Using the ALMA Science Archive data

Marcella Massardi

EUROPEAN ARC
ALMA Regional Centre || Italian

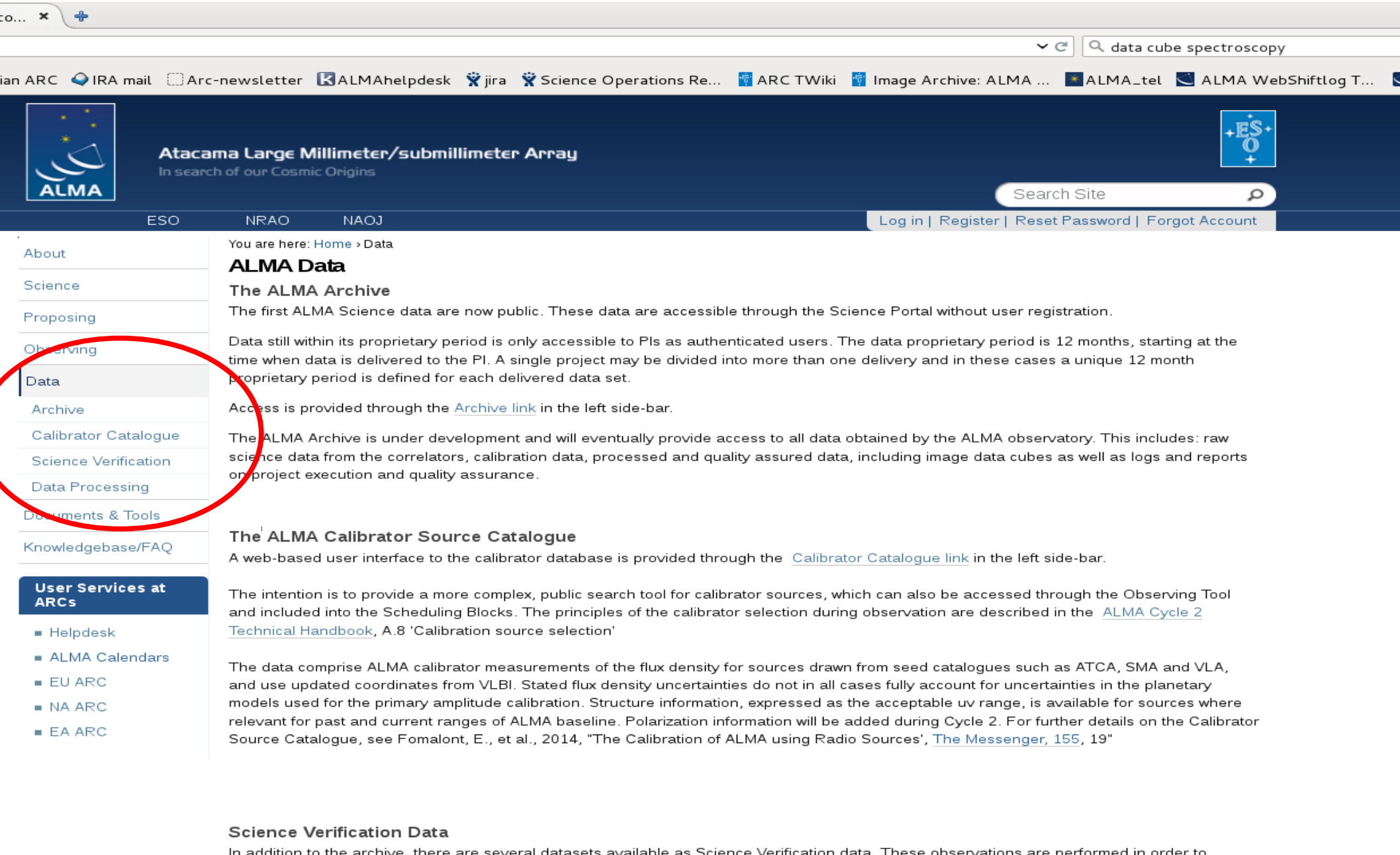
Reasons to use archived data

- 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- Extracting unpublished information from existing data (e.g. finding additional spectral lines, targets in the same region/time of other observations,)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**

	Proposal submission	Archive mining
Time to get data	✖	✚
Amount of data	✖	✚
Data homogeneity	✚	✖
Adherence to idea	✚	✖

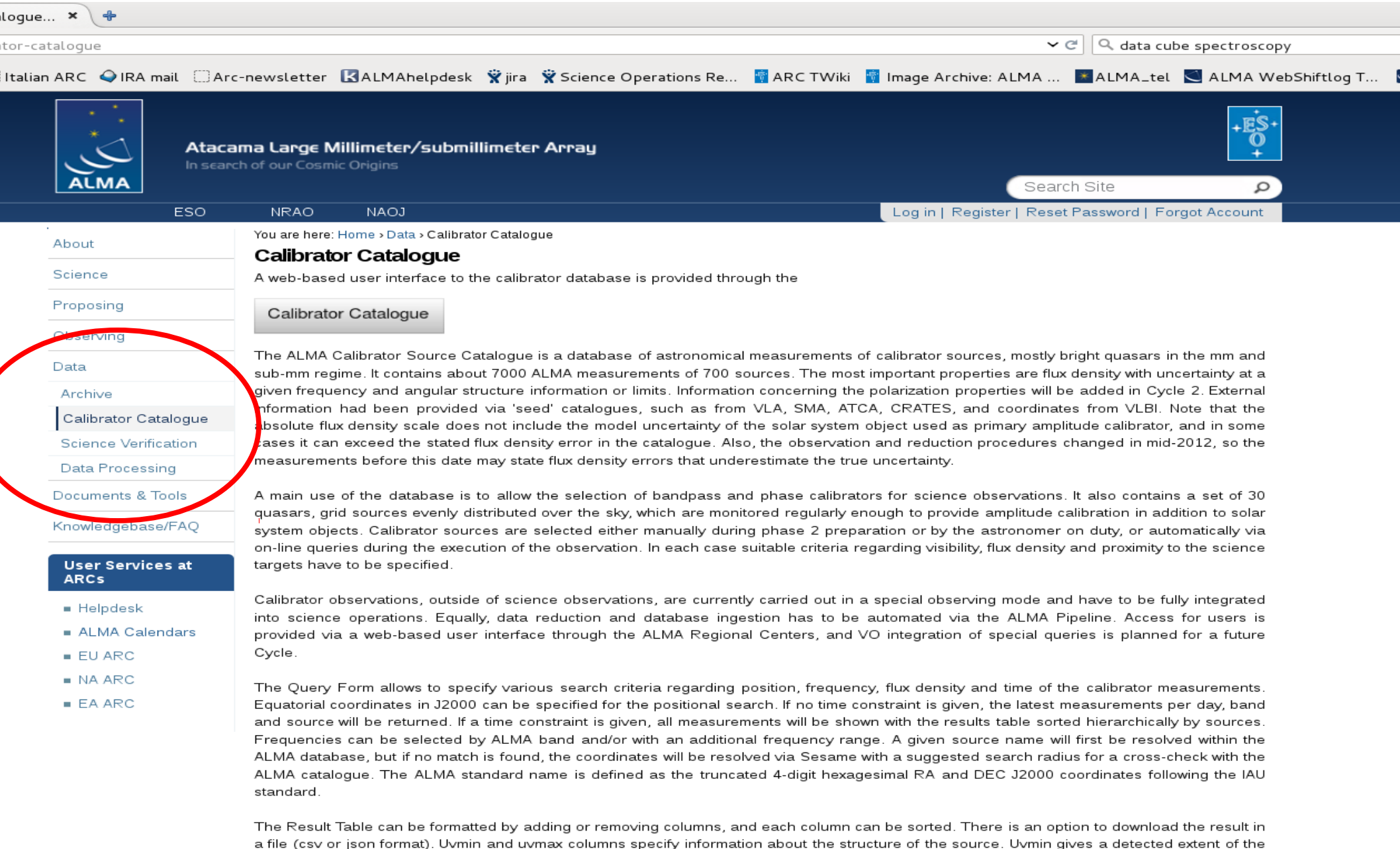
ALMA data on the Science Portal

https://almascience.eso.org/alma-data



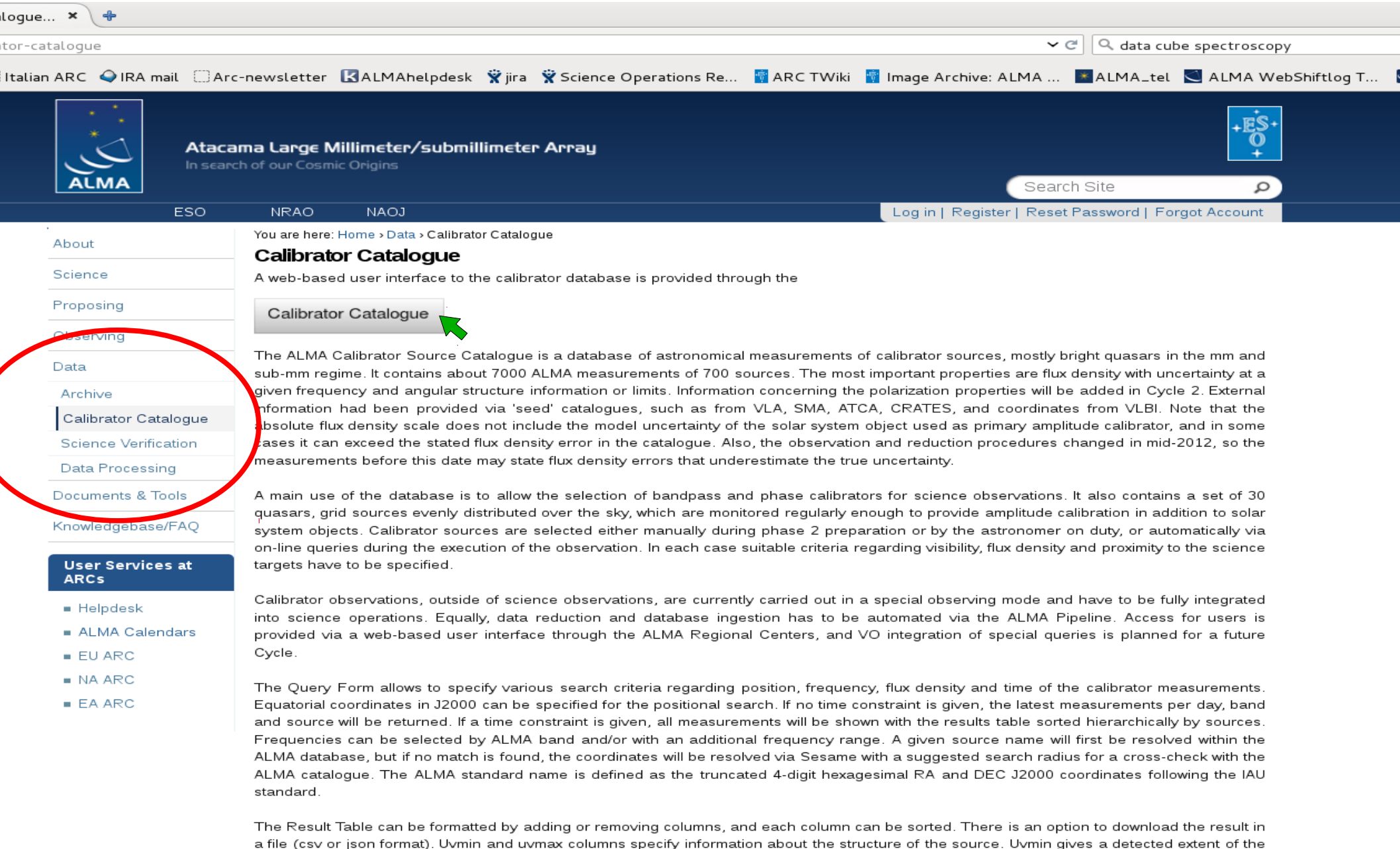
ALMA Calibrators

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



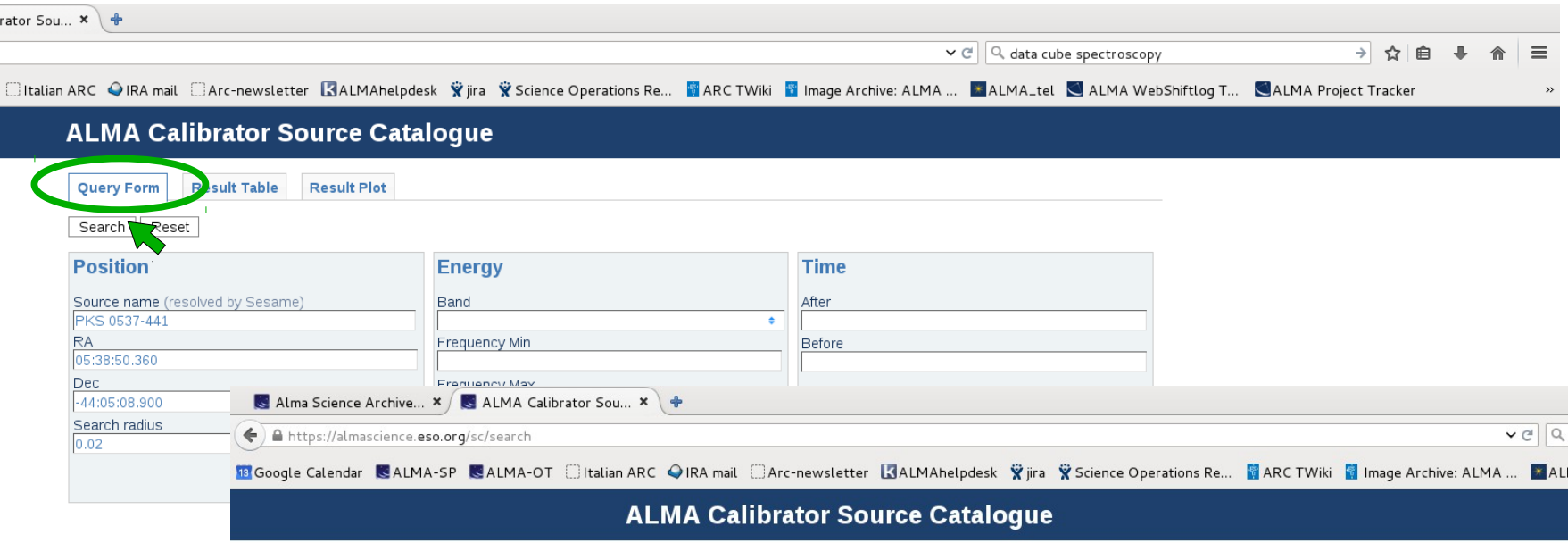
ALMA Calibrators

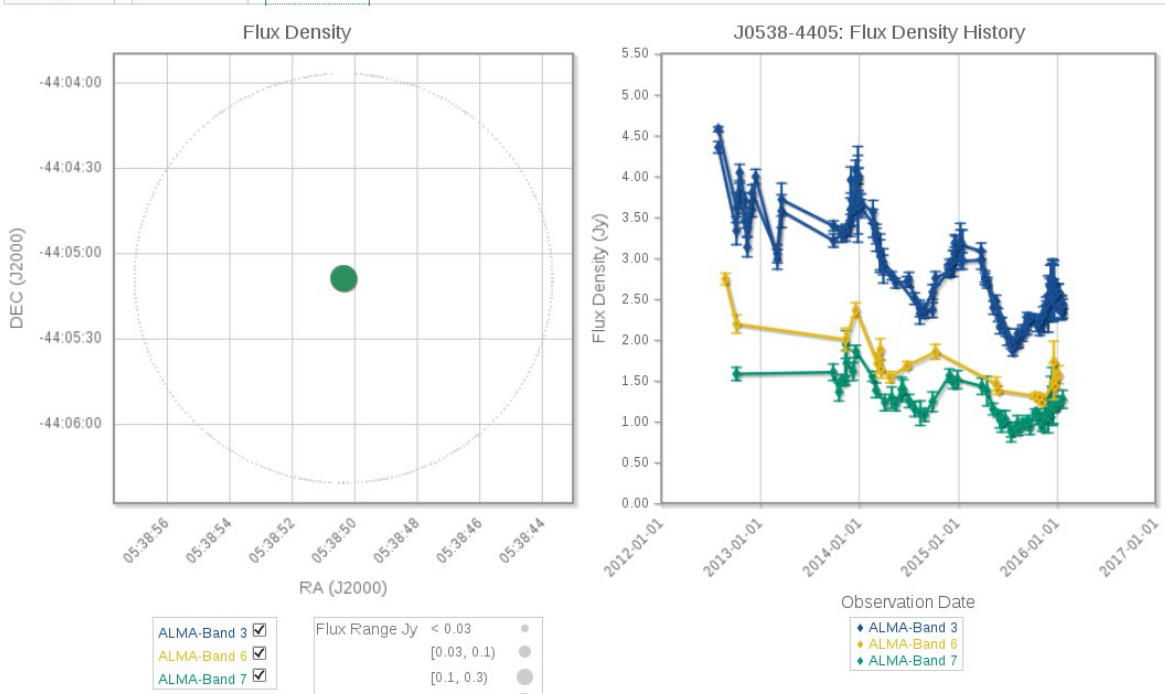
https://almascience.eso.org/alma-data/calibrator-catalogue



ALMA Calibrators

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

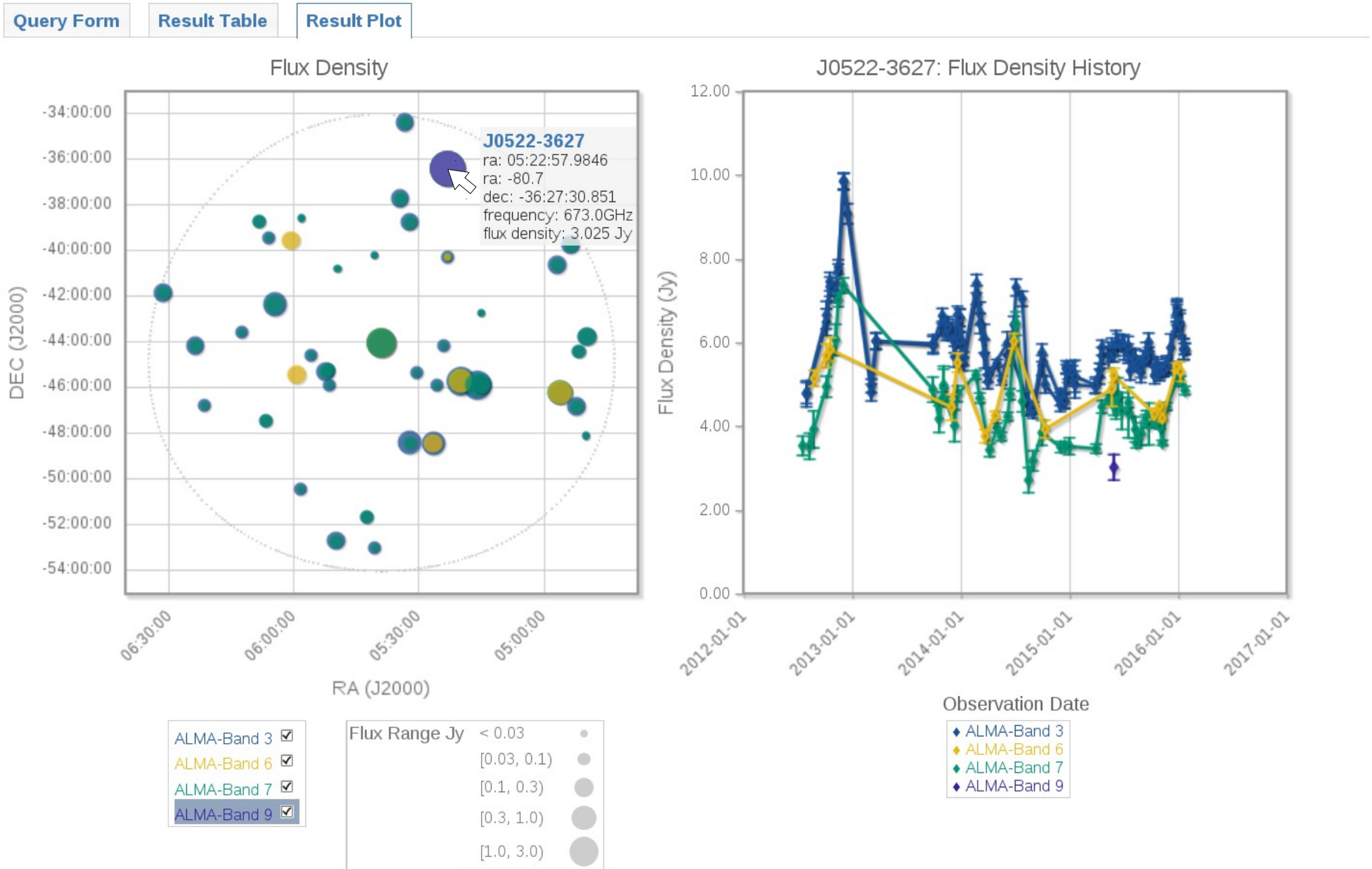




It is possible to download (cvs format)
only the most recent and/or the data from all the epochs available for selected calibrators or in regions.

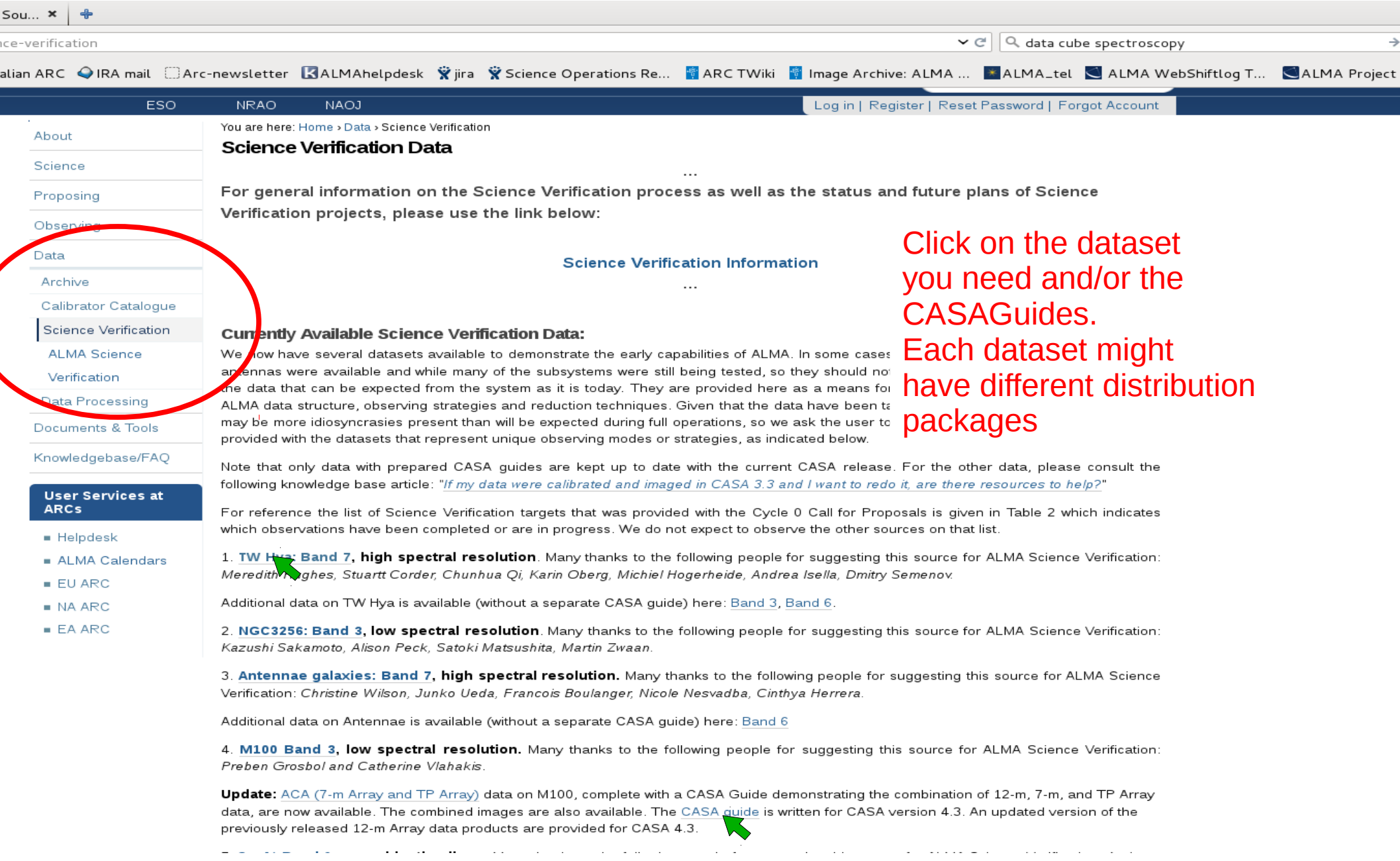
ALMA Calibrators

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



ALMA Science Verification Data

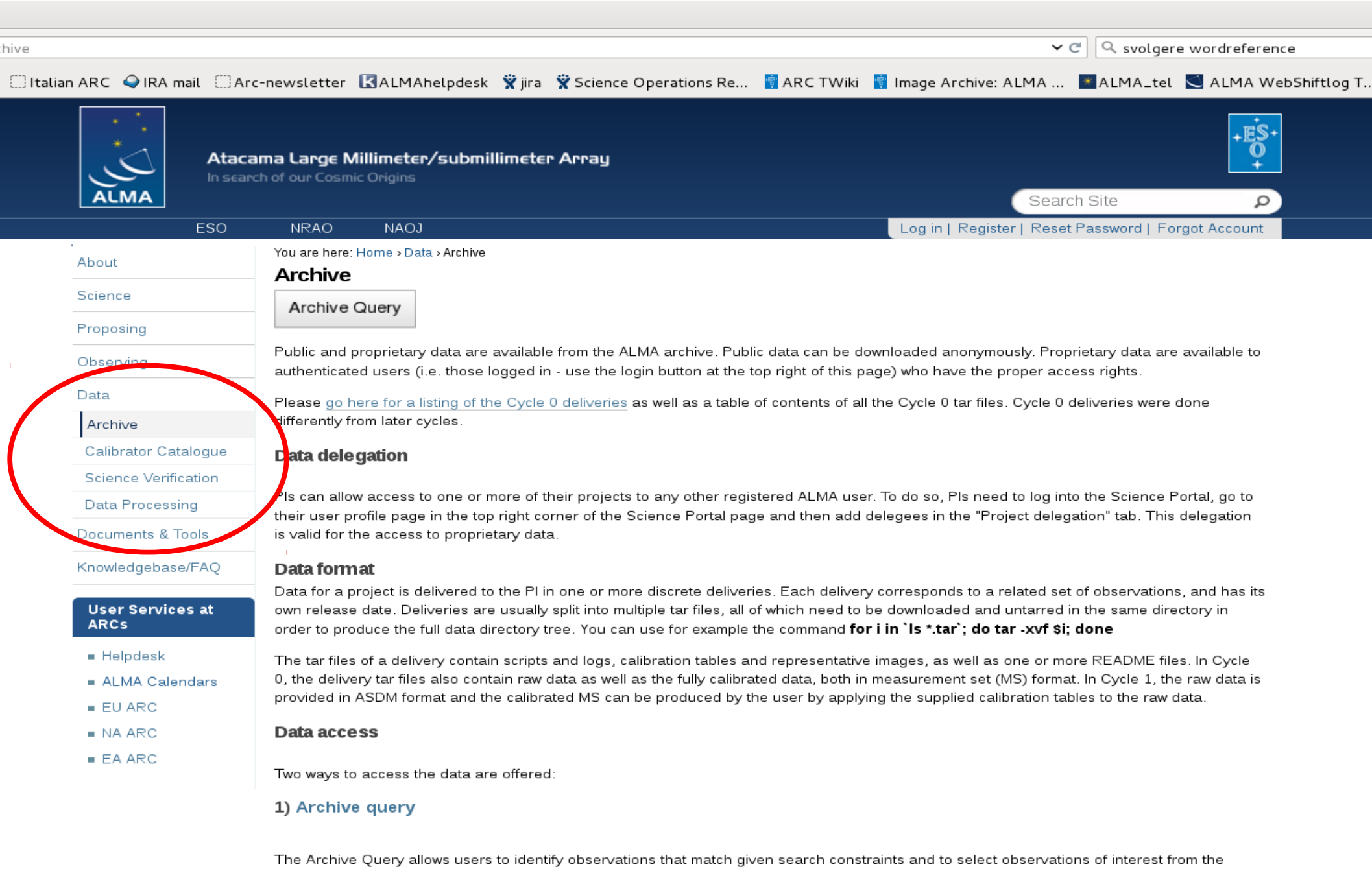
https://almascience.eso.org/alma-data/science-verification



Click on the dataset you need and/or the CASAGuides.
Each dataset might have different distribution packages

Tha Science ALMA Archive

https://almascience.eso.org/alma-data/archive



Definitions: SG, SB and EB

At proposal stage the PI makes some choices and requests needed to reach his purposes.
The PI splits the project in

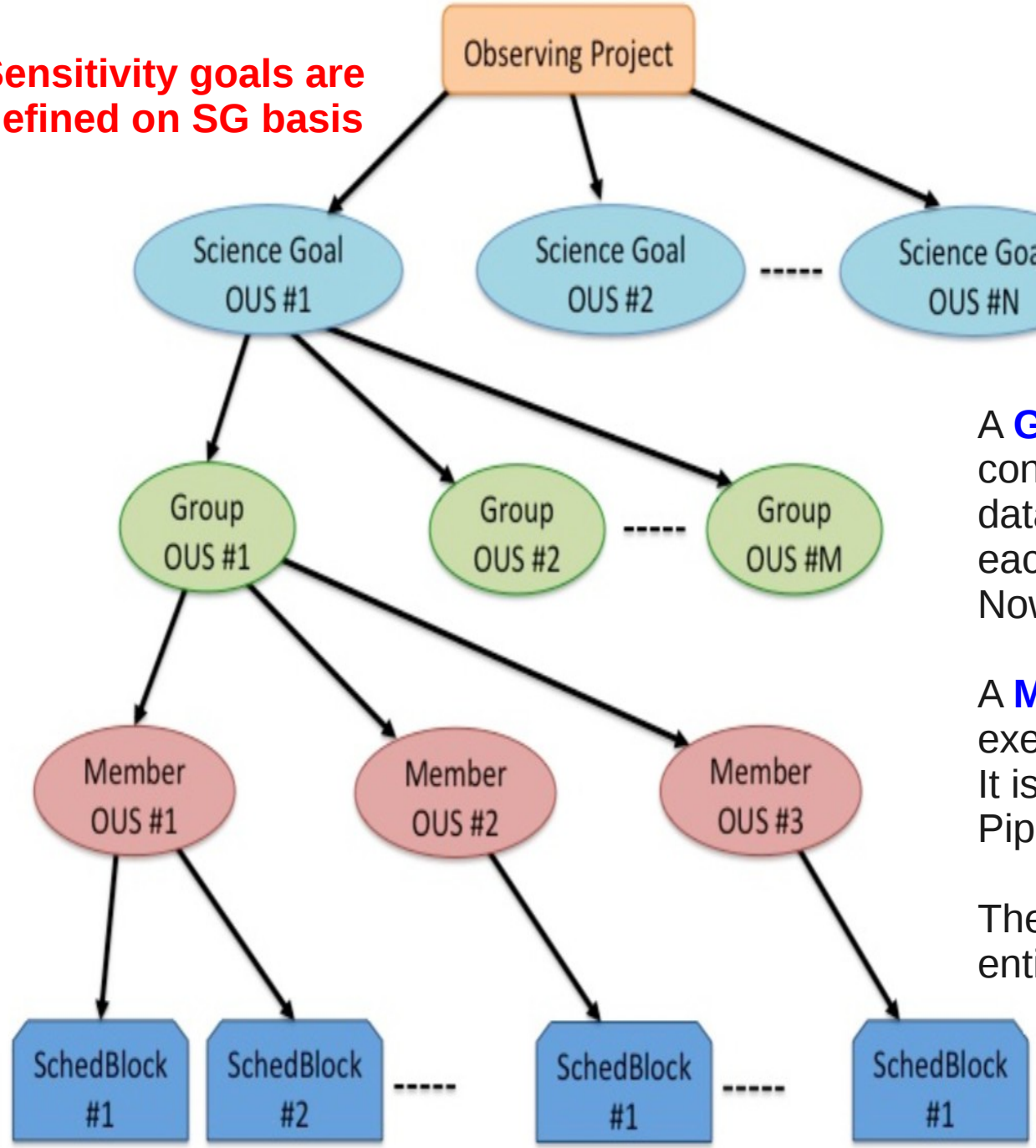
Science Goals **Minimum proposed observational unit including targets**
in the same sky region that roughly share the same calibration
aimed at reaching a requested sensitivity
in a given angular (resolution and LAS) and spectral setup
e.g. Different bands on the same target are in different SG
Different 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 observative 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!!!

Data structure

Sensitivity goals are defined on SG basis



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

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**

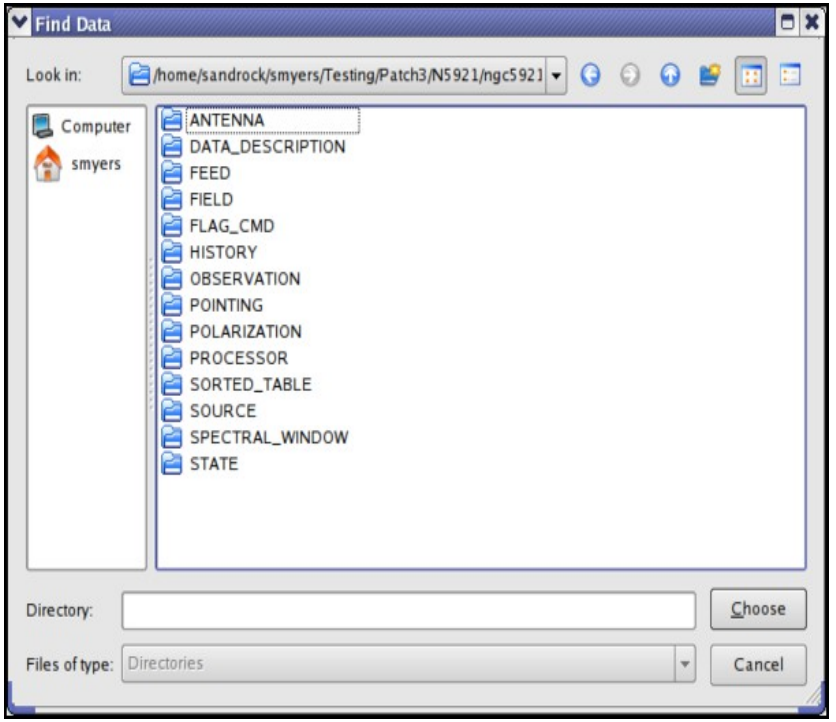
Each has an unique hexadecimal name
(eg uid://A002/X2fed6/X3f).
Each contains the meta-data (headers,
descriptions of the observation setup, etc),
and the binary data (the raw data)

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.

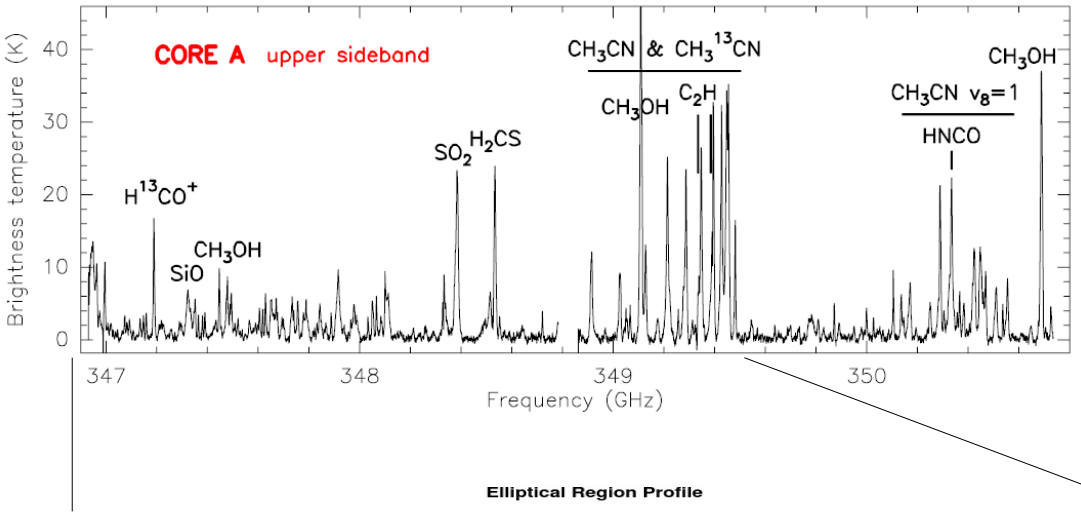
As a consequence of the process only data that passed QA2 (at least in part) are in the archive

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

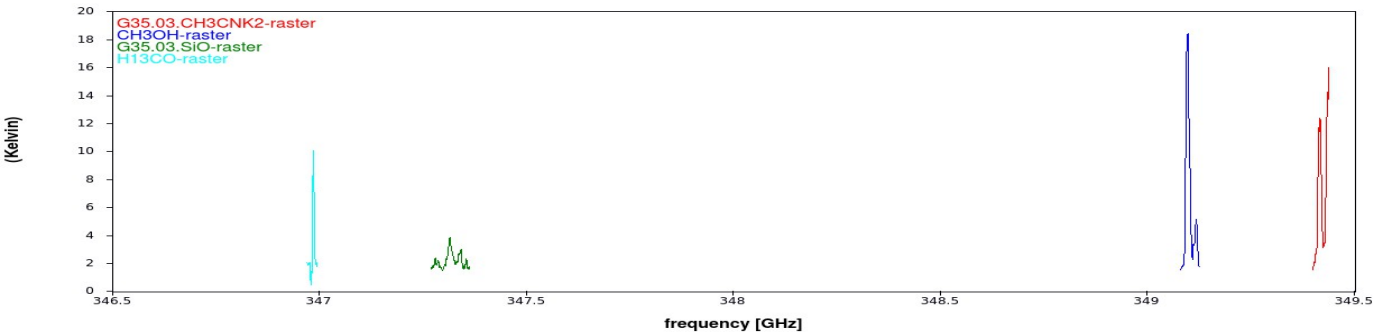
Images in the archive are provided as starting point on the way to obtain the final images and a valuable basis for archive researchers
(i.e. they are not considered science-ready!!!)

What is in the archive?

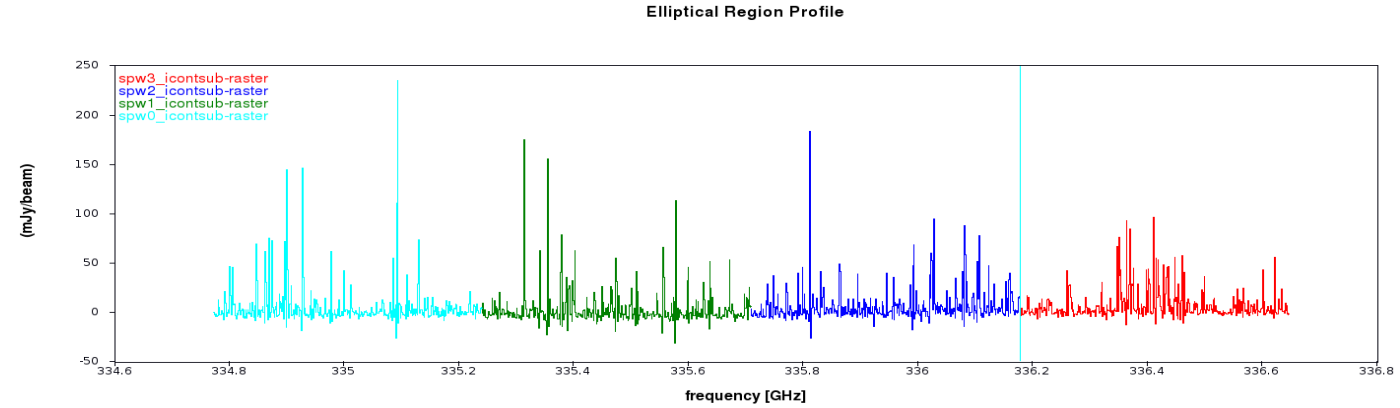


Images in the archive might cover only a fraction of the spectrum available

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:

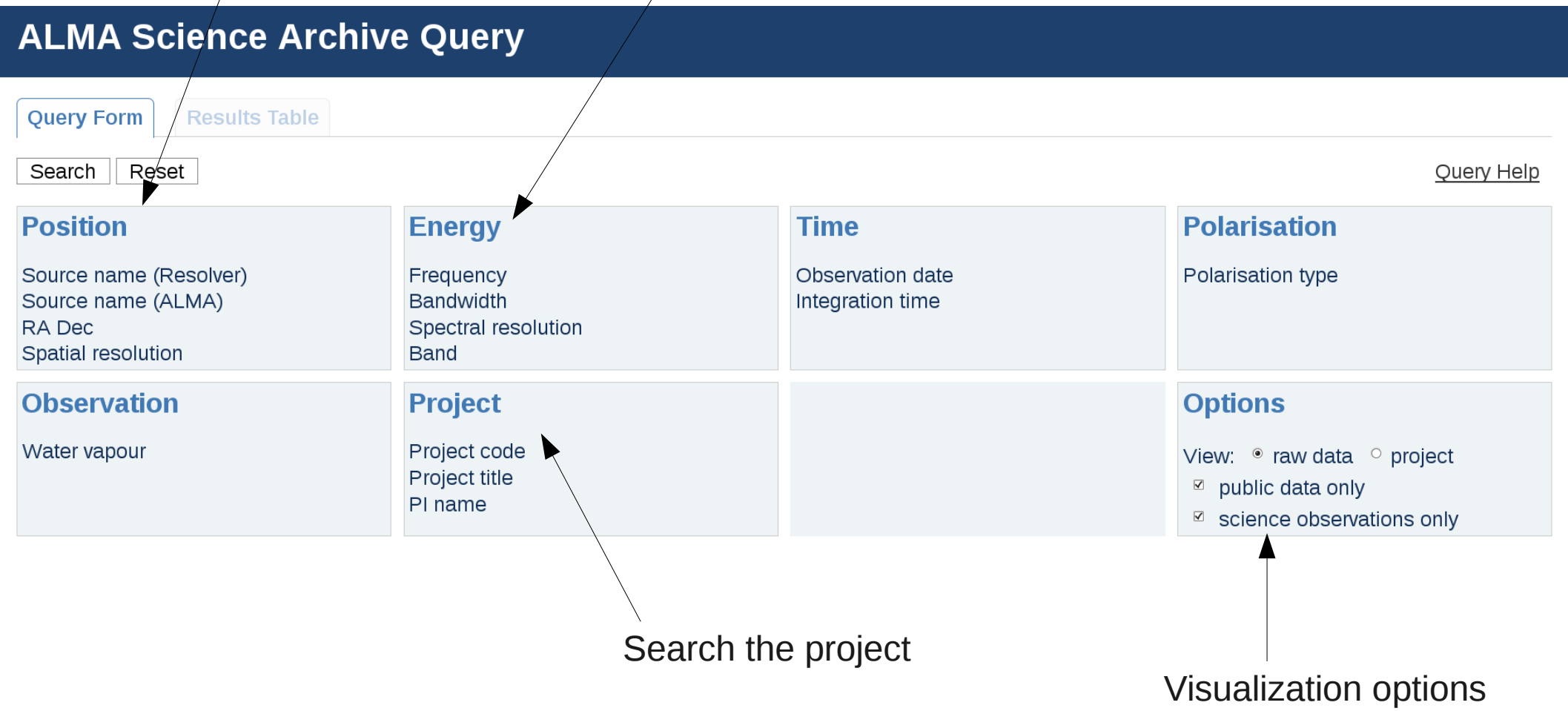
- Allow download of “ligth” (<1GB) products (QA2 images and READme) and/or “heavy” (>10GB) raw data
- Check in the downloaded README files and script the CASA version used

→ **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**

The ALMA archive: query

Search per name or position or within a radius

Search the spectral setup



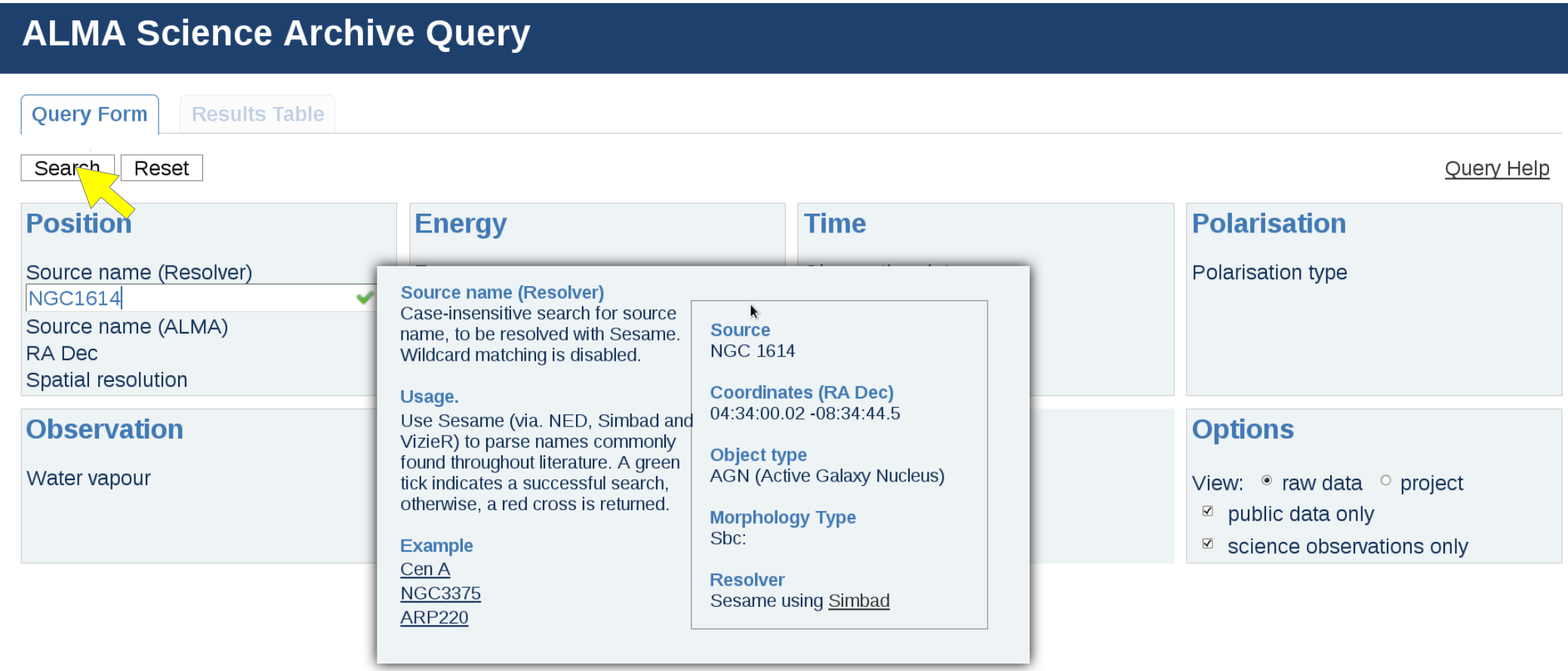
Search the project

Visualization options

The query will change to allow more criteria!!!

The ALMA archive: help

1) Search with the criteria you need and click Search



Contextual help for each tab

The ALMA archive: result table

2) Select the project/execution blocks you need and click "Submit the download request"

ALMA Science Archive Query

Query Form | Results Table

Submit download request

Results Bookmark | Export Table | Results Help

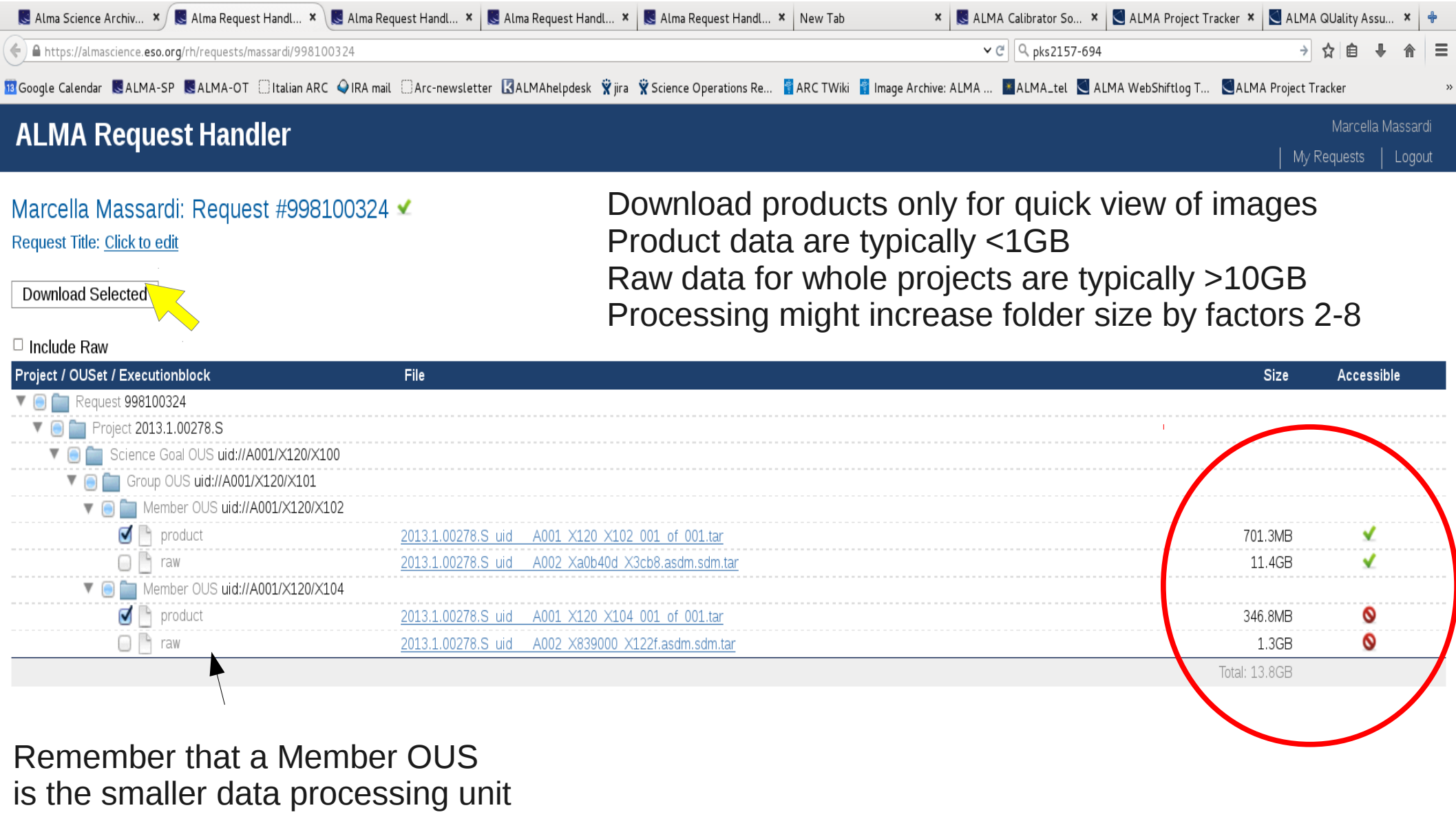
Showing 7 rows (7 before filtering). More columns

	Project code	Source name	RA	Dec	Band	Integration	*Release date*	Velocity resolution	Frequency support
Filter:								m/s	
☐	2011.0.00020.S	NGC 1614	04:34:00.03	-08:34:44.6	7	484.557	2013-01-12	834.09	344.15..357.85GHz
☐	2011.0.00020.S	NGC 1614	04:34:00.03	-08:34:44.6	7	382.854	2013-01-12	851.55	336.17..351.86GHz
☐	2011.0.00768.S	NGC1614	04:34:00.03	-08:34:44.6	7	463.612	2013-10-15	846.76	337.97..353.59GHz
☐	2011.0.00768.S	NGC1614	04:34:00.03	-08:34:44.6	7	464.391	2013-10-15	846.76	337.97..353.59GHz
☐	2011.0.00768.S	NGC1614	04:34:00.03	-08:34:44.6	7	463.991	2013-10-15	846.76	337.97..353.59GHz
☑	2011.0.00182.S	NGC 1614	04:34:00.03	-08:34:45.2	9	697.859	2013-12-21	13784.20	675.83..683.30GHz
☑	2011.0.00182.S	NGC 1614	04:34:00.03	-08:34:45.2	9	702.437	2013-12-21	13784.20	675.83..683.30GHz

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 returns many lines.
Columns values are only indicative. Data structure can be more complex than what shown.

The ALMA archive: download manager

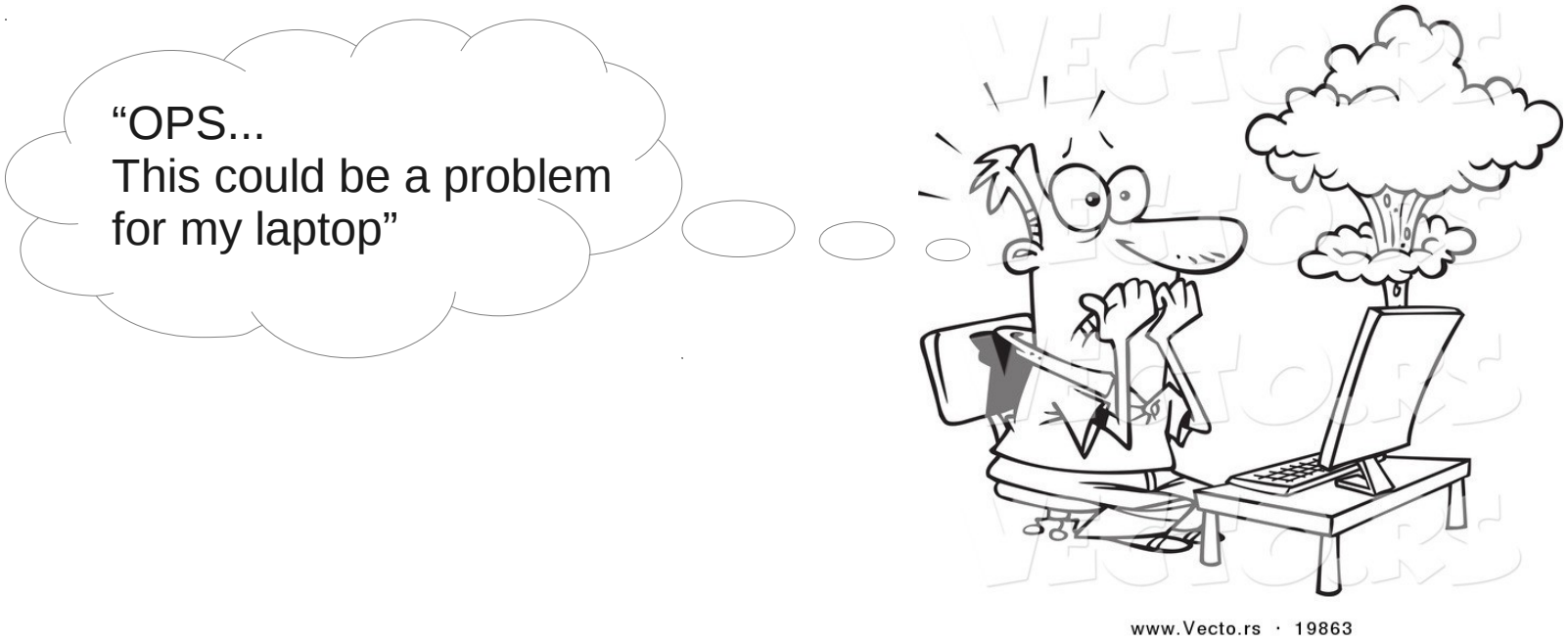
3) Select the data you want



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

Remember that a Member OUS
is the smaller data processing unit

10 GB or more?????

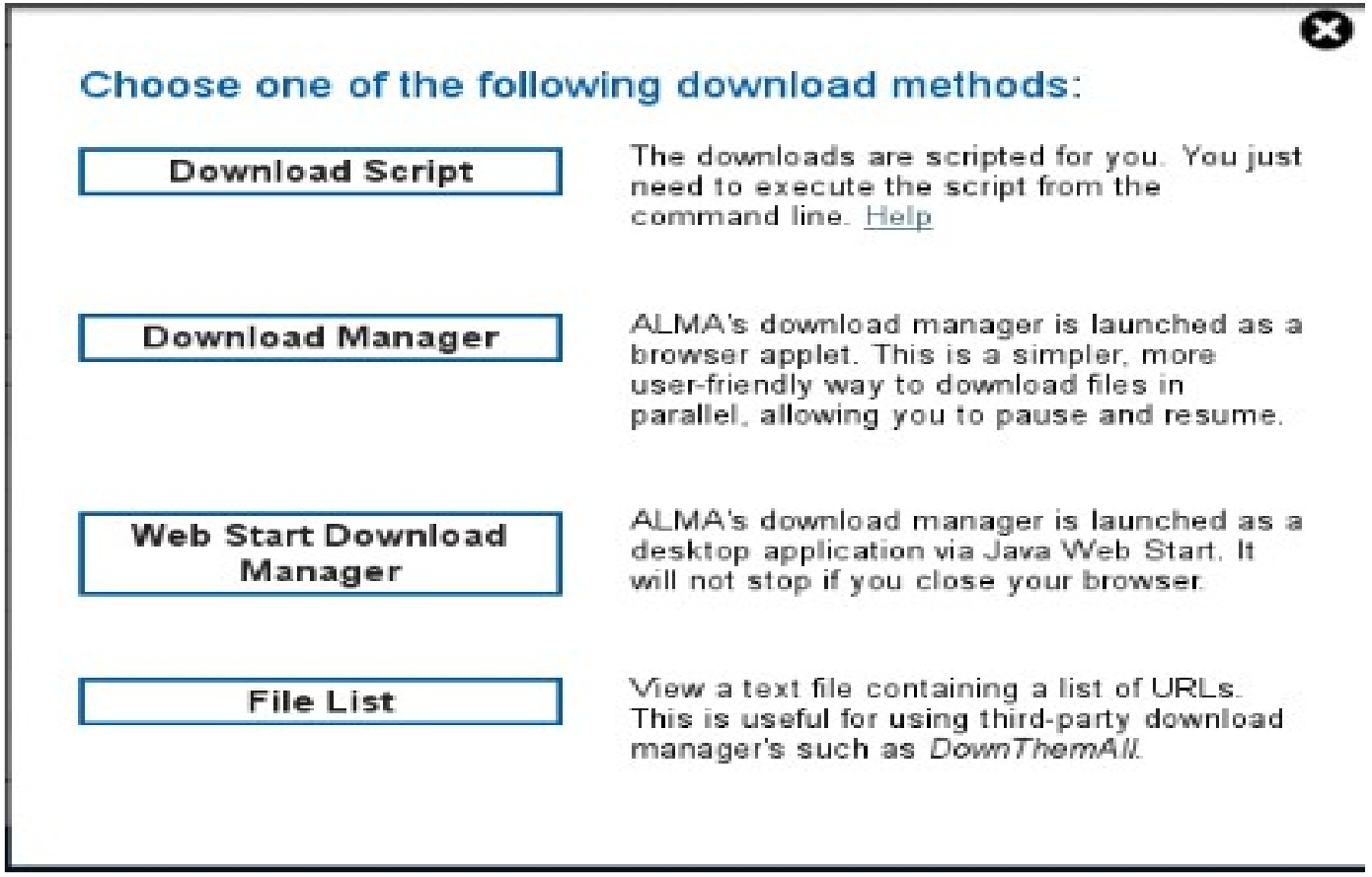


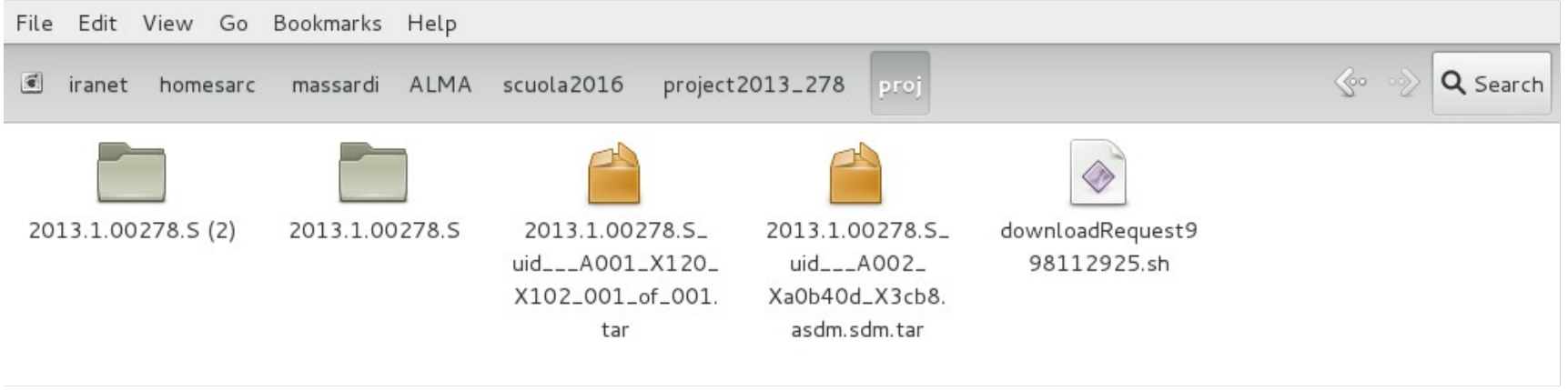
Don't struggle on it!!!
You can ask an account on our ARC cluster to deal with ALMA data!

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

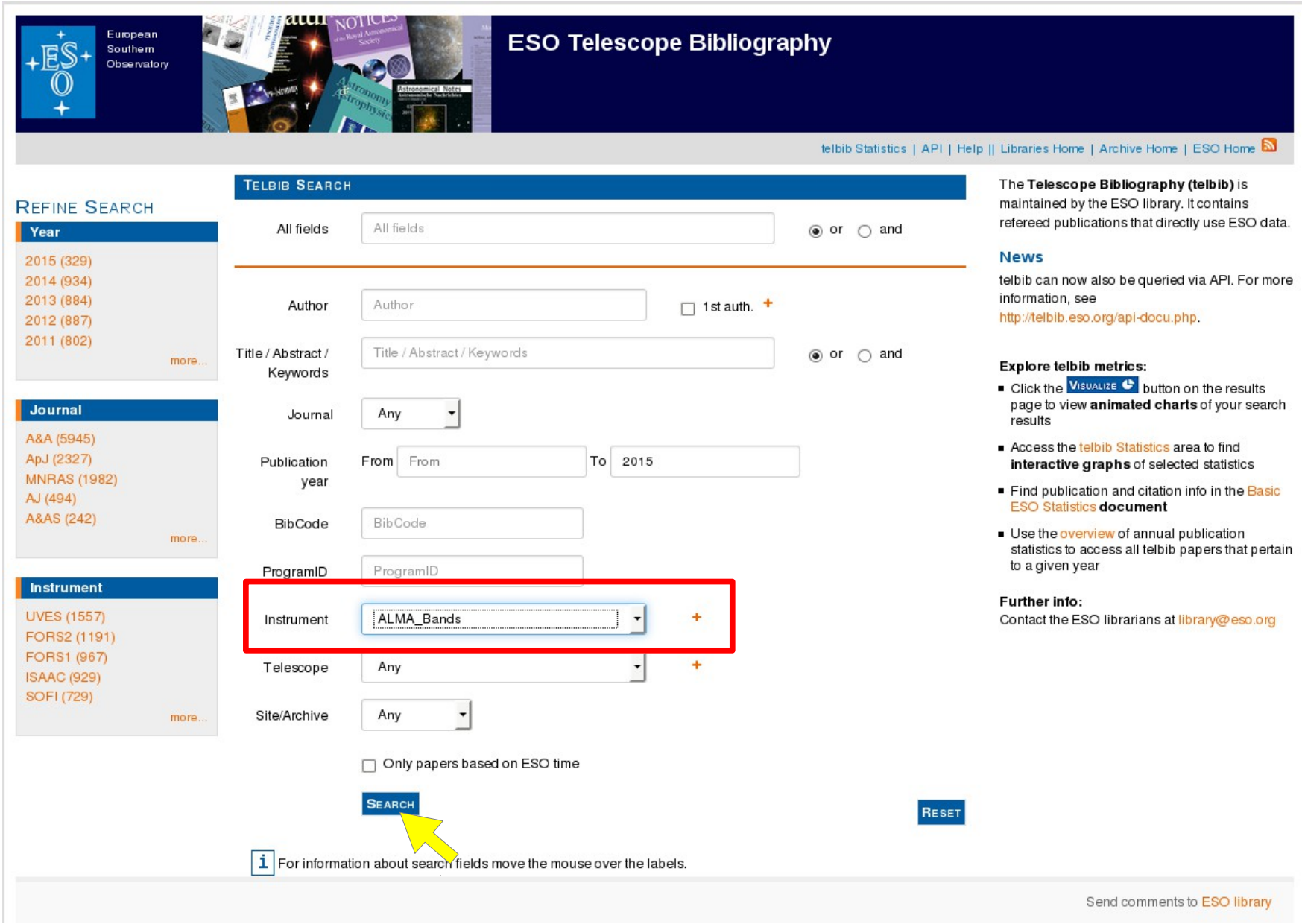
4) Choose the download method





The ESO telbib

http://telbib.eso.org/



The ESO telbib

REFINE SEARCH

Year

2015 (40)
2014 (97)
2013 (65)
2012 (20)

Journal

ApJ (121)
A&A (54)
MNRAS (16)
Nature (11)
PASJ (6)
more...

Instrument

ALMA_Bands (222)
LABOCA (14)
XSHOOTER (6)
FORS2 (5)
SHFI (5)
more...

NEW SEARCH | EDIT SEARCH | VISUALIZE | EXPORT

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

« Previous Next »

YEAR ▼	AUTHOR	TITLE	INSTRUMENTS	ACCESS TO DATA	FULLTEXT ADS
2015	Sakai, Yusuke et al.	An ALMA Imaging Study of Methyl Formate (HCOOCH3) in Torsionally Excited States toward Orion KL	***ALMA_Bands***	2011.0.00009.SV	2015ApJ...803...97S
2015	Brouillet, N. et al.	Antifreeze in the hot core of Orion. First detection of ethylene glycol in Orion-KL	***ALMA_Bands***	2011.0.00009.SV	2015A&A...576A.129B
2015	Saito, Toshiki et al.	ALMA Multi-line Observations of the IR-bright Merger VV 114	***ALMA_Bands***	2011.0.00467.S	2015ApJ...803...60S
2015	Olofsson, H. et al.	ALMA view of the circumstellar environment of the post-common-envelope-evolution binary system HD 101584	***ALMA_Bands***	2012.1.00248.S	2015A&A...576L..15O
2015	Sakai, Takeshi et al.	ALMA Observations of the IRDC Clump G34.43+00.24 MM3: DNC/HNC Ratio	***ALMA_Bands***	2011.0.00656.S	2015ApJ...803...70S
2015	Gullberg, B. et al.	The nature of the [C II] emission in dusty star-forming galaxies from the SPT survey	***ALMA_Bands***	2011.0.00957.S, 2011.0.00958.S, 2012.1.00844.S	2015MNRAS.449.2883G
2015	Rathborne, J. M. et al.	A Cluster in the Making: ALMA Reveals the Initial Conditions for High-mass Cluster Formation	***ALMA_Bands***	2011.0.00217.S	2015ApJ...802..125R

Direct link to the ALMA Archive

ASTROquery.alma
(https://astroquery.readthedocs.org/en/latest/alma/alma.html)

- 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...)

What is in the packages?

When untarred, the Product Package standard directory structure contains

Untarred products only

```
|-- 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
```

Untarred ASDM (raw data)

```
| | | | |-- raw/ .....for calibration move it in the products folder at the
                                                    right level  (follow the README)
```

Untarred ASDM (Raw data)

Untarred products

What to do after download?

```
[massardi@arcbl02 member.uid___A001_X120_X102]$ cd script/
[massardi@arcbl02 script]$ casapy-setup 42.2.30986-pipe-1-64b
[massardi@arcbl02 script]$ casapy --pipeline
...

CASA <2>: execfile('scriptForPI.py')
```

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

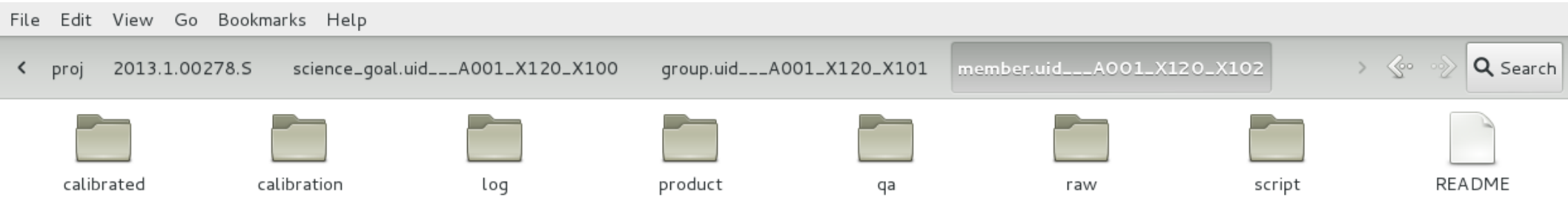
Down loaded data

```
|-- 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
```

New folder produced by script



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!

Acknoweledgement 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)