

Archive interfaces

Felix Stoehr, Marcella Massardi, George Bendo, Toma Bădescu

ALMA Science Archive School
Italian ARC node headquarters, 5-7 October 2022

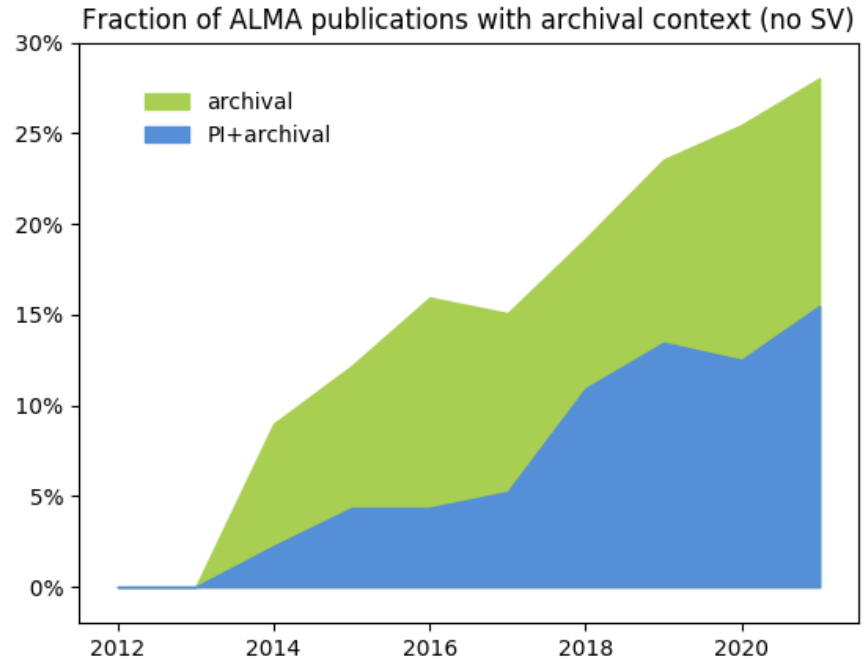


1. introduction

Context

Science Archives

- The Principle Investigators suggest observations and if granted time, get those observed.
- After typically 12 months they become public
- Science Archives are a major part of modern astronomical facilities
- Enables archival researches to work on huge amounts of data immediately



In a nutshell

ALMA Science Archive

- 10 years of observations collected
- Science categories from the solar system to cosmology
- 1.4 PB of data
- 45 000 observations are already publicly accessible
- >10 000 of those have not yet been published

- The ASA is a **treasure-trove** for archival research!

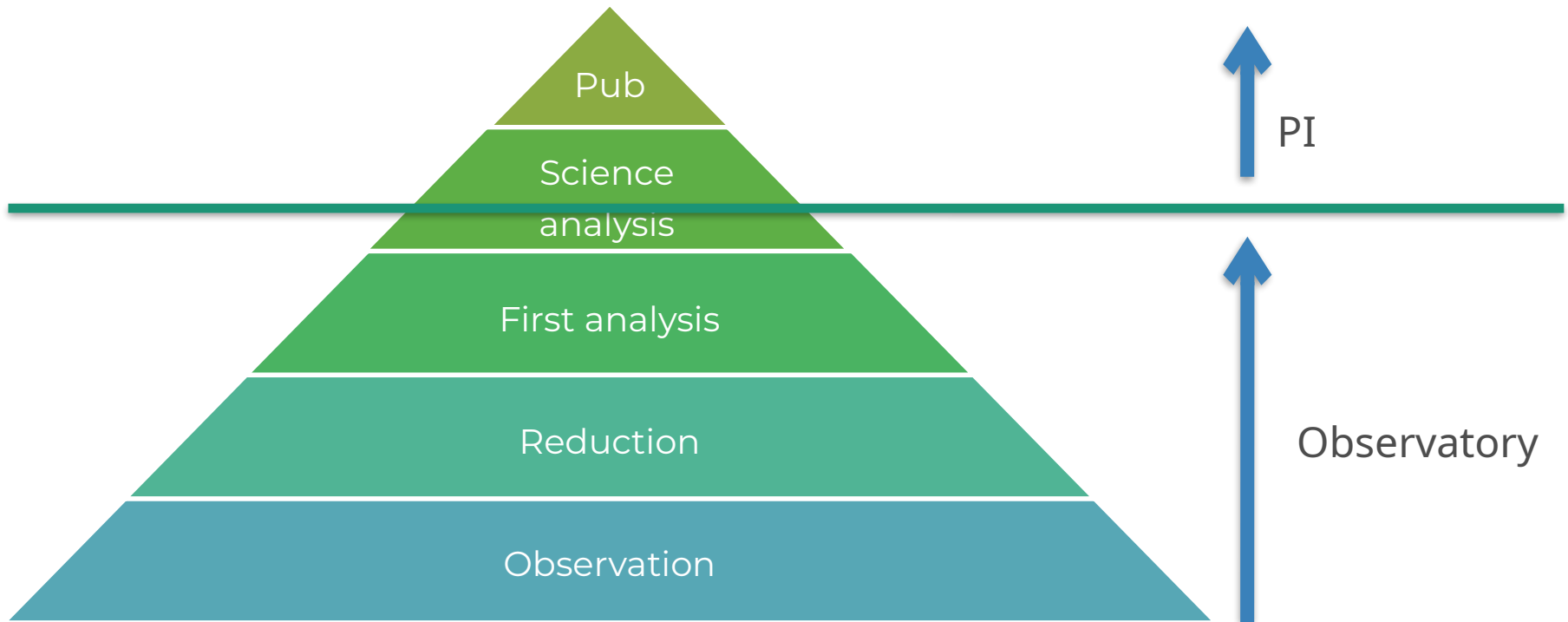
Context

Best practices 2022

- Physical quantities
- Unscoped search
- Observations, Proposals, Publications
- Target-list upload
- Previews
- Modern user-experience
- Programmatic access (VO)
- Metadata are public
- Science-grade products + PL
- Anonymous downloads
- Self-describing FITS files
- Parallel downloads
- Authors must cite data-use
- Frequent Reprocessing
- **NEW**: Science platforms

Vision

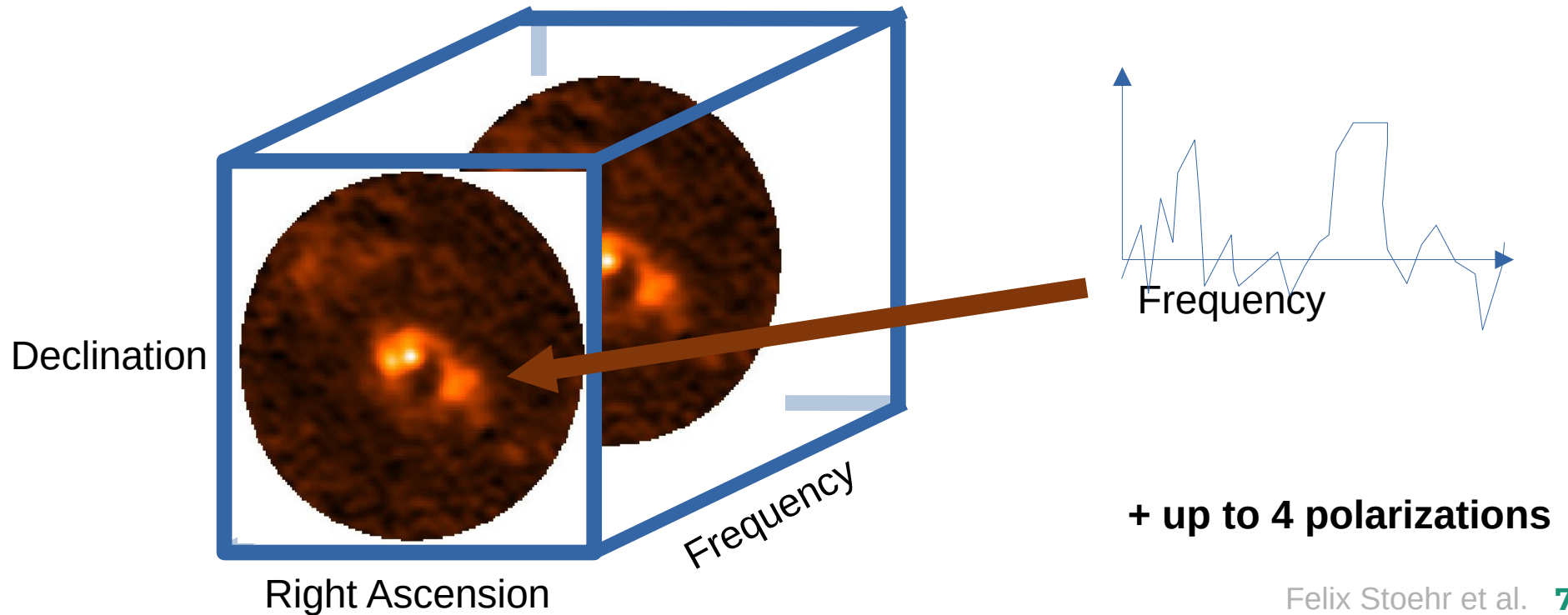
Observatories: more responsibility



Vision

3D data challenge

- ALMA is producing 2D images and **3D data cubes**



Vision

3D data challenge

- ALMA is producing 2D images and **3D data cubes**
- More and more other facilities from x-ray to optical, too
- Huge challenges for visualization and analysis (3D source-extraction)

- ALMA Science Archive:
 - Spectral coverage viewer
 - Previews
 - CARTA visualization

Vision

fastronomy

It is not enough that people can do what they need to do.
They need to be able to do it **fast!**



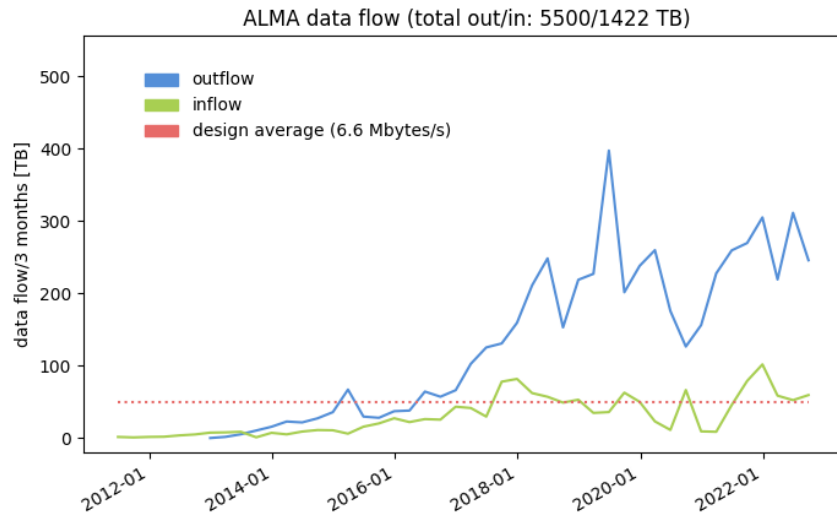
fastronomy

Let scientists concentrate on science

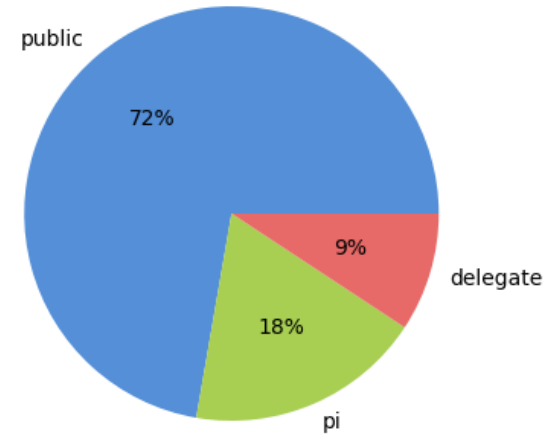
2. ALMA Science Archive

ALMA Science Archive

Access



Downloaded ALMA data (total: 18262 users)



ALMA Science Archive

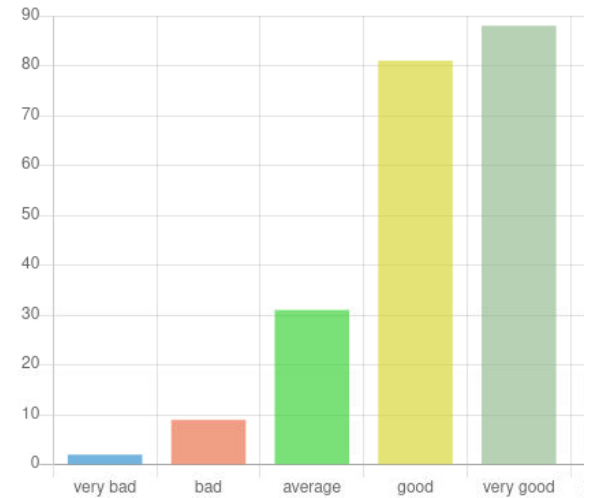
Towards ALMA Science Archive 1.0

Recent developments

- CARTA
- Text-based similarity search
- Complete VO suite
- Jupyter Notebook tutorials
- Object-type search
- Previews

User survey 2019
statistics and 43 pages
of comments

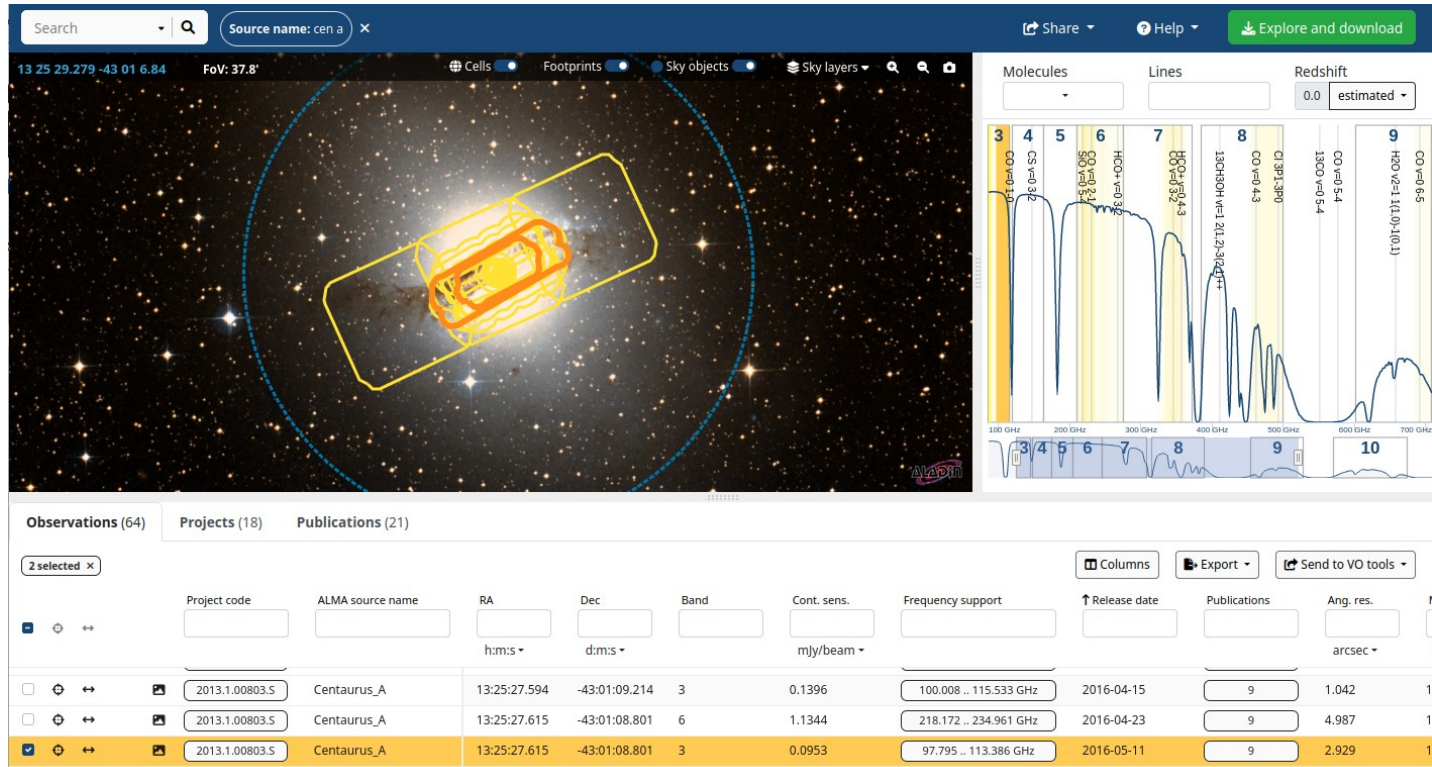
Please rate our new query interface at
<https://almascience.org/asax>



ALMA Science Archive

almascience.org/aq

(live demo)



ALMA Science Archive

Object-type search

All SIMBAD and NED objects that fall into ALMA observations can be **queried** for on the ASA and are **displayed** on the interface.

First Author **Object type**

Authors Clear all Best match

Enter object type

- (Micro)Lensing Event (3)
- Absorption Line system (537)
- Absorption line system (647)
- Active Galaxy Nucleus (1035)
- Association of Stars (457)
- Asymptotic Giant Branch Star (...)

Observations (2682)

Band	Cont. sens.	Frequency support	Release date	Publications
5	0.1181	343.08..358.84GHz	2012-12-06	2
4	0.9115	330.25..346.11GHz	2012-12-06	5
0	0.1136	337.01..353.00GHz	2012-12-06	2
3	0.4848	337.03..353.01GHz	2012-12-20	3
3	0.4848	337.02..353.01GHz	2012-12-20	3
3	0.5346	337.01..352.99GHz	2012-12-20	3

Object type
Only show ALMA observations for which SIMBAD or NED contain at least one object with the given object type in ALMA's observation footprint. If 'Best match' is selected, then show those ALMA observations, where the selected object type matches the object type of the SIMBAD or NED source that has most likely been targeted by the ALMA PI.

WARNING: Use 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.

Description
The SIMBAD or NED objects that fall within an ALMA observation's footprint are retrieved. That ALMA observation is shown in the result table in case any SIMBAD or NED object with the selected object type falls within the ALMA observation's footprint.

ALMA observations mostly target a single object per footprint. We try to identify the 'best' object out of the many SIMBAD or NED objects for each ALMA observation by taking the object that has the best combination of having a close name to the ALMA source name given by the PI, being a strong source and finally of being close to the RA/Dec position given by the PI.

Search Source name: Cen A

13 25 27.009 -43 01 9.15 FoV: 46.27"

Cells Footprints Sky objects Sky layers

● Sky object [HWK2003] SJ1

Main id: [HWK2003] SJ1
RA Dec: 13:25:27.000 -43:01:09.000
Long type: Radio-source
Resolver type: SIMBAD

Observations (81) Projects (20) Publications (24)

ALMA Science Archive

VO tools: Aladin and Topcat

- Start Aladin, Topcat or both
 - `java -jar topcat-full.jar` (download link)
 - `java -jar Aladin.jar` (download link)
- On the ALMA Science Archive Interface click on the rocket symbol and choose between all displayed rows and the selected rows
- In order to see the observation footprints in Aladin, click on the “ALMA Science Archive” entry on the right and then scroll until the FoV appears. Click on those.

ALMA Science Archive

Complete VO suite

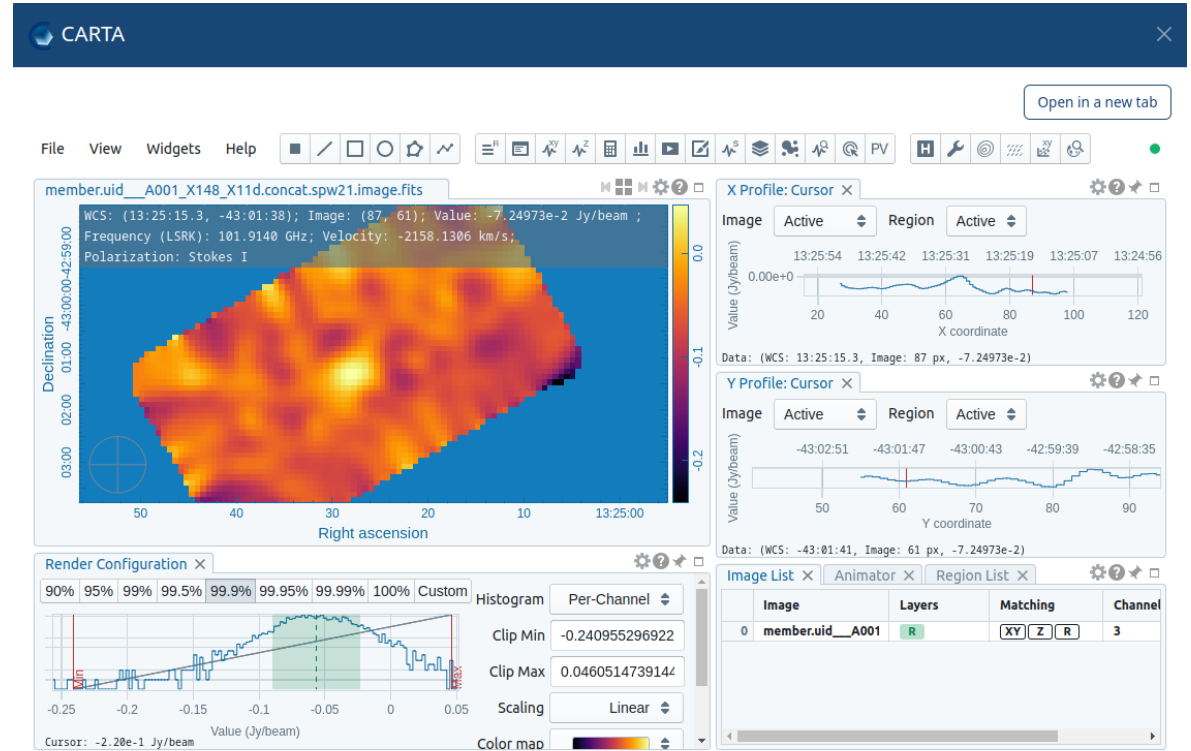
We now have a **full suite of VO tools** in place including

- TAP (ObsCore)
- SIAv2
- DataLink (now used in TAP and SIAv2 outputs)
- SODA for cut-outs

ALMA Science Archive

CARTA

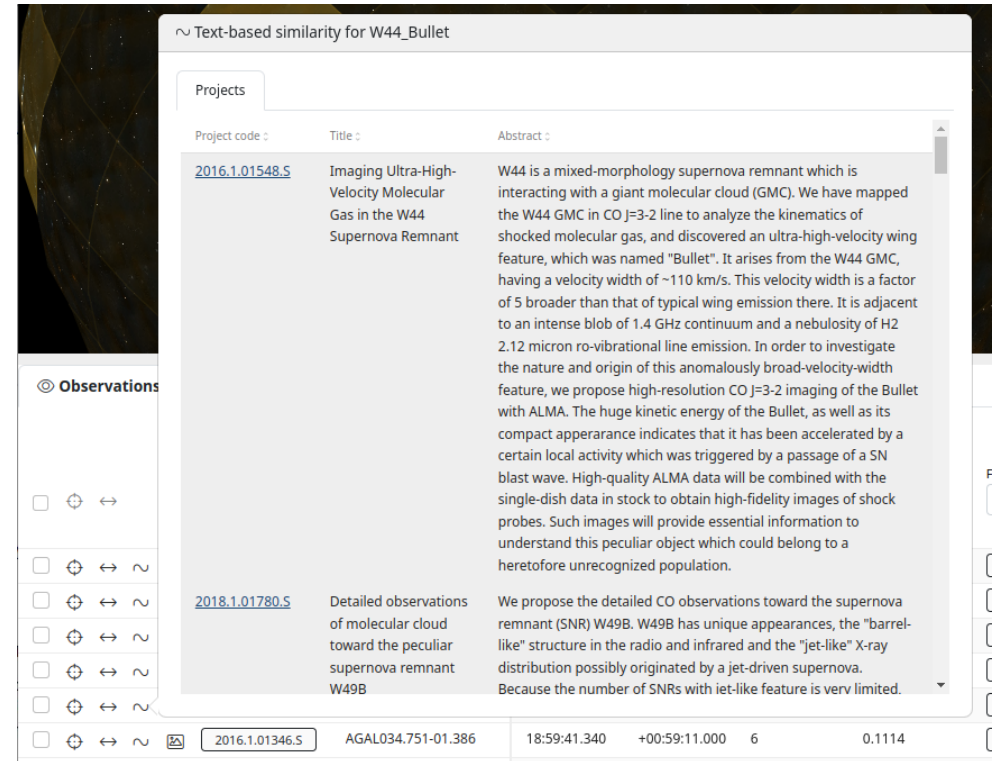
CARTA 3.0 is available on the ASA on the query page as well as on the download page.



ALMA Science Archive

Text-based similarity

The ASA can show projects or publications that are similar to a given one based on **state-of-the-art text similarity**. (Idea and proof-of-concept: Alejandro Barrientos)



The screenshot displays a web interface titled "Text-based similarity for W44_Bullet". It features a table with columns for "Project code", "Title", and "Abstract". The table lists two projects:

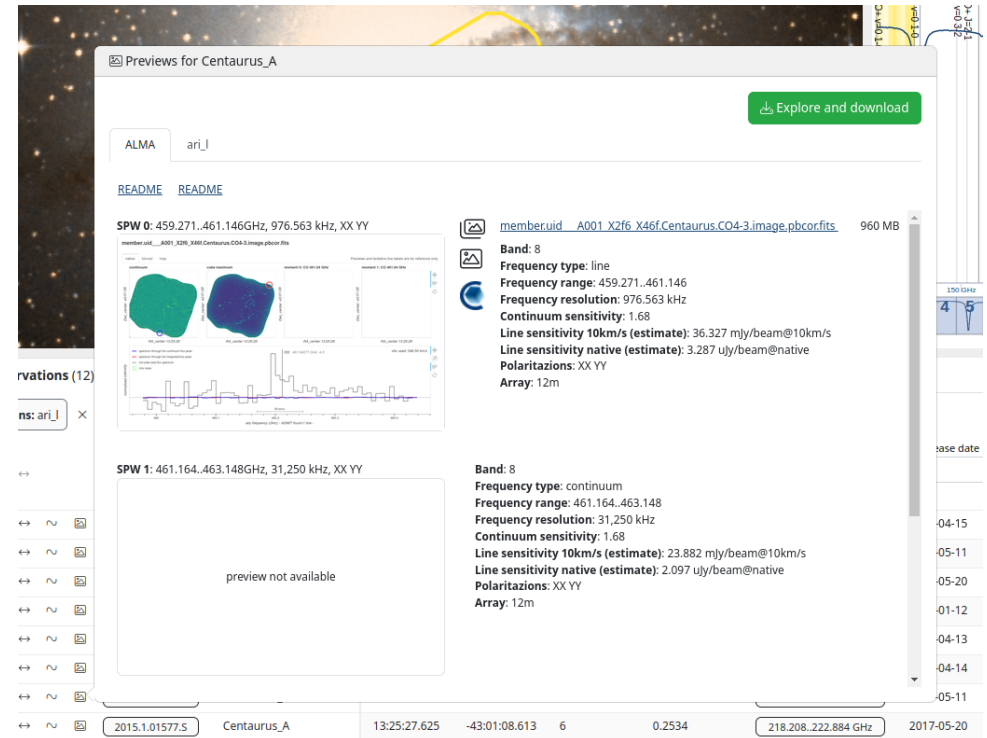
Project code	Title	Abstract
2016.1.01548.S	Imaging Ultra-High-Velocity Molecular Gas in the W44 Supernova Remnant	W44 is a mixed-morphology supernova remnant which is interacting with a giant molecular cloud (GMC). We have mapped the W44 GMC in CO J=3-2 line to analyze the kinematics of shocked molecular gas, and discovered an ultra-high-velocity wing feature, which was named "Bullet". It arises from the W44 GMC, having a velocity width of ~110 km/s. This velocity width is a factor of 5 broader than that of typical wing emission there. It is adjacent to an intense blob of 1.4 GHz continuum and a nebulosity of H2 2.12 micron ro-vibrational line emission. In order to investigate the nature and origin of this anomalously broad-velocity-width feature, we propose high-resolution CO J=3-2 imaging of the Bullet with ALMA. The huge kinetic energy of the Bullet, as well as its compact appearance indicates that it has been accelerated by a certain local activity which was triggered by a passage of a SN blast wave. High-quality ALMA data will be combined with the single-dish data in stock to obtain high-fidelity images of shock probes. Such images will provide essential information to understand this peculiar object which could belong to a heretofore unrecognized population.
2018.1.01780.S	Detailed observations of molecular cloud toward the peculiar supernova remnant W49B	We propose the detailed CO observations toward the supernova remnant (SNR) W49B. W49B has unique appearances, the "barrel-like" structure in the radio and infrared and the "jet-like" X-ray distribution possibly originated by a jet-driven supernova. Because the number of SNRs with jet-like feature is very limited.

Below the table, there is a section for "Observations" with a list of observation IDs and their corresponding project codes. At the bottom, there is a search bar and a table of observation details for the selected project.

ALMA Science Archive

Previews

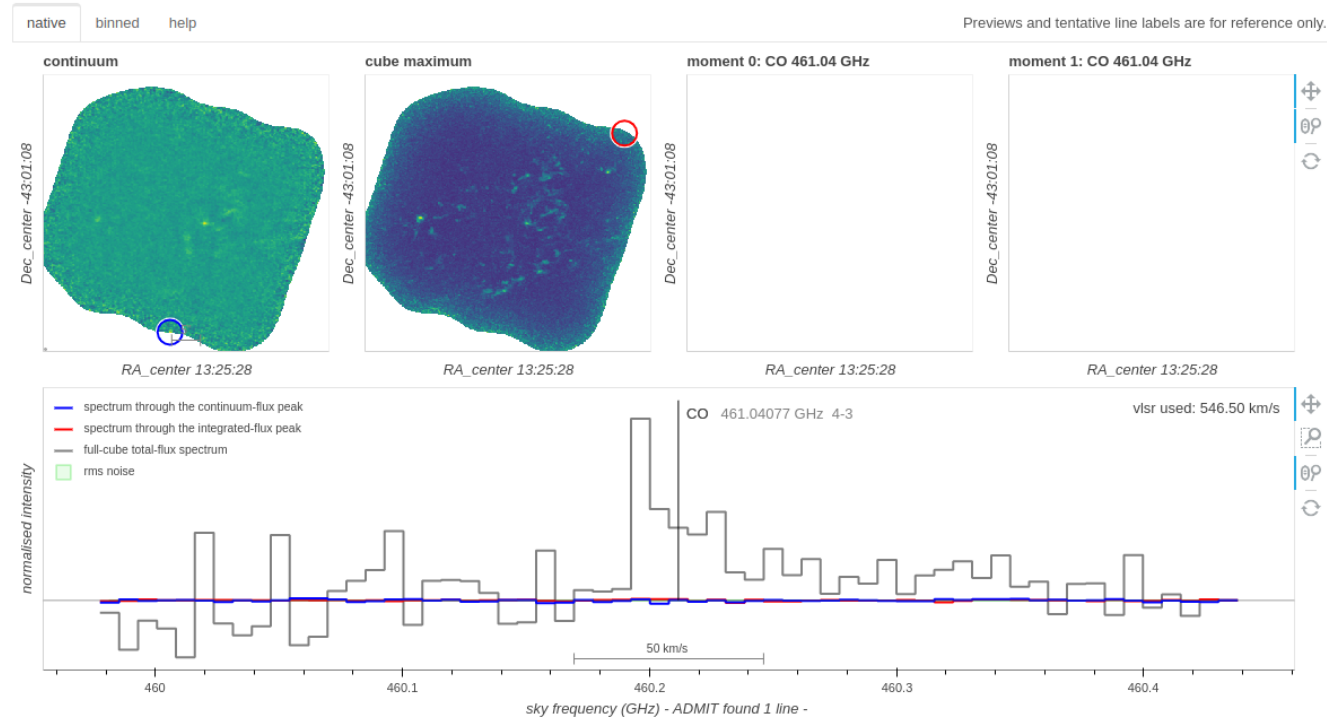
Previews for (nearly all) ALMA FITS files are available directly from the query form. The previews are static files but nevertheless fully **interactive**. We use the ALMA Data Mining Tool-Kit (ADMIT) to run line-finding and tentative **line-identification** on the ALMA images.



ALMA Science Archive

Previews

member.uid A001_X2f6_X46f.Centaurus.CO4-3.image.pbcor.fits



ALMA Science Archive

Some other noteworthy features

- Spectral coverage viewer (see 3D challenge), including lines that might be covered
- Search for real physical quantities (e.g. sensitivity, resolutions)
- Search the entire n-dimensional data cube. I.e. observations, projects and publications together “show all public but not yet published observations”)
- Slide smoothly through sky backgrounds
- Get a calendar event when proprietary data will become public
- ...
- Data downloads will be discussed in George’s presentation

4. Documentation

Documentation

ASA manuals

- Archive Manual

<https://almascience.eso.org/alma-data/documents-and-tools/latest/science-archive-manual>

- Video tutorials

<https://almascience.eso.org/alma-data/archive/archive-video-tutorials>

- ALMA Archival data – a Primer

<https://almascience.eso.org/documents-and-tools/cycle9/archive-primer>

- Jupyter notebooks

<https://almascience.eso.org/alma-data/archive/archive-notebooks>

Documentation

Jupyter notebook tutorials

Tutorials for programmatic access to the ASA have been published on the Science Portal as Jupyter Notebooks.



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
 - 1. Query one source
 - 2. Query a catalogue of sources
 - 3. Query by proposal and IDs
 - 4. Query by science keyword
 - 5. Query by spatial resolution
 - 6. Query by covered frequency range
 - 7. Query by Sensitivity
 - 8. Query using Astroquery.ALMA
 - 9. Download data after query
-

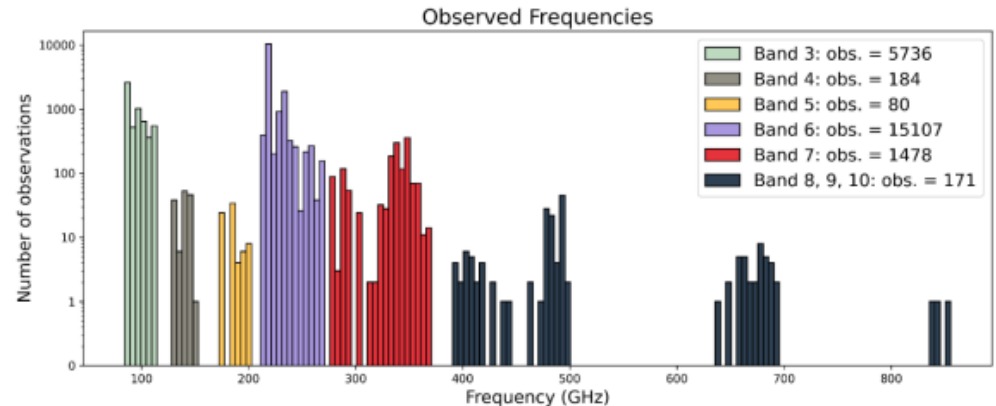
4. Additional tools

Additional tools

ALMiner

<https://github.com/emerge-erc/ALminer>

alminer is a Python-based code to effectively query, analyse, and visualize the ALMA science archive. It also allows users to directly download ALMA data products and/or raw data for further image processing.



Additional tools

ALMA Data Mining Toolkit (ADMIT)

<https://admit.astro.umd.edu>

```
git clone --branch python3 https://github.com/astroumd/admit.git
```

A toolkit that works on ALMA data using CASA to

- detect sources
- find lines
- identify lines

The ALMA previews make use of ADMIT for the line finding and line identification.