0.00656.S</td>
<td>2015ApJ...803...70S</td>
</tr>
</tbody>
</table>

Direct link to the ALMA Archive
ASTROquery.alma

- Python library for archival query (not developed by ALMA)
- Allows batch searches (i.e. lists of sources) and full download
- Allows the same criteria as the archive
- Allows scripting for downloads (if anyone needs we can distribute examples)
- However, it is computer sensitive (download may not work...)
When untarred, the Product Package standard directory structure contains:

|-- project_id/
 | |-- sg_ouss_id/
 | | |-- group_ouss_id/
 | | | |-- member_ouss_id/
 | | | | |-- README ..............................
 | | | | |-- product/ .......................... all the imaging products as result of QA2
 | | | | |-- calibration/ ................. calibration and flagging tables
 | | | | |-- qa/ .......................... diagnostic plots generated during QA2
 | | | | |-- script/ ..................... the scripts necessary to regenerate the products
 | | | | |-- log/ ............................. CASA log files from QA2 calibration and imaging
 | | | |-- raw/ .......................... for calibration move it in the products folder at the right level (follow the README)

What is in the packages?

Untarred products only

Untarred ASDM (raw data)

2013.1.00278.S (2)
Untarred ASDM (Raw data)

2013.1.00278.S
Untarred products

2013.1.00278.S_uid___A001_X120_X102_001_of_001.tar

2013.1.00278.S_uid___A002_Xa0b40d_X3cb8.asdm.sdm.tar

downloadRequest9
98112925.sh
What to do after download?

1) Untar the packages

2) Look at weblog and/or QA reports

3) Read the README file and follow the instructions: typically
   - Launch the correct CASA (with pipeline) version in the script folder
   - Run the “Script_for_PI” to generate the calibrated MS
   - Run the “Script_for_Imaging” to regenerate the images

4) Edit the scripts where needed according to your purposes
What is in the packages?

After running the ScriptForPI.py

|-- project_id/
 | |-- sg_ouss_id/
 | | |-- group_ouss_id/
 | | | |-- member_ouss_id/
 | | | | |-- README ...................... important summary of the contents
 | | | | |-- product/ .................... all the imaging products as result of QA2
 | | | | |-- calibration/ .................. calibration and flagging tables
 | | | | |-- qa/ ................................ diagnostic plots generated during QA2
 | | | |-- script/ ........................ the scripts necessary to regenerate the products
 | | | |-- log/ ............................ CASA log files from QA2 calibration and imaging
 | | |-- raw/ ............................. moved in the main folder from raw data download
 | |-- calibrated/ ........................ calibrated ms, flagging and calibration tables
Caveats

• This is the CURRENT version of the archive

• Philosophy will remain the same, access interfaces and information available might change in the future

• There are differences between cycles

• There are differences between pipeline and manual data reduction, calibration and imaging

• Images are not science-ready!!!
In publications with ALMA data!

Acknowledgement Statement:

“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.”

(Can be found in the SP, on the ‘ALMA-Data’ page or in the Archive)