# THE ALMA SCIENCE ARCHIVE

M. MASSARDI & K. RYGL

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 similar objects (e.g. statistics of populations, …)

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- Extracting unpublished information from existing data (e.g. finding additional spectral lines, targets in the same region/time of other observations,)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

The ALMA archive mirrors
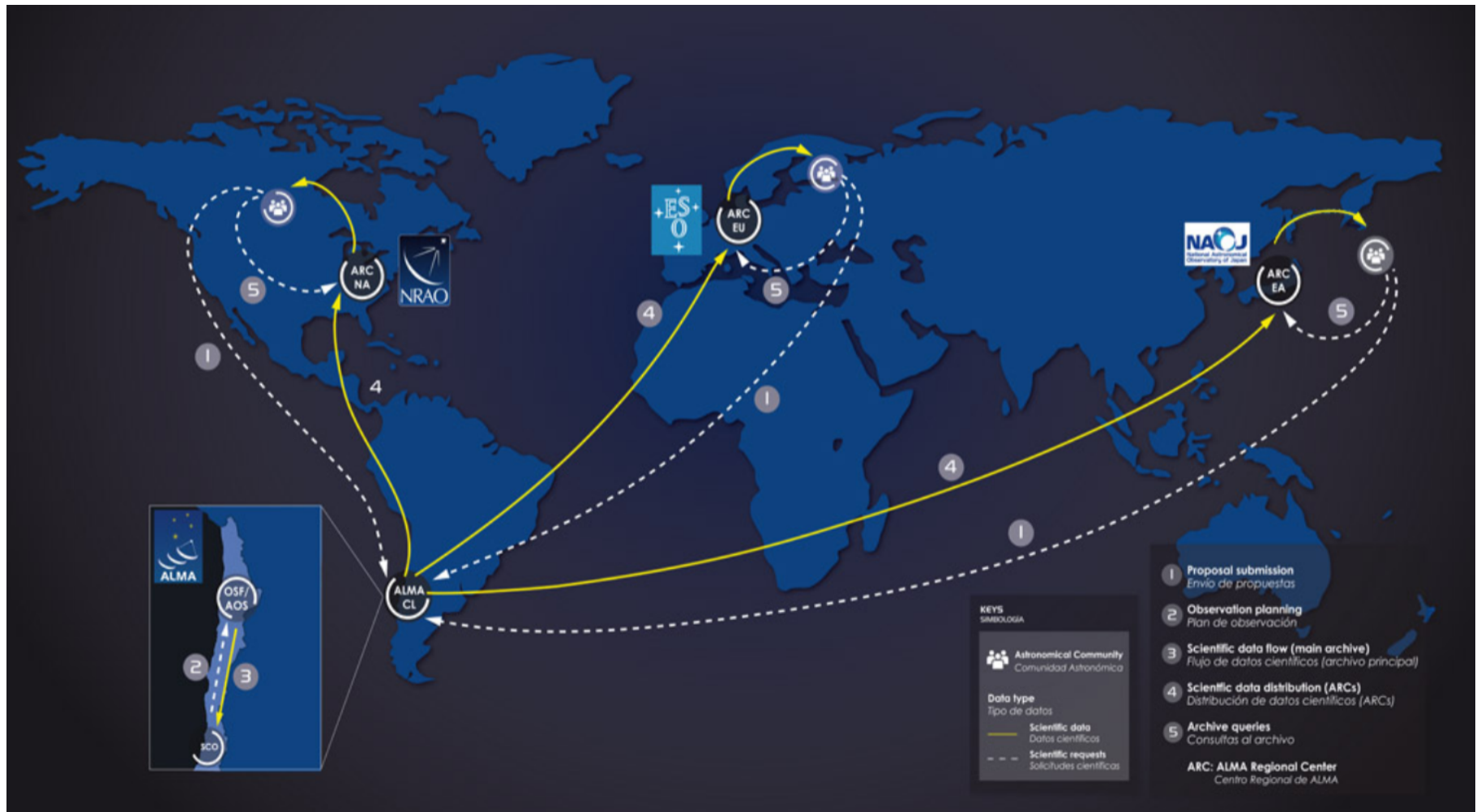
Data is collected, reduced and archived.

Each ARC hosts an archive mirror.

What is in the archive?

For each project the main deliverables are

Raw Data, Calibration Scripts and tables, Images and Quality summary

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

Some Imaging Products are delivered too, as result of QA 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:
- continuum image combining all SPWs

The main purpose is to measure the rms and verify the achievement of PI requests.

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!!!)

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 science goals, 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 distributed at the JAO and the ARCs with the help of a semi-automatic CASA pipeline. **Results are archived and given to the PI. Only "QA2-passed" or "QA2-semipassed" data sets are archived.**

QA3: (optional) PIs may request rereduction, problem fixes, possibly reobservation

The ALMA Reduction procedures

Manual Calibration
- Always available
-Applied on old cycles and non standard mode observations (at the time of QA2)
-To be restored **requires specific versions of CASA (see the README file!)**
-An **editable script is made available**
-Generates calibrated data and a png-txt Summary

Manual Imaging
- Always available
-Applied on old cycles and non standard mode observations (at the time of QA2)
-An **editable script is made available** as hint
- Typically **only fraction of data are imaged** as it is time consuming for analysts

Calibration Pipeline
-Available since Cycle 3
-Applied on standard mode observations ("standard" at the time of QA2)
-To be restored **requires specific versions of CASA (with pipeline!)**
-Can be edited by expert people (usually manual processing is faster!)
-Generates calibrated data and weblog Summary

Imaging Pipeline
-Available since mid Cycle 4 (now more than 75% of the QA2)
-Applied on pipeline calibrated data
-To run requires specific versions of CASA (with pipeline!)
-Can be edited by expert people (manual imaging is faster!)
-Generates full cubes at the native resolution
-**No additional script is made available** as hint for manual reimaging

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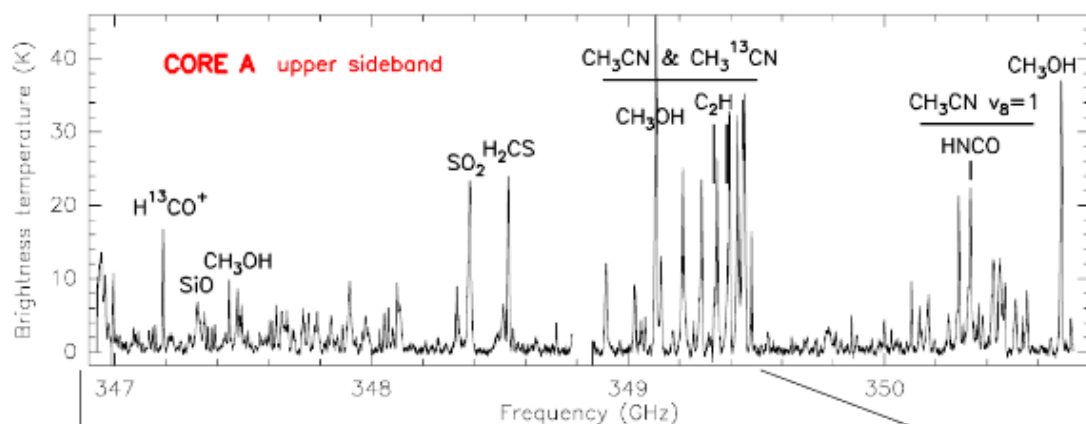
Images in the archive might cover only a fraction of the spectrum available

Spectrum observed
(and available in the raw data)
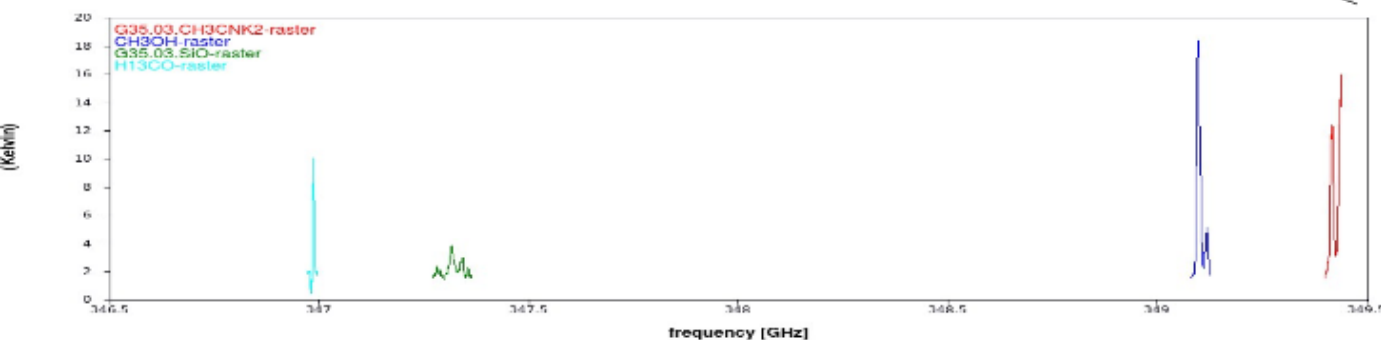
Imaged data for QA2
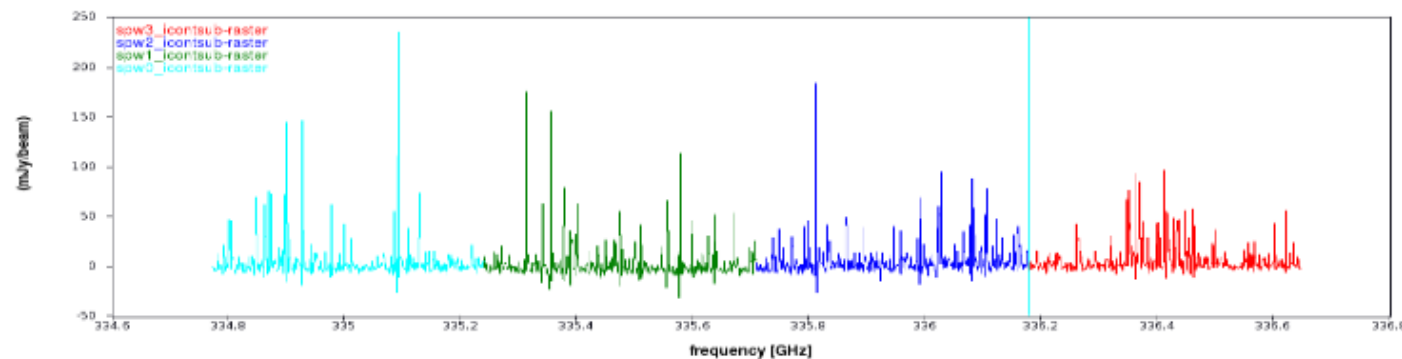
Different data and PI requests
on different sources
generate different products
In the archived images
but raw data contain the full
spectral windows

Data structure
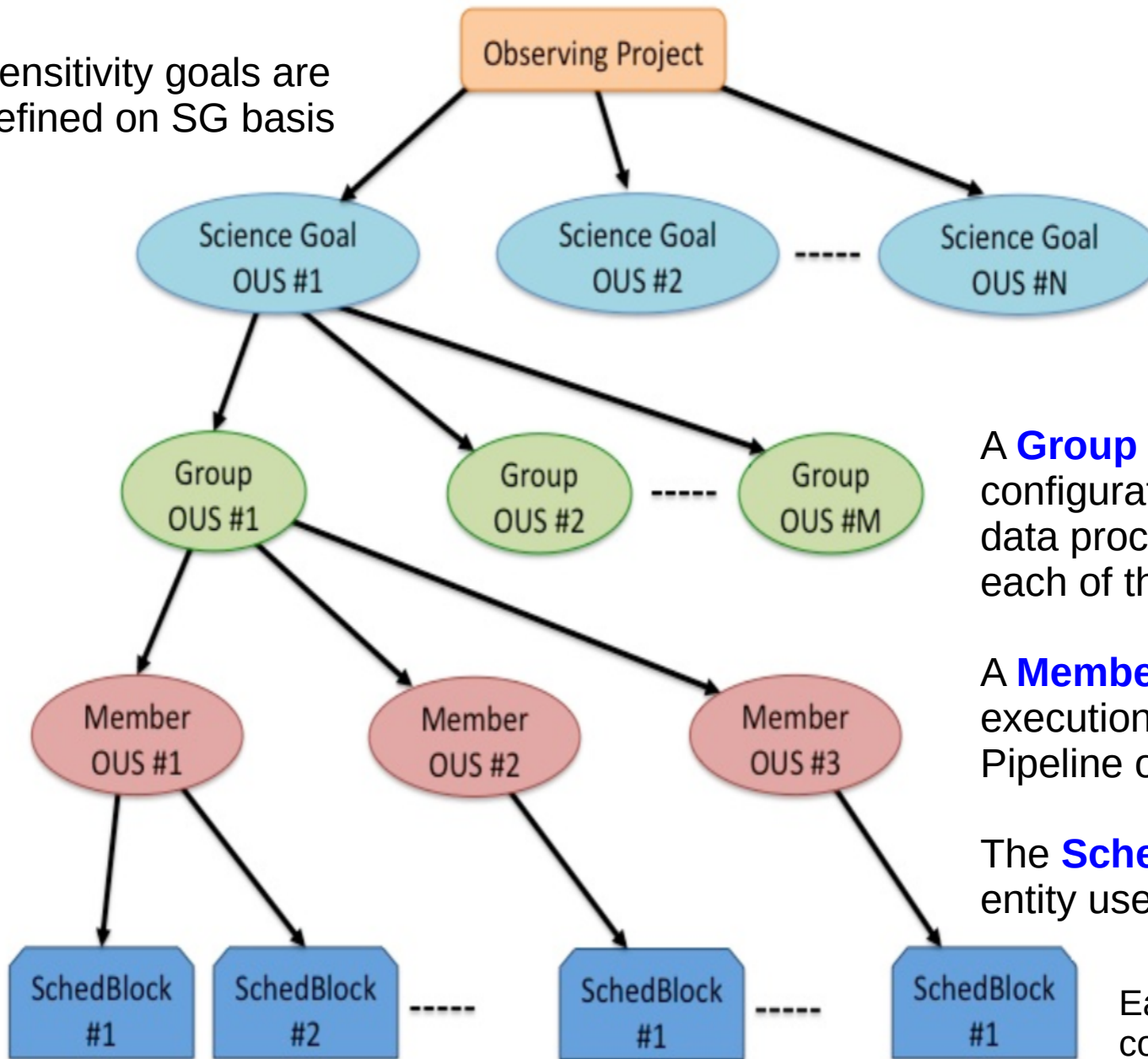
Sensitivity goals are defined on SG basis

Science goal:
Group of sources in the same sky region that share the same spectral setup

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.

A **Member** can contain multiple executions of a Scheduling Block. Pipeline operates at this level.

The **Scheduling Block** is the smallest entity used for observing

Each repetition of a Scheduling Block constitutes an **Execution Block**

Observing Project

Science Goal
OUS #1

Science Goal
OUS #2

Science Goal
OUS #N

Group
OUS #1

Group
OUS #2

Group
OUS #M

Member
OUS #1

Member
OUS #2

Member
OUS #3

SchedBlock
#1

SchedBlock
#2

SchedBlock
#1

SchedBlock
#1

The ALMA Science Archive

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

The ALMA archive: query

Search target name or position within a radius

Search the spectral setup

Visualization options for the result table

Search a project

Search data associated to papers

The query display will change after the ALMA deadline!!!

The ALMA archive: help

1) Search with the criteria you need and click Search

Contextual help for each tab

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 MOUS you need and click "Submit the download request"

ALMA Science Archive

Query Form | Results Table

Close Viewer   Results Bookmark   Export Table   Results Help

Submit download request

J2000   13 25 44.059 -29 50 1.24

FoV: 2.24'

ALADIN

More columns   Showing 4 of 4 rows.   Seelect the columns to show

	Project code	Source name	RA	Dec	Band	▲Release date	Velocity resolution	Frequency support	Pub	Angular res
Filter:			H:M:S ⌄	D:M:S ⌄			m/s ⌄			
☐	2013.1.00243.S	NGC_5135	13:25:44.06	-29:50:01.2	6	2016-01-27	1206.86	226.47..242.55GHz	2	1.16
☐	2013.1.00243.S	NGC_5135	13:25:44.06	-29:50:01.2	6	2016-08-22	1207.26	226.47..242.55GHz	2	0.23
☐	2013.1.00524.S	NGC5135	13:25:43.99	-29:50:01.1	9	2016-10-01	13682.87	677.29..684.77GHz	7	0.16
☐	2018.1.01344.S	NGC5135	13:25:44.06	-29:50:01.2	6	2020-01-03	8842.47	245.67..264.86GHz	0	0.26

Restricted
access

Click to see abstract details

Notice multiple MOUS
on the same proj (different resolution)

The ALMA archive: download manager

3) Select the data you want: which MOUS products, raw data

Zipped raw data sizes for whole projects are typically >10GB

The ALMA archive: download manager

4) Choose the download method

Choose one of the following download methods:

Download Script — The download scripts are available for Linux and MacOS only.

Java 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. You must have Java installed on your computer.

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

Request 2146910448196

total size of files: 1.2GB

This page does not use single-sign on. It uses basic authentication and is intended to be scriptable.

PLEASE do not attempt to chop a single download into pieces to make it faster. This places a significant load on our servers and may result in your downloads being throttled. For example, in the Firefox plugin *DownThemAll*, make sure that the property "Max. Number of Segments Per Downloads" is set to 1. Likewise, it's easy to download more than 5 files in parallel. Please don't.

- https://almascience.eso.org/dataPortal/requests/anonymous/2146910448196/ALMA/2013.1.00243.S_uid___A001_X122_X48c_001_of_001.tar/2013.1.00243.S_uid___A001_X122_X48c_001_of_001.tar
- https://almascience.eso.org/dataPortal/requests/anonymous/2146910448196/ALMA/2013.1.00243.S_uid___A001_X122_X48e_001_of_001.tar/2013.1.00243.S_uid___A001_X122_X48e_001_of_001.tar

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)
```

**Do not change the folder structure and names if you want to run the scripts!
ALWAYS READ THE README FILE!!!!!!**

Look to QA summary for data quality

Look to the weblog or to the png summary plots and details in the "qa" folder to verify the quality of the data.

For more info come to the ARC or have a look to

http://www.alma.inaf.it/index.php/ALMA_Data_Handling_Workshop

Look to products with CASAviewer

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
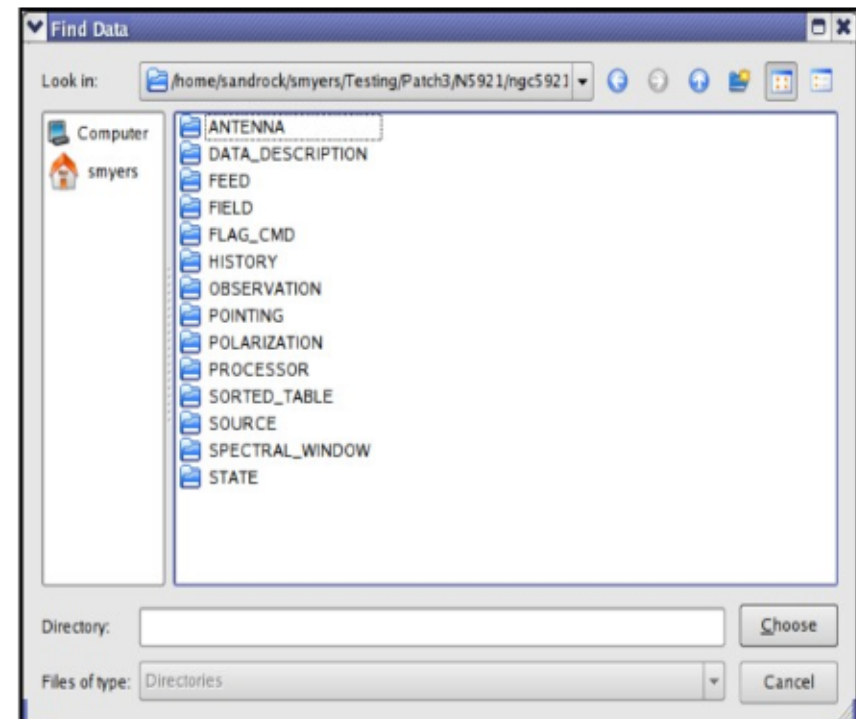
What to do after download?

```
[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
    (In case of pipeline imaging there is no Script_for_Imaging)
4) Edit the scripts where needed according to your purposes

Sabatini et al. (2018)
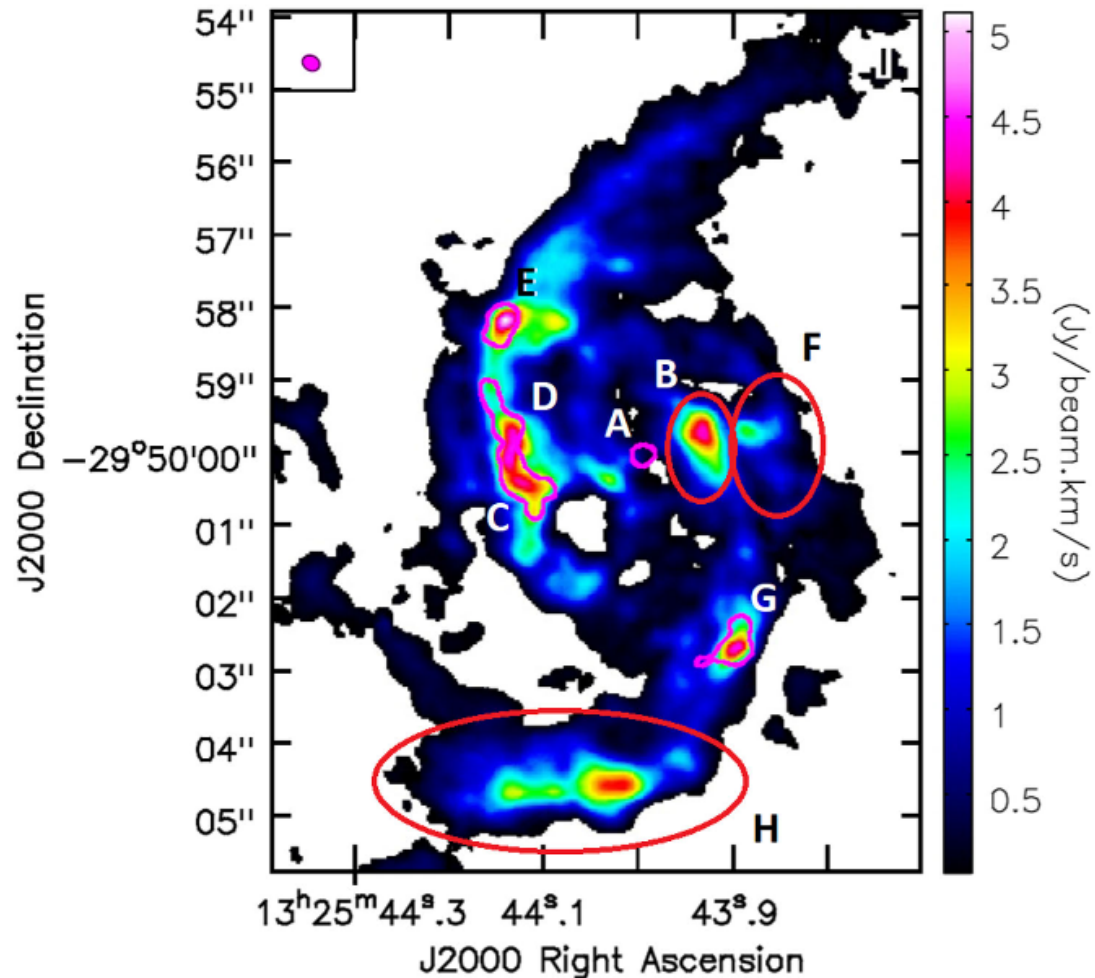
!!!Disclaimer!!!

Images are not supposed to be science ready in the ASA (even if sometimes they are pretty good).

No combination is done for multiple configurations.

Consider to re-image the archive products for science purposes.

Use products only as indication of data content.

THE ARI-L DEVELOPMENT PROJECT
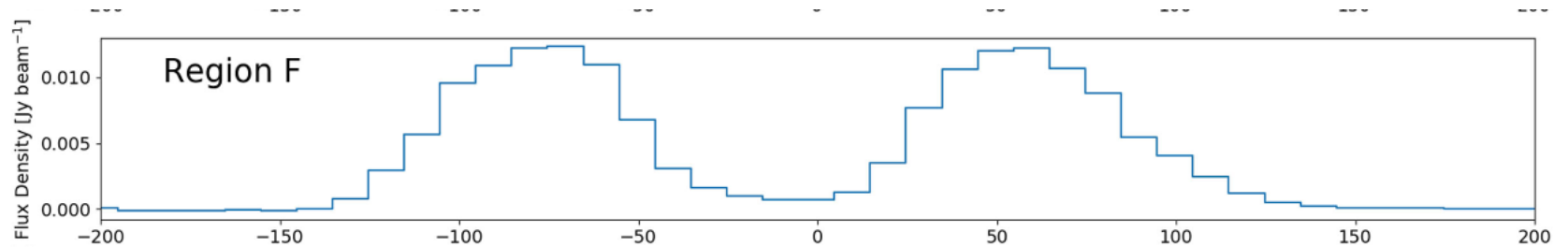
**The "Additional Representative Images for Legacy (ARI-L) in the ASA" project will produce and ingest into the ASA a set of additional image products representative of the whole data content for more than 70% of the observing projects in cycles 2-4 that can be processed through the ALMA Imaging Pipeline, to complement the QA2-generated images.
In addition calibrated MS of the processed dataset will be released too.**

The ARI-L images will:

- constitute a complete set of imaging products highly relevant for all science-cases and enhance the possibilities of exploitation of archival data

- facilitate the archive access and the data usage for science purposes also to non-expert data-miners,

- provide a homogeneous view of all data to compare datasets and to do a more conscious download selection,

- allow to more profitably link the archive to several tools of visualization and analysis (e.g. VO, CARTA, ADMIT, KAFE, ...),

- allow to generate previews similar to those planned for the following cycles.

The ARI-L products will not modify the current deliverables, but will be included as "externally contributed products". The project will produce a cube and a continuum image for each source (including calibrators) for at least 70% of the 3476 MOUS in cycle 2-4.

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)

●

The ESO telbib

http://telbib.eso.org/

The ESO telbib