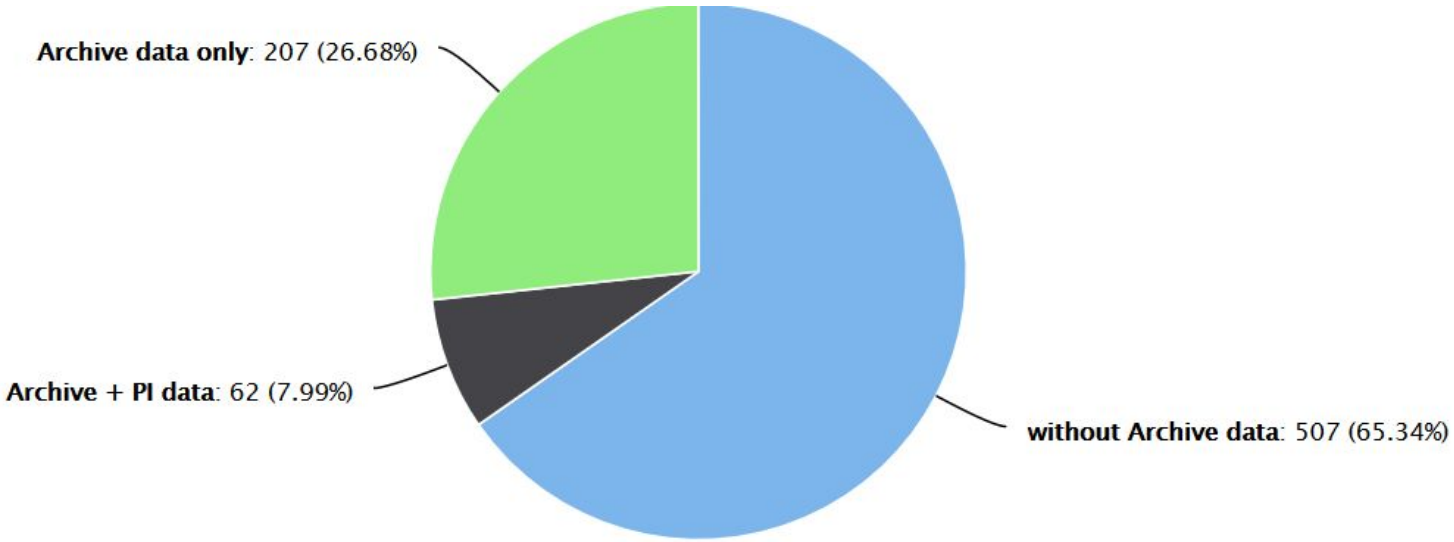
The Additional Representative Images for Legacy (ARI-L) Development Project

- **Exploiting Archives for Radio Astronomy in the SKA-era**

15th december 2020 - M. Massardi

THE ALMA Science Archive in the ALMA 2030 roadmap

...the ALMA Science Archive will become the primary source for an increasing number of publications. The ability to efficiently mine the archive contents is therefore vital for the community and ALMA's future...



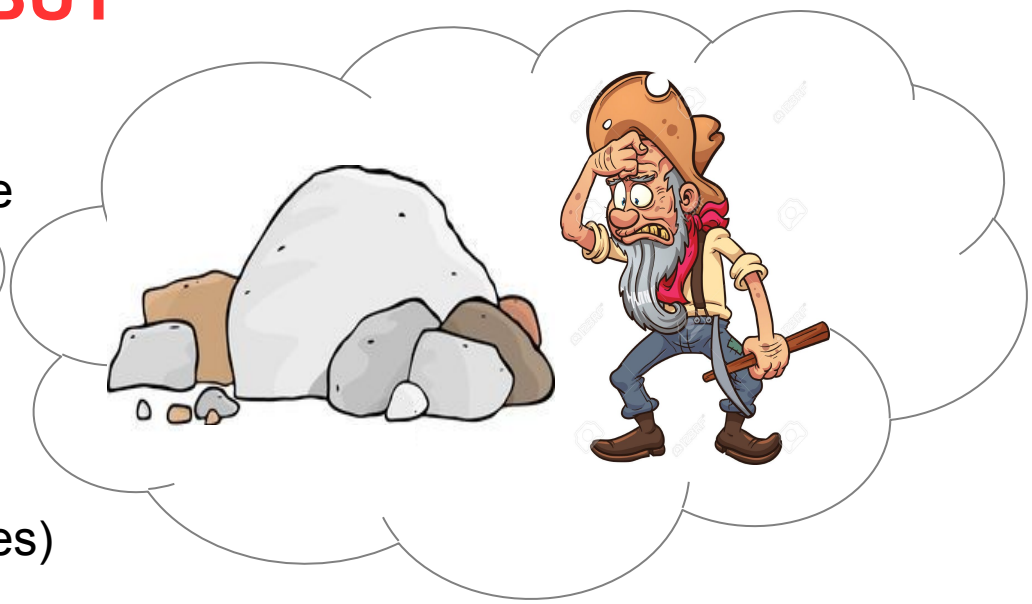
Mining efficiency affects the science exploitation efficiency

The ALMA Science Archive from a miner perspective

THE ASA IS A GOLD MINE

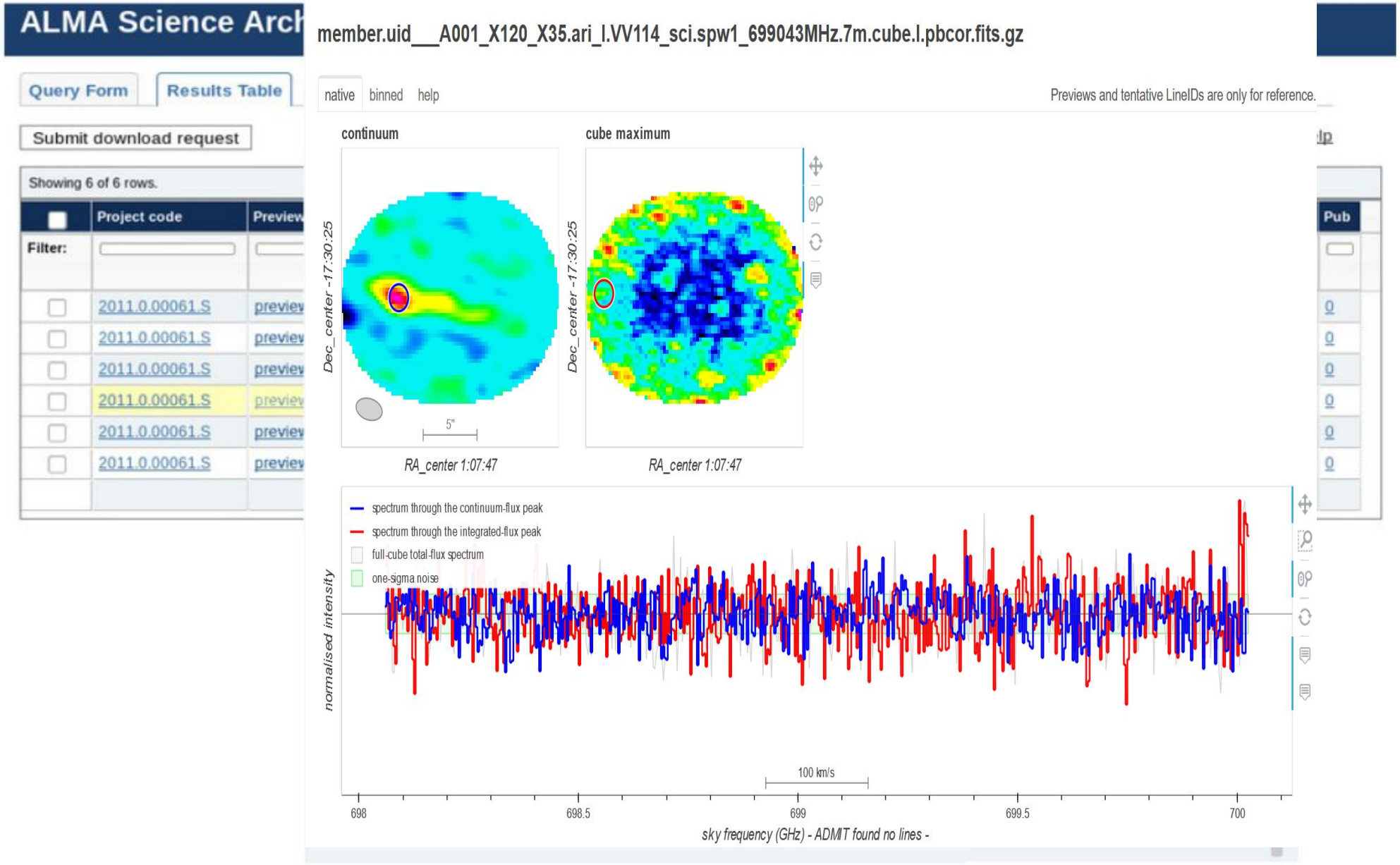
- 3348 projects observed and archived in 8 observing cycles
- data are already calibrated
- not all the possibilities have been exploited
- archive will keep growing

BUT



- dataset are big (>100MB in some cases)
- images available are made for quality assurance
  - = incomplete (<10 % in Cy 0-4, better in later cycles, thanks to pipeline)
  - = inhomogeneous across cycles
  - = not easily comparable
- to understand scripts users need interferometry induction/skill (luckily there are the ARC nodes)
- to run scripts CASA is needed (old versions …)

The ASA tools need images: ASA previews



The ASA tools need images: JVO

ALMA FITS Archive : Target Info

Target Name : NGC5135

▷ **Filter by Frequency**

☐ Show all the data including calibration (*.flux.fits, *.pb.fits, target=J####[+-]####), duplicated, a

Number of data per page : 20

Ordered by dataset_id (desc)

Total number : 4

* **WebQLv3 will end at the end of March in 2019.**

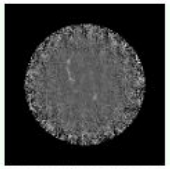
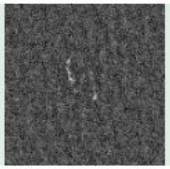
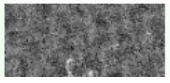
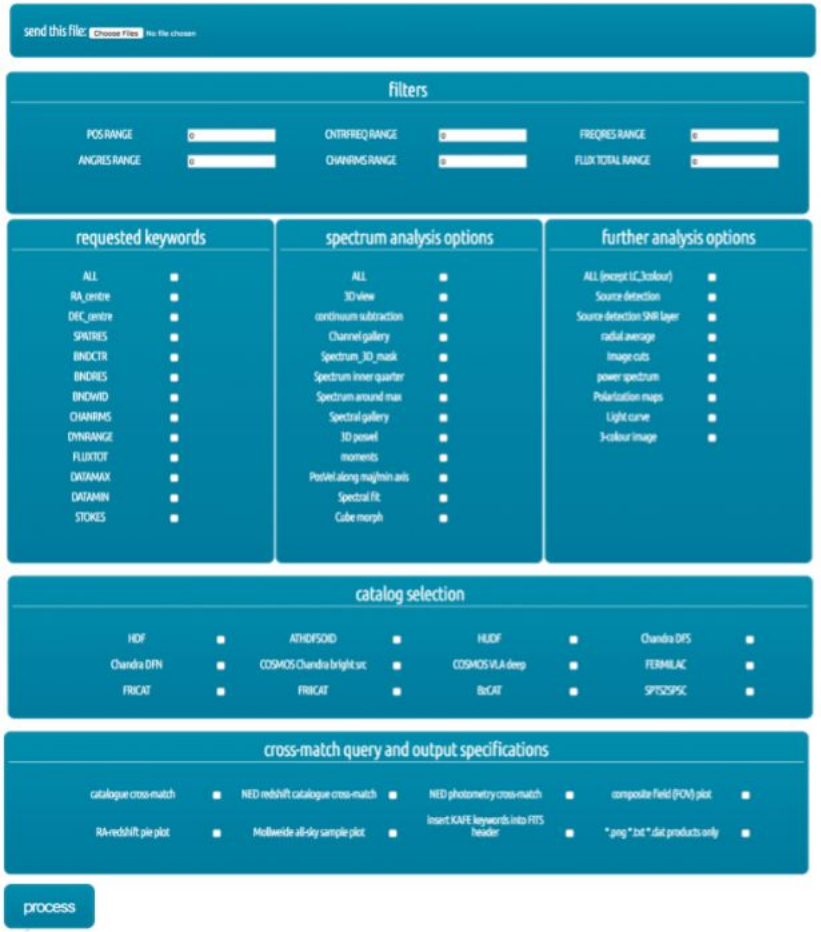
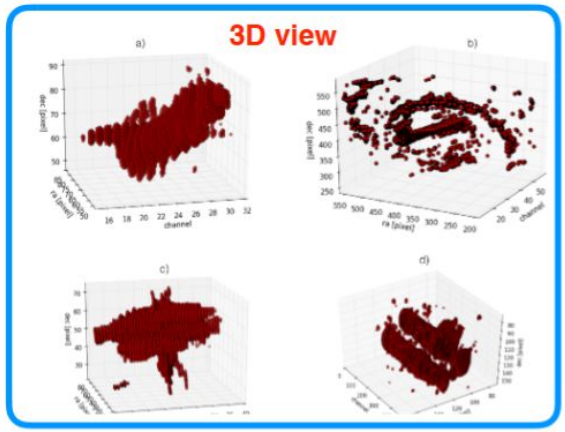
#	dataset id	project code	☐ all	Download all the checked data	image	spect
1	ALMA01081334	2013.1.00524.S	☐	Download WebQLv4		
2	ALMA01081333	2013.1.00524.S	☐	Download WebQLv4		
				Download		

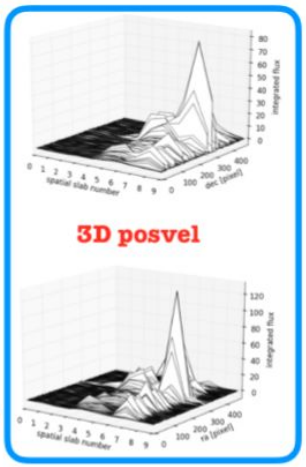
Image analysis tools

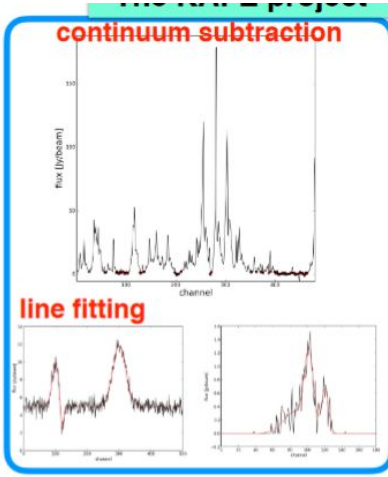


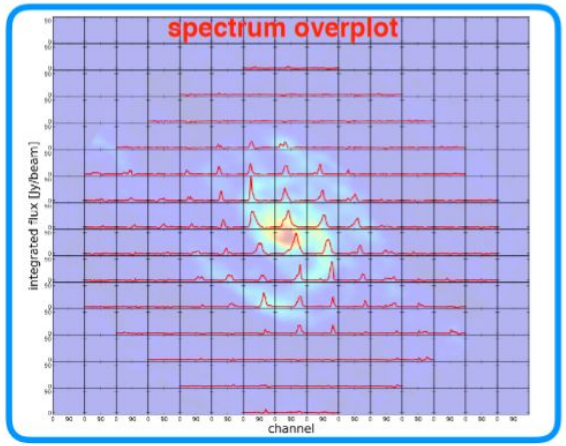
(Burkutean et al. 2018)

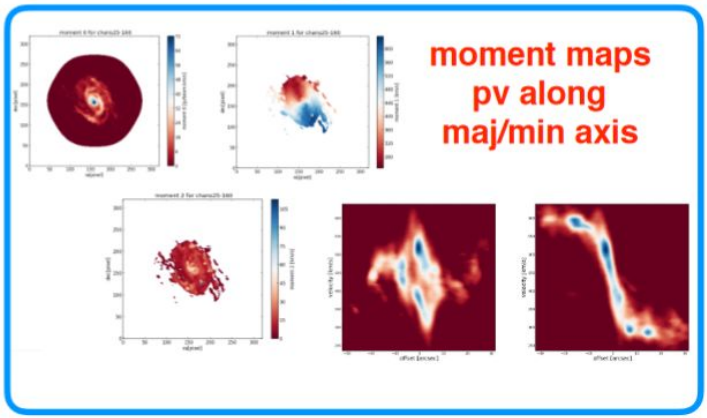
Please write to
kafe@ira.inaf.it
for access information
to the web interface.



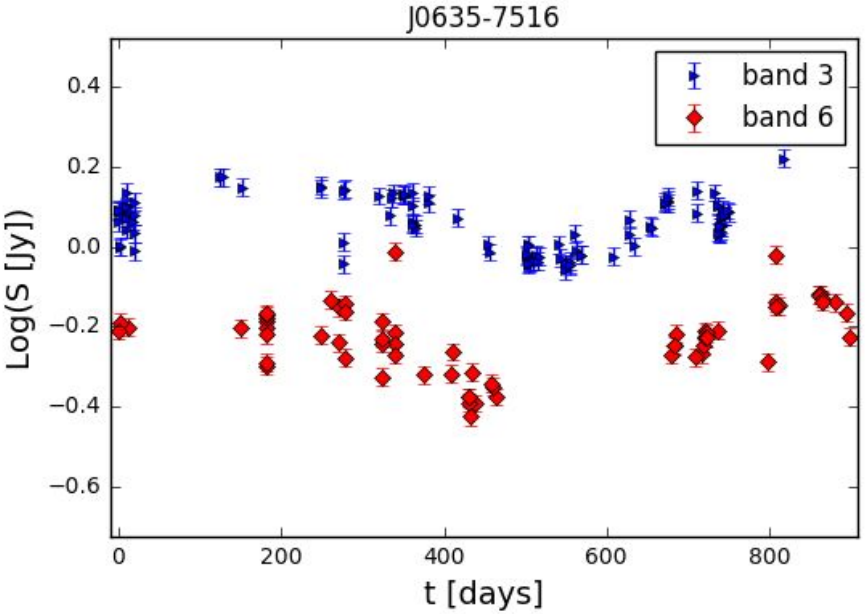


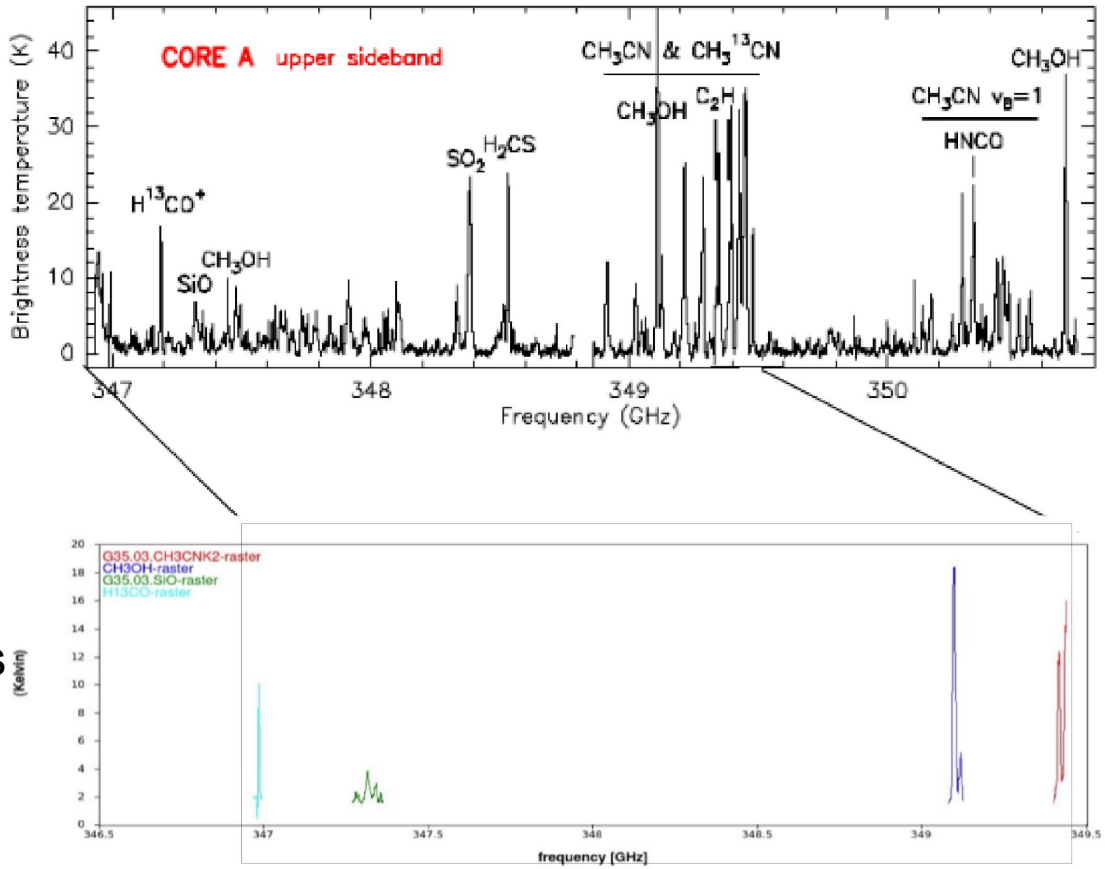






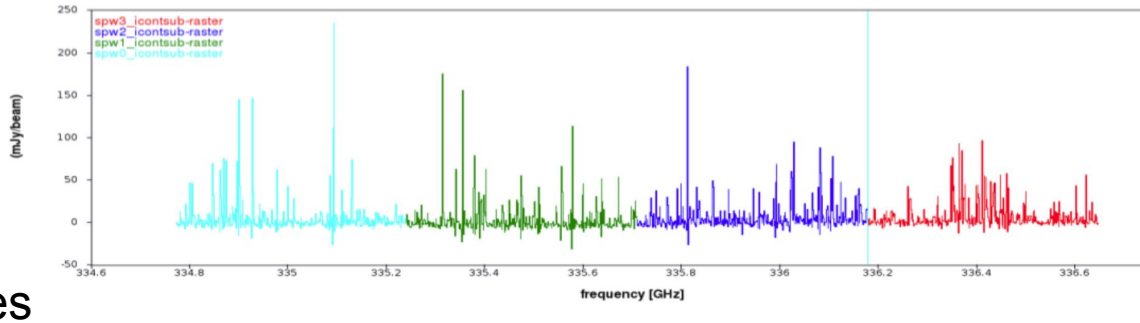
Science needs images!





Statistical analysis requiring many datasets are impossibly time comsuming
(see the timeline of J0635-7516, calibrator In hundreds of projects)

Images in the archive are incomplete and Not representative of the data content
(see the comparison of published and archived spectra for a pre-stellar object)



Science exploitation need complete, homogeneous, and easily accessible images

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 philosophy:

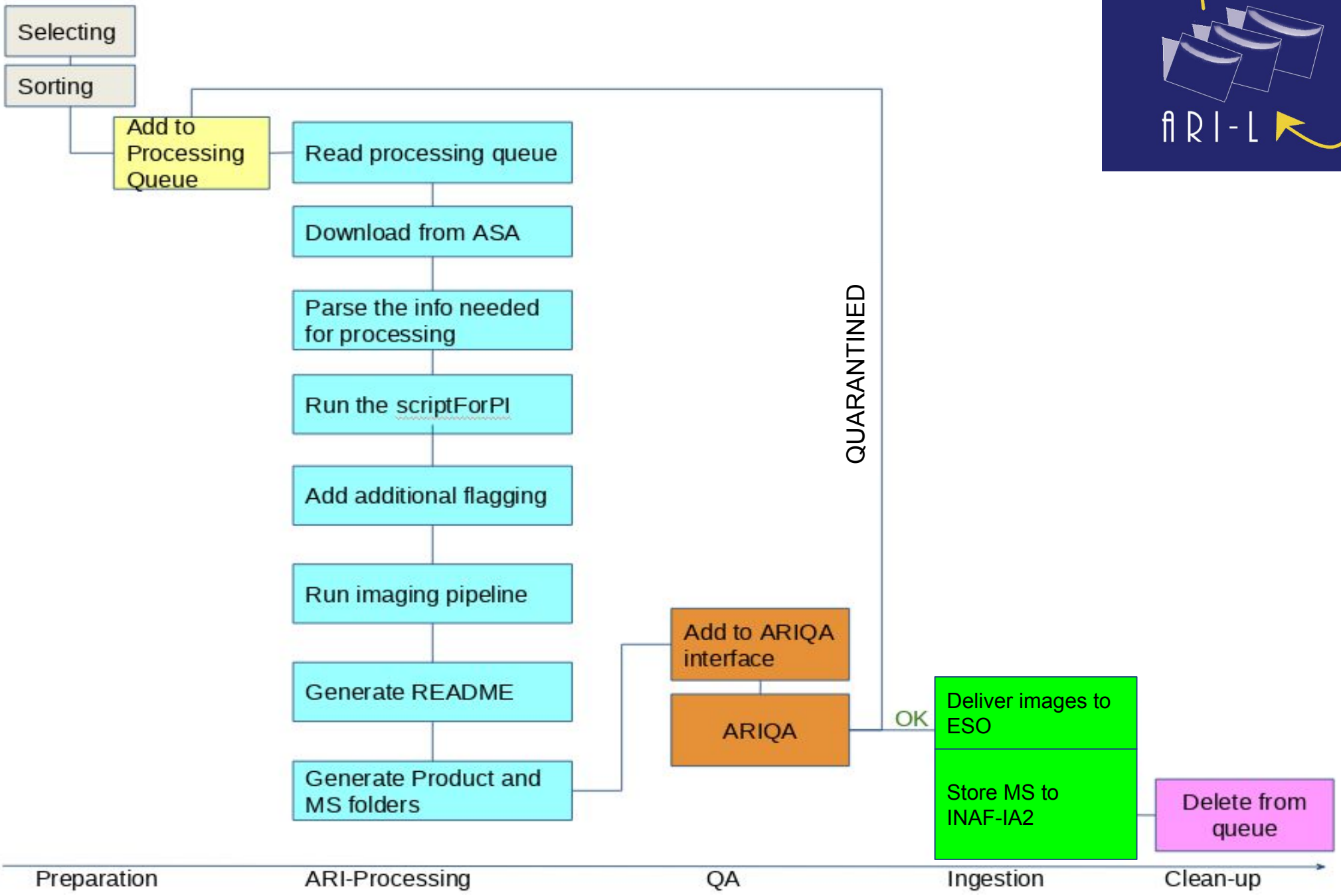
- **THINK OF THE MINERS**: we will produce imaging products highly relevant for all science-cases and enhance the possibilities of exploitation of archival data also to non-expert data-miners,

- **HOMOGENEITY** we shall provide a homogeneous view of archive data content within ARI-L and wrt the following Cycles to compare datasets and to make a more conscious download selection,

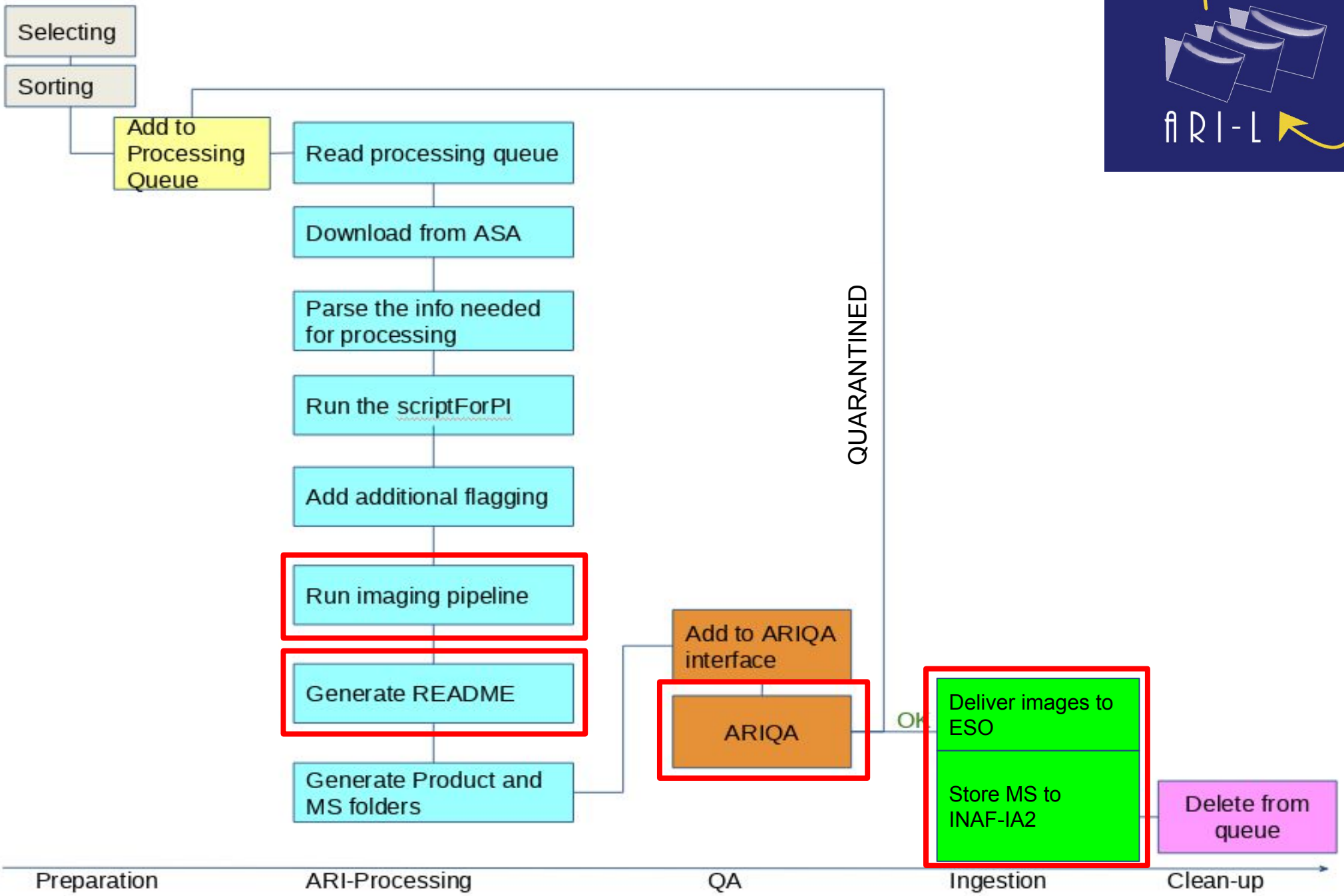
- **COMPLETENESS** we rate the 70% goal on the number of MOUs but we will try to complete as many projects as possible to complement the ASA resources,

- **ADD VALUE TO THE ASA** we aim at providing additional products that will complement and add value to the ASA, hence we have the responsibility of the quality of what we deliver to be ingested.

THE ARI-L PROJECT : WORKFLOW



THE ARI-L PROJECT : WORKFLOW



THE ARI-L PROJECT : PIPELINE & PRODUCTS

ALMA Pipeline became available for imaging in Cycle 4 (2017).

Later cycles are mostly imaged with the pipeline, but features became more and more efficient with time: not all the datasets were pipelineable.

Earlier cycles were manually imaged.
When introduced, the pipeline was not commissioned to be back compatible, so it can be used with earlier cycles but with care: our study demonstrated the ranges and conditions of applicability and the rate of success for the pipeline applied to earlier cycles (Massardi et al. 2018).

We, on first attempt, use the same parameters used in later cycles, aiming for homogeneity of the archival products.

Hence, **like for later cycles we produce in any analyzed MOUS:**

- **1 aggregated continuum image**
- **1 mfs image for each spw**
- **1 cube at the native resolution for each spw**

for the phase, bandpass and check source calibrator and for each science target (notice that later cycles products include only continuum for calibrators).

README

Additional Representative Images for Legacy (ARI-L) products

###

The ARI-L project products consist of a uniform set of continuum images and native spectral resolution cubes generated with the ALMA Imaging Pipeline for each source included in processable MOUS of cycles 2-4. The project is described in the ARI-L documentation at https://almascience.org/alma-data/aril. We stress that these images should be considered indicative of the data content but are not intended to be science-ready.

Information about the content, processing and quality of the data-reduction can be found in the *weblog.tgz file.

GENERAL ON ARI-L

###

MOUS description

Cycle: 4
Project code: 2016.1.00798.S
SB name:2016.1.00798.S
PI name: V. Mainieri
Project title: SUPER-ALMA: gas fractions and depletion timescales in AGN hosts at z~2
Configuration:C40-4
Proposed rms:0.5 mJy in 100 km/s
Proposed beam size: 1 arcsec
CASA version used for reduction: 4.7.0-1
QA2 Result: PASS
Total Number of Member SBs in this OUS Group:1

BASIC PROJECT INFO

###

Notes of the ARI-L process

ARI-L processing CASA version: 5.6.1-8

ARI-L QA NOTES

EB INFO: TIME, ARRAY, SPW RANGE & RES., UVRANGE

###

List of Execution Blocks in this MOUS:

#id	$ EB_name	$ Timerange[MJD]	$ Array	$ #ofAntennas	$ SPW_ranges[GHz]	$ ch_res[MHz]
1	$ uid___A002_Xbb154b_X947	$ 57717.9368234-57717.9551974	$	$ 42	$ [103.11-104.99, 102.05-103.93, 90.05-91.93, 91.55-93.43]	$ [7.813, 7.813, 7.813, 7.8

###

Source properties

Source	$ J0006-0623	$ J0202-0559	$ X-N-128-48
ra[h]	$ 0.10386	$ 2.03524	$ 2.10376
dec[deg]	$ -6.39315	$ -5.98337	$ -4.09534
EB_id	$ [1]	$ [1]	$ [1]
NumberOfPointings	$ [1]	$ [1]	$ [1]
LAS[arcsec]	$ 23.61333	$ 23.61333	$ 23.61333
Resolution[arcsec]	$ 0.50641	$ 0.50641	$ 0.50641
Intent	$ ['BANDPASS#ON_SOURCE', 'FLUX#ON_SOURCE', 'WVR#ON_SOURCE']	$ ['PHASE#ON_SOURCE', 'WVR#ON_SOURCE']	$ ['OBSERVE_TARGET#ON_SOURCE']

TARGET INFO: RA, DEC, INTENT, ID, LAS, ANGULAR RES.

THE ARIQA

3-levels of Quality Assurance:

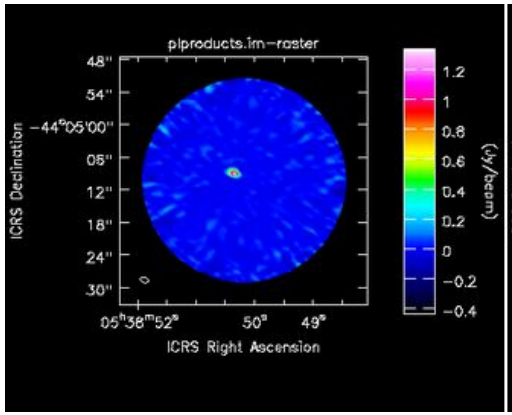
- **check** that the ARI-L code run was performed correctly
- perform a **weblog review** of the imaging pipeline products
- **compare** ARI-L products with QA2 products on rms, peak flux, dynamic range, having extracted the portion of the ARI-L products corresponding to what is available as QA2 products, smooth them to the same resolutions.

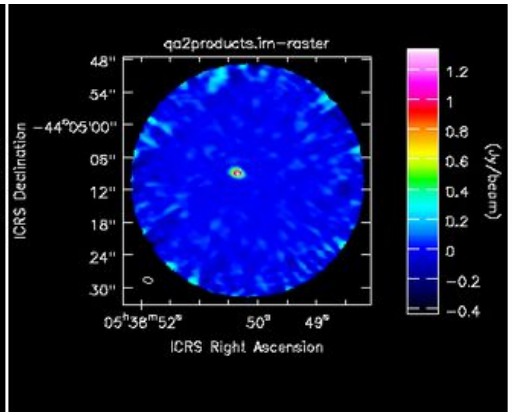
Num	Experiment Name	Date&Time	Quality Summary	Compare	Counting Exp. | Prod. | tt1 | status	README	Notes	Discard	Accept
1	2016.1.00627.S_uid___A001_X879_X8ad	2020-12-11 09:38:37	Task Summary	Compare FITS	README	| |	Comment experiment...	Quarantine	OK
92	2015.1.00144.S_uid___A001_X2fb_X1f0	2020-12-03 09:09:19	Task Summary	Compare FITS	27 | 30 | 3 | OK	README			ACCEPTED
93	2015.1.00144.S_uid___A001_X2fb_X1e8	2020-12-03 09:09:18	Task Summary	Compare FITS	27 | 30 | 3 | OK	README			ACCEPTED
94	2015.1.00370.S_uid___A001_X2fe_X450	2020-12-03 09:09:18	Task Summary	Compare FITS	36 | 36 | 0 | OK	README			ACCEPTED
95	2016.2.00014.S_uid___A001_X1234_X31c	2020-12-03 09:09:16	Task Summary	Compare FITS	27 | 30 | 3 | OK	README			ACCEPTED
96	2015.1.00932.S_uid___A001_X2f6_X1f1	2020-12-03 09:09:16	Task Summary	Compare FITS	27 | 27 | 0 | OK	README			ACCEPTED

Home | By Topic | By Task | Project Code N/A

Task Summaries

Task		QA Score		Duration
1. hifa_importdata: Register measurement sets with the pipeline	1/1 have HISTORY		0.50	0:03:44
2. hif_mstransform: Create science target MS			1.00	0:05:02
3. hifa_flagtargets: ALMA Target flagging			1.00	0:00:27
4. hifa_imageprecheck: ImagePreCheck			1.00	0:13:45
5. hif_checkproductsize: Check product size			1.00	0:00:53
6. hif_makeimlist: Set-up parameters for target per-spw continuum imaging			1.00	0:00:14
7. hif_findcont: Detect continuum frequency ranges			1.00	0:25:43
8. hif_uvcontfit: UV continuum fitting			1.00	0:03:38
9. hif_uvcontsub: UV continuum subtraction			1.00	0:04:41





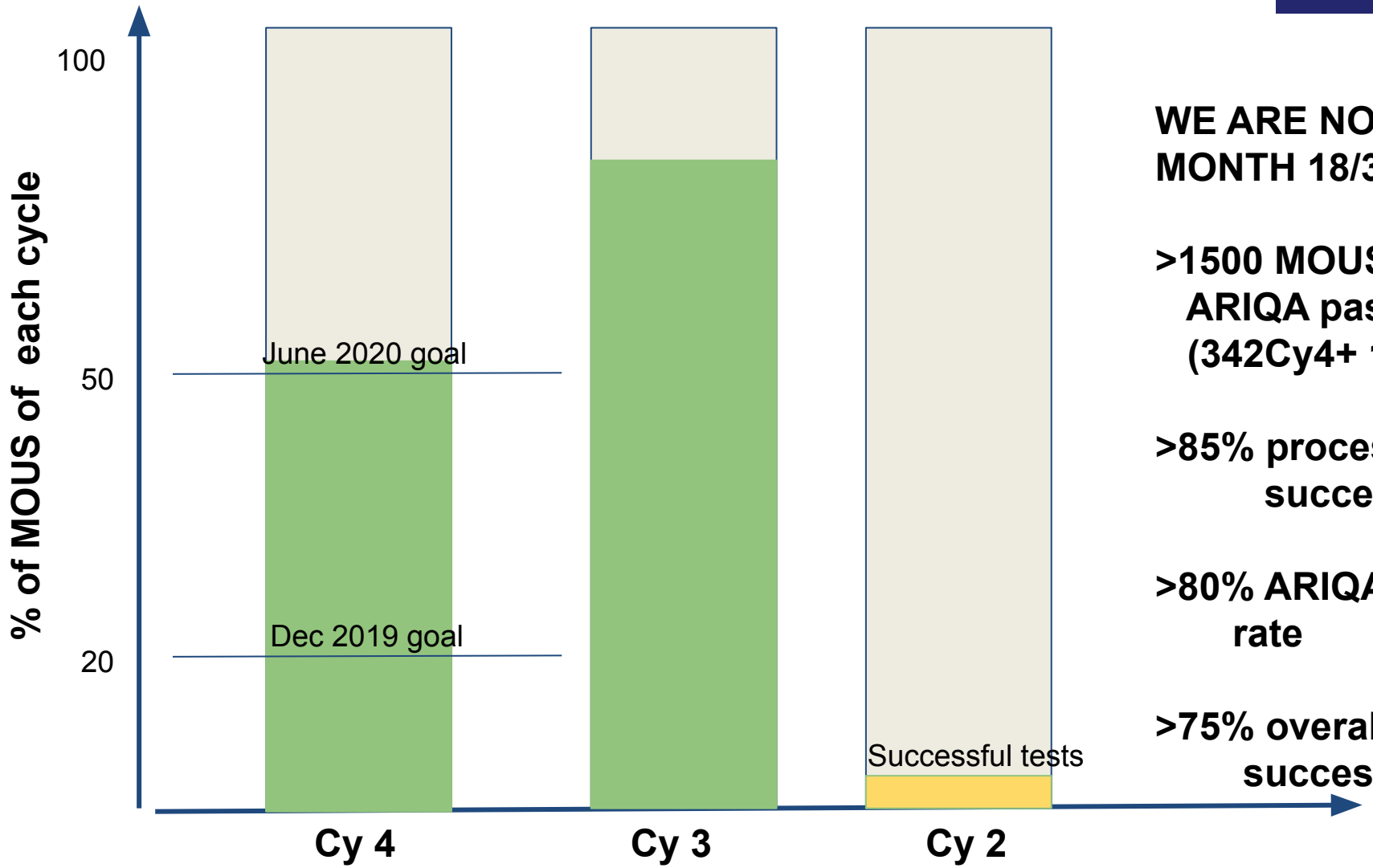
rms: 0.063
dyn: 20.941
max: 1.323
min: -0.055
flux: 1.812

rms: 0.071
dyn: 18.796
max: 1.340
min: -0.054
flux: 1.600

rms: -6.013% **dyn:** 5.399% **flux:** 6.223%

CURRENT STATUS: rates

TOTAL MOUS: 672 + 1719 + 1085 = 3476



WE ARE NOW AT MONTH 18/36

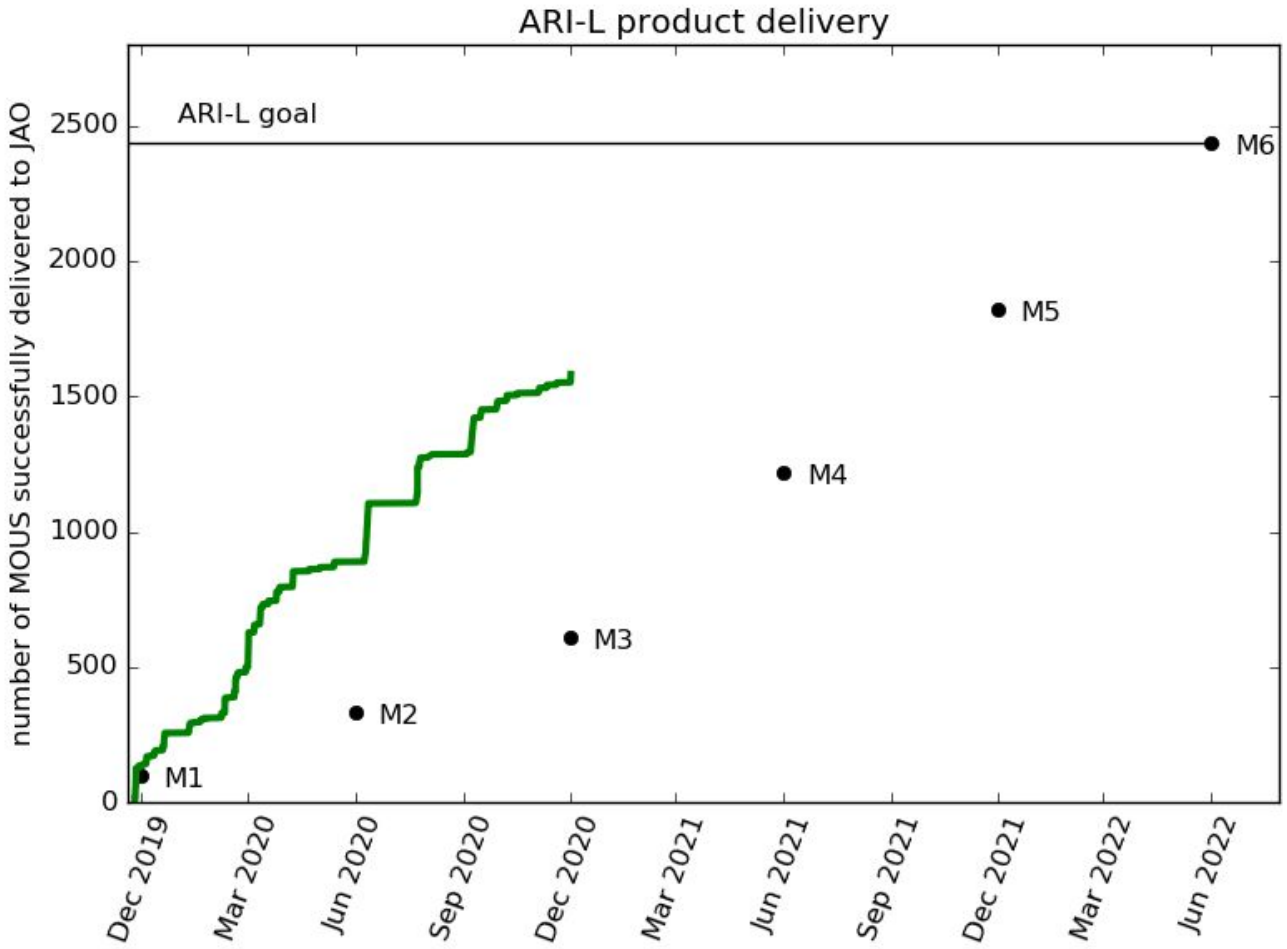
>1500 MOUS ARIQA passed (342Cy4+ 1200Cy3)

>85% processing success rate

>80% ARIQA success rate

>75% overall current success rate

CURRENT STATUS: delivery



M6: 2434 MOUS JUNE 2022

M5: 1824 MOUS DEC 2021

M4: 1217 MOUS JUNE 2021

M3: 608 MOUS DEC 2020

M2: 336 CY4 JUNE 2020

M1: 100 MOUS DEC 2019

ARI-L PRODUCTS IN THE ASA

The ARI-L products do not modify the current deliverables,
but are included into ASA as "**externally contributed products**".

More than 75000 ARI-L images are currently available

ALMA Request Handler

Login

Anonymous User: Request #3000000754199 ✓

Request Title: click to edit

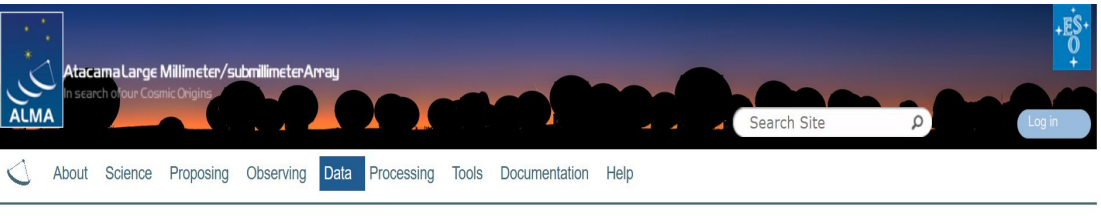
Download Selected

☐ readme ☐ product ☐ auxiliary ☐ raw ☐ raw (semipass) ☑ external

Project / OUSet / Executionblock	File	Size	Accessible	Actions
Request 3000000754199		1 GiB		
Project 2015.1.01362.S		1 GiB		
Science Goal OUS uid://A001/X2d6/X224		1 GiB		
Group OUS uid://A001/X2d6/X225		1 GiB		
Member OUS uid://A001/X2d6/X226		1 GiB		
SB SDP11_a_09_TE				
readme	member.uid___A001_X2d6_X226.README.txt	14 KiB	✓	
product	2015.1.01362.S_uid___A001_X2d6_X226_001_of_001.tar	45 MiB	✓	
auxiliary	2015.1.01362.S_uid___A001_X2d6_X226_auxiliary.tar	60 MiB	✓	
raw	2015.1.01362.S_uid___A002_Xbab09c_X2c2d.asdm.sdm.tar	8 GiB	✓	
external	2015.1.01362.S_uid___A001_X2d6_X226_external_001_of_001.tar	1 GiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_1_2_3_681205MHz.12m.cont.l.mask.fits.gz	5 KiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_1_2_3_681205MHz.12m.cont.l.pb.fits.gz	186 KiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_1_2_3_681205MHz.12m.cont.l.pbcor.fits	366 KiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_682161MHz.12m.cube.l.pb.fits.gz	20 MiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_682161MHz.12m.cube.l.pbcor.fits	41 MiB	✓	
external	member.uid___A001_X2d6_X226.ari_l.J0854p2006_bp.spw0_682161MHz.12m.mfs.l.mask.fits.gz	5 KiB	✓	

ARI-L WEBSITE

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



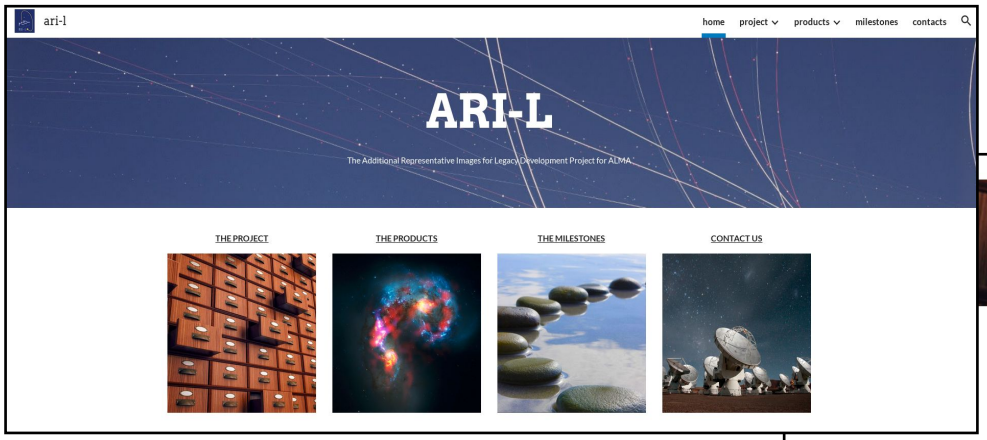
ARI-L

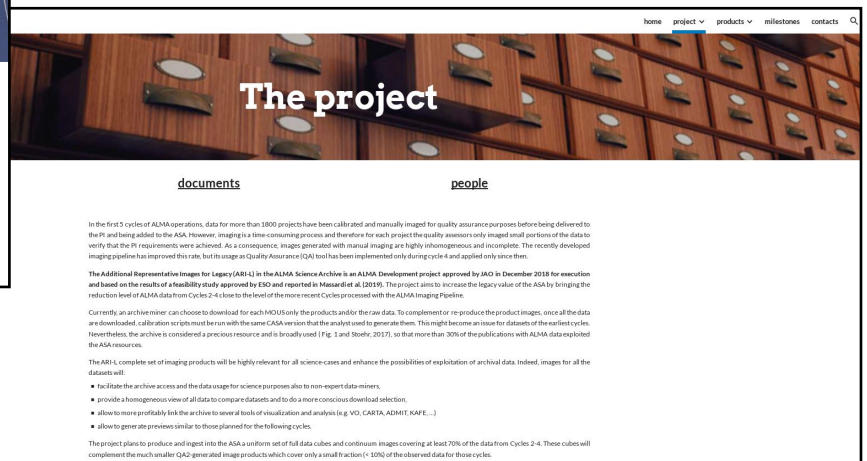
The Additional Representative Images for Legacy (ARI-L) in the ALMA Science Archive (ASA) is an Development project for ALMA upgrade approved by the ALMA Director in December 2018. It is based on the results of a feasibility study approved by ESO, as reported in Massardi et al. (2019). The project aims to increase the legacy value of the ALMA Science Archive by bringing the reduction level of ALMA data from Cycles 2-4 close to that of what is processed with the ALMA Imaging Pipeline in more recent Cycles.

The complete set of ARI-L imaging products will be very relevant for all science fields and enhance the possibilities of exploitation of archival data. Indeed, making these images available for all the datasets will:

- facilitate archive access and data usage for science purposes, also to non-expert data miners,
- provide a homogeneous view of all data, in order to compare datasets and enable a more conscious download selection,
- permit to more profitably link the archive to several visualisation and analysis tools (e.g. VO, CARTA, ADMIT, KAFE, ...),
- allow users to generate previews similar to those planned for future cycles.

https://sites.google.com/inaf.it/ari-l

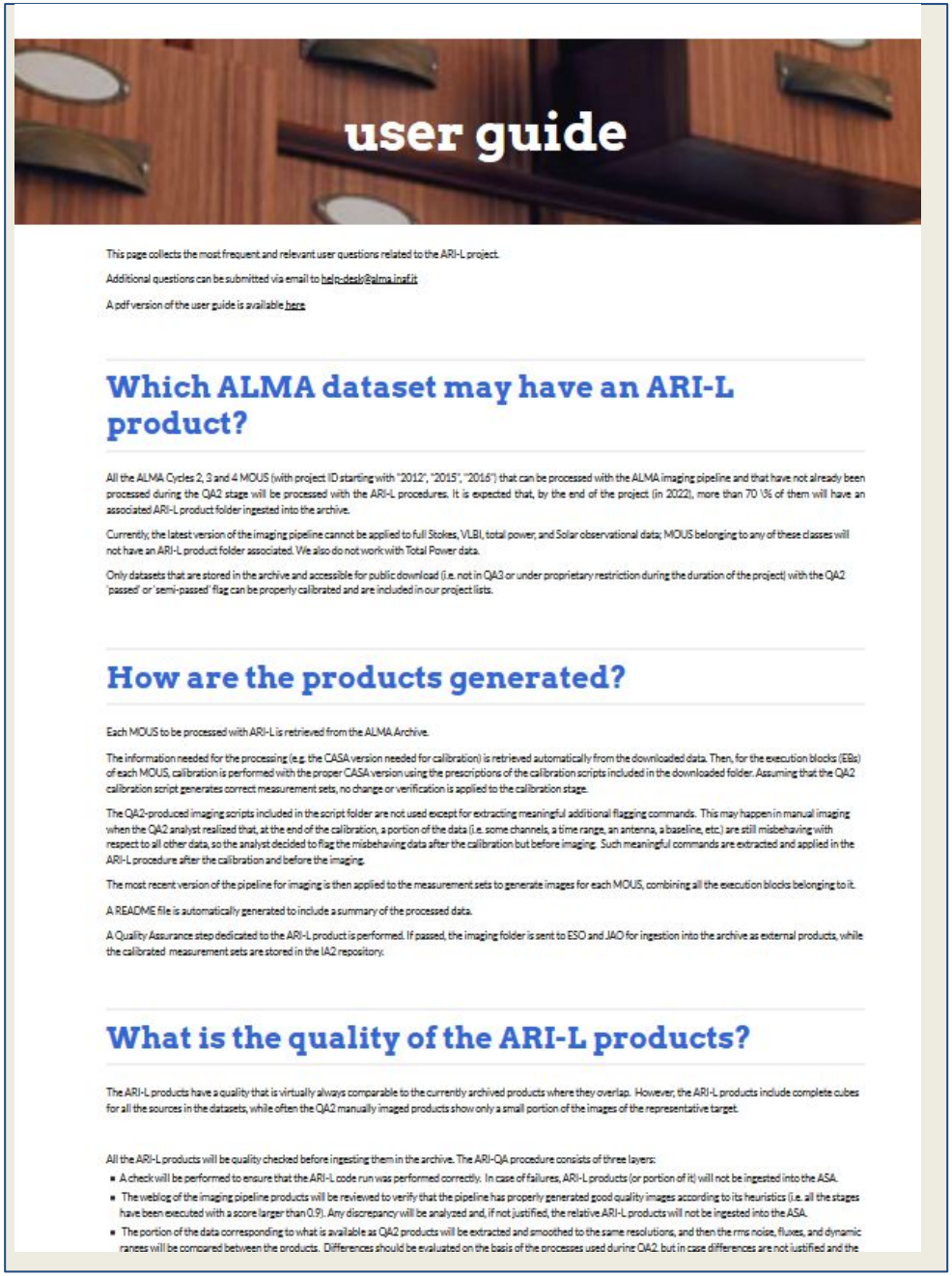




Information for ARI-L products User:

- **project rationale:** why ARI-L
- **project details:** people, timelines, workflow, user guide
- **status of the products and project**
- **contact details for support** help-desk@alma.inaf.it

DOCUMENTATION: User Guide



user guide

This page collects the most frequent and relevant user questions related to the ARI-L project.

Additional questions can be submitted via email to help-desk@alma.inaf.it

A pdf version of the user guide is available here

Which ALMA dataset may have an ARI-L product?

All the ALMA Cycles 2, 3 and 4 MOUS (with project ID starting with "2012", "2015", "2016") that can be processed with the ALMA imaging pipeline and that have not already been processed during the QA2 stage will be processed with the ARI-L procedures. It is expected that, by the end of the project (in 2022), more than 70 \3% of them will have an associated ARI-L product folder ingested into the archive.

Currently, the latest version of the imaging pipeline cannot be applied to full Stokes, VLBI, total power, and Solar observational data; MOUS belonging to any of these classes will not have an ARI-L product folder associated. We also do not work with Total Power data.

Only datasets that are stored in the archive and accessible for public download (i.e. not in QA3 or under proprietary restriction during the duration of the project) with the QA2 'passed' or 'semi-passed' flag can be properly calibrated and are included in our project lists.

How are the products generated?

Each MOUS to be processed with ARI-L is retrieved from the ALMA Archive.

The information needed for the processing (e.g. the CASA version needed for calibration) is retrieved automatically from the downloaded data. Then, for the execution blocks (EBs) of each MOUS, calibration is performed with the proper CASA version using the prescriptions of the calibration scripts included in the downloaded folder. Assuming that the QA2 calibration script generates correct measurement sets, no change or verification is applied to the calibration stage.

The QA2-produced imaging scripts included in the script folder are not used except for extracting meaningful additional flagging commands. This may happen in manual imaging when the QA2 analyst realized that, at the end of the calibration, a portion of the data (i.e. some channels, a time range, an antenna, a baseline, etc.) are still misbehaving with respect to all other data, so the analyst decided to flag the misbehaving data after the calibration but before imaging. Such meaningful commands are extracted and applied in the ARI-L procedure after the calibration and before the imaging.

The most recent version of the pipeline for imaging is then applied to the measurement sets to generate images for each MOUS, combining all the execution blocks belonging to it.

A README file is automatically generated to include a summary of the processed data.

A Quality Assurance step dedicated to the ARI-L product is performed. If passed, the imaging folder is sent to ESO and JAO for ingestion into the archive as external products, while the calibrated measurement sets are stored in the IA2 repository.

What is the quality of the ARI-L products?

The ARI-L products have a quality that is virtually always comparable to the currently archived products where they overlap. However, the ARI-L products include complete cubes for all the sources in the datasets, while often the QA2 manually imaged products show only a small portion of the images of the representative target.

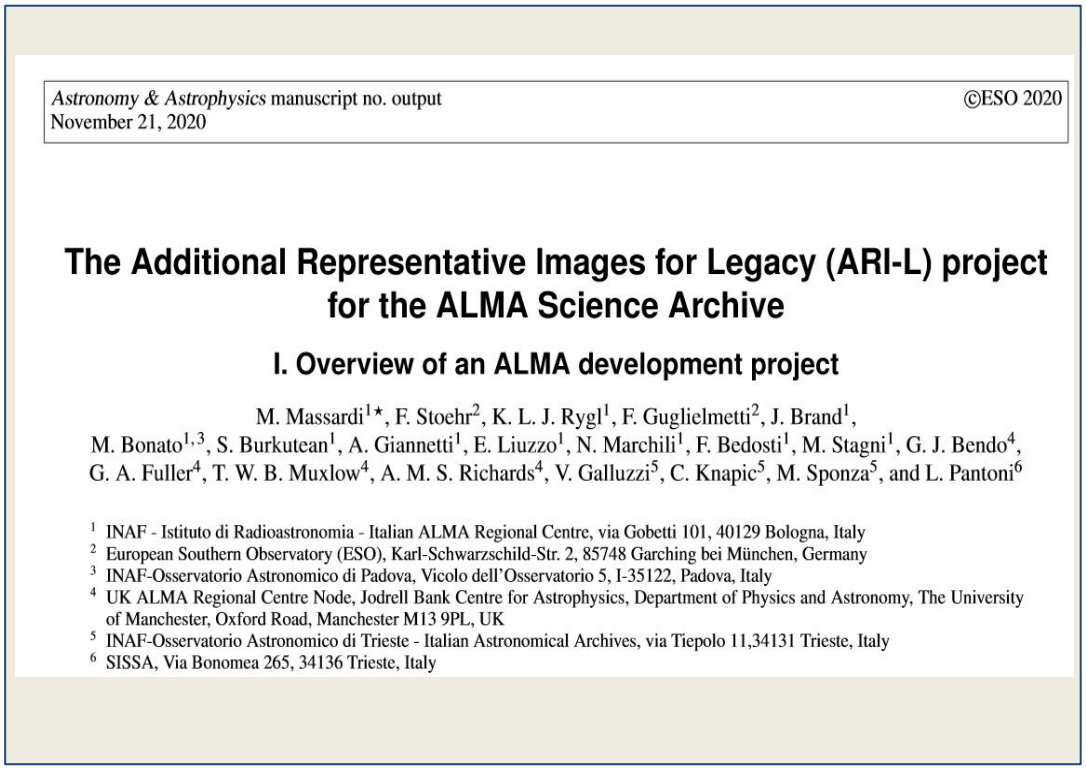
All the ARI-L products will be quality checked before ingesting them in the archive. The ARI-QA procedure consists of three layers:

- A check will be performed to ensure that the ARI-L code run was performed correctly. In case of failures, ARI-L products (or portion of it) will not be ingested into the ASA.
- The weblog of the imaging pipeline products will be reviewed to verify that the pipeline has properly generated good quality images according to its heuristics (i.e. all the stages have been executed with a score larger than 0.9). Any discrepancy will be analyzed and, if not justified, the relative ARI-L products will not be ingested into the ASA.
- The portion of the data corresponding to what is available as QA2 products will be extracted and smoothed to the same resolutions, and then the rms noise, fluxes, and dynamic ranges will be compared between the products. Differences should be evaluated on the basis of the processes used during QA2, but in case differences are not justified and the

Collection of FAQ for quick recovery of informations.
Periodically updated

https://sites.google.com/inaf.it/ari-l/project/user-guide

DOCUMENTATION: first ARI-L paper



Astronomy & Astrophysics manuscript no. output
November 21, 2020
©ESO 2020

The Additional Representative Images for Legacy (ARI-L) project for the ALMA Science Archive

I. Overview of an ALMA development project

M. Massardi[1]⋆, F. Stoehr[2], K. L. J. Rygl[1], F. Guglielmetti[2], J. Brand[1], M. Bonato[1,3], S. Burkutean[1], A. Giannetti[1], E. Liuzzo[1], N. Marchili[1], F. Bedosti[1], M. Stagni[1], G. J. Bendo[4], G. A. Fuller[4], T. W. B. Muxlow[4], A. M. S. Richards[4], V. Galluzzi[5], C. Knapic[5], M. Sponza[5], and L. Pantoni[6]

[1] INAF - Istituto di Radioastronomia - Italian ALMA Regional Centre, via Gobetti 101, 40129 Bologna, Italy
[2] European Southern Observatory (ESO), Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
[3] INAF-Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy
[4] UK ALMA Regional Centre Node, Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
[5] INAF-Osservatorio Astronomico di Trieste - Italian Astronomical Archives, via Tiepolo 11,34131 Trieste, Italy
[6] SISSA, Via Bonomea 265, 34136 Trieste, Italy

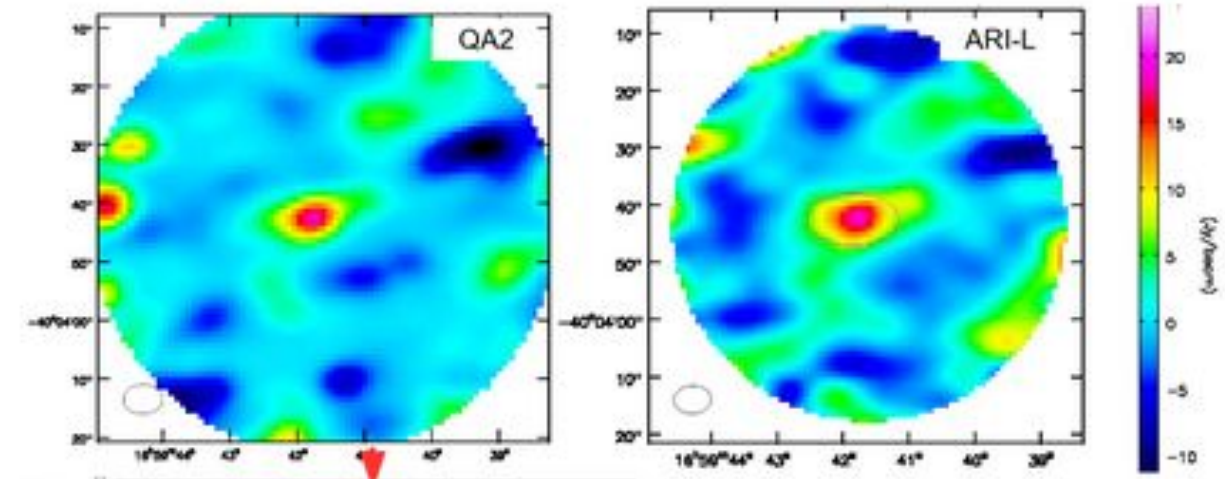
Information for general archival User:

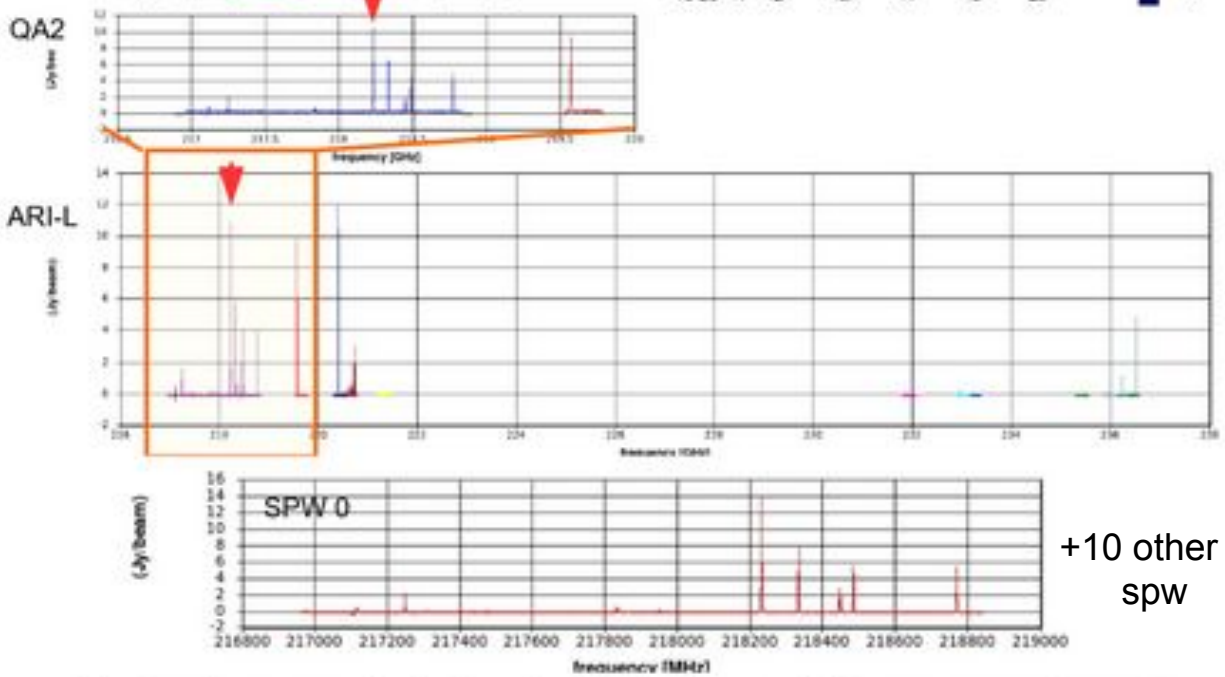
- **project rationale:** why ARI-L
- **project activities:** what, how, QA
- **description of the products and status**
- **examples of applications (with emphasis on cy4 and cy3 data)**

Submitted on nov 1st 2020.

Users of ARI-L products are encouraged to acknowledge this paper.

JUST A FLAVOUR OF ARI-L SCIENCE





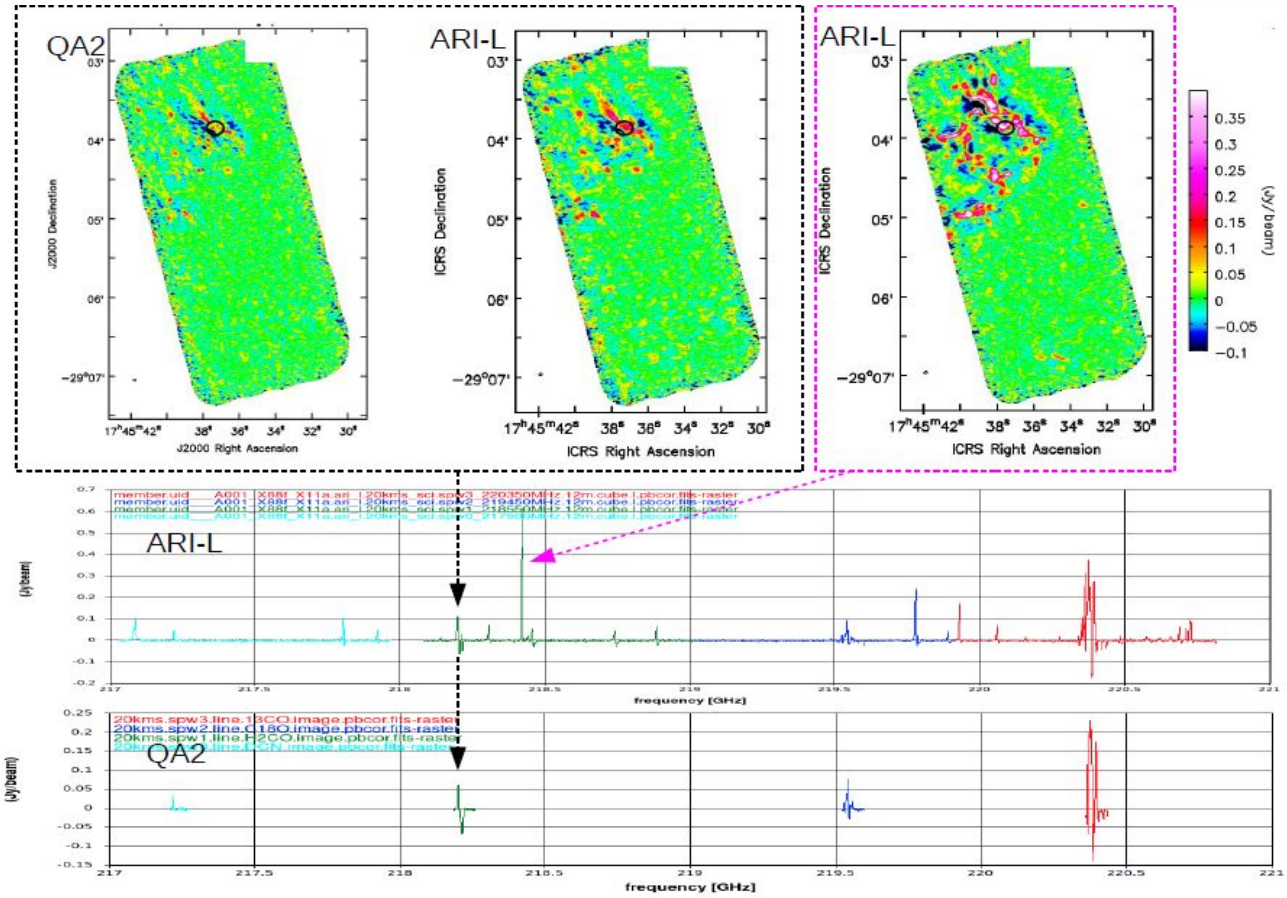
+10 other spw

Massive star formation in O-type young stellar object (G345.49+1.47) (PI: Cesaroni)

The complete spectral coverage of the ARI-L cubes shows the rich chemistry of this object, useful to investigate the presence of Keplerian disks or outflows and constrain the mechanisms of massive star formation.

Massardi et al. subm.

JUST A FLAVOUR OF ARI-L SCIENCE



Massardi et al. subm.

Filamentary structure in star forming regions (PI: Kauffmann).

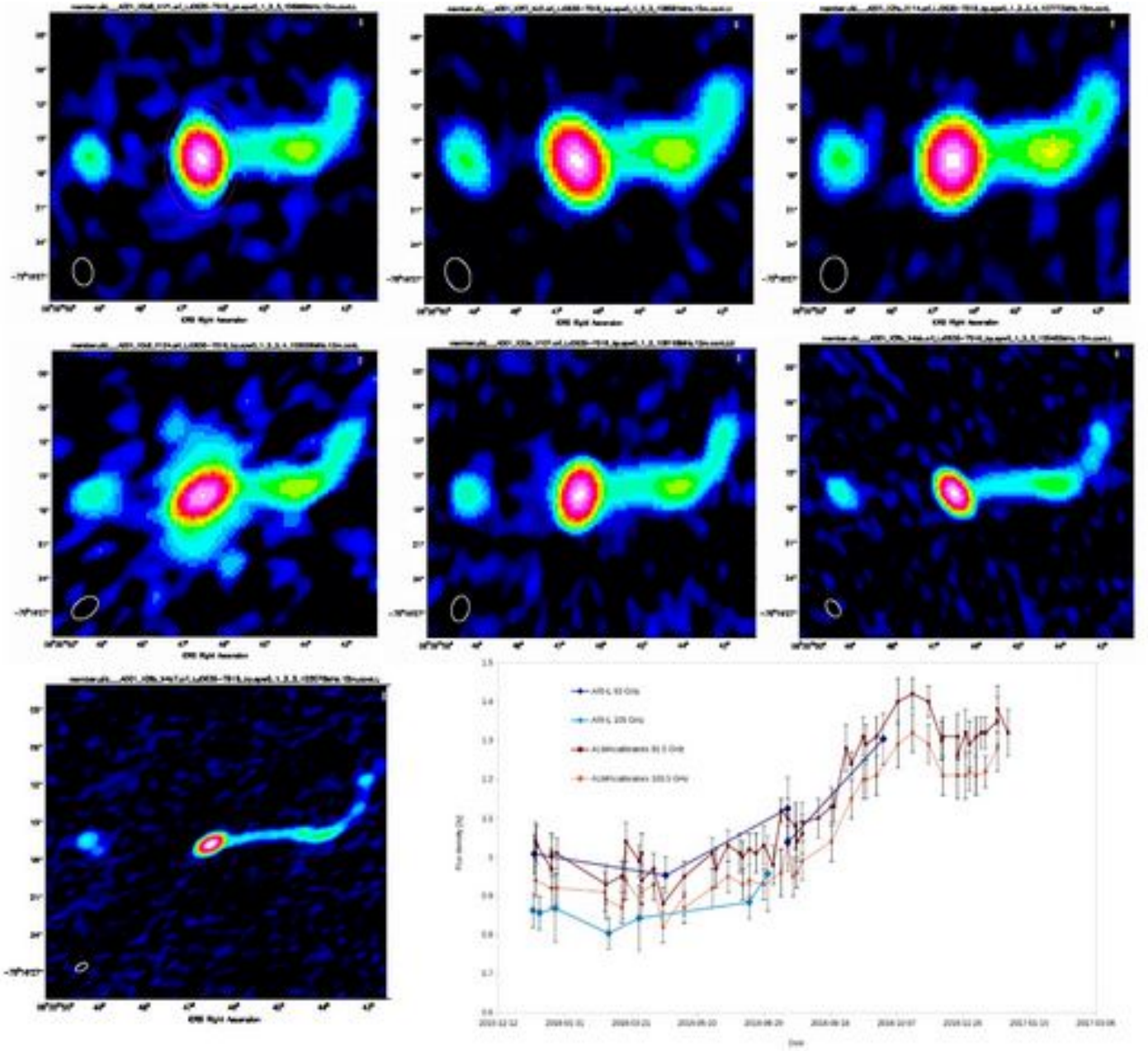
The QA2 images collects only one spectral line for each spectral windows (as requested to QA2 analysts to establish the data quality).

The automasking used by the imaging pipeline exploited in ARI-L extracts structure to higher significance with respect to the QA2 image.

The ARI-L complete spectrum can identify more spectral features that are present in the data.

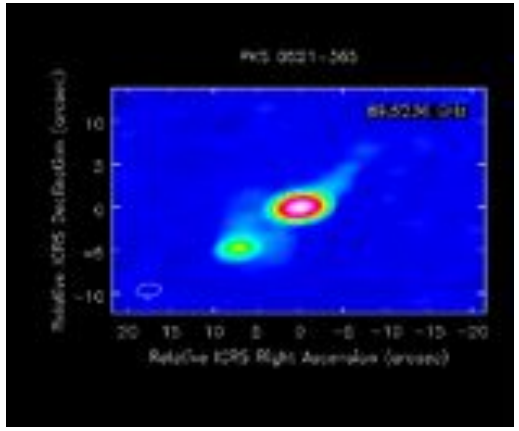
JUST A FLAVOUR OF ARI-L SCIENCE

PKS 0635-752

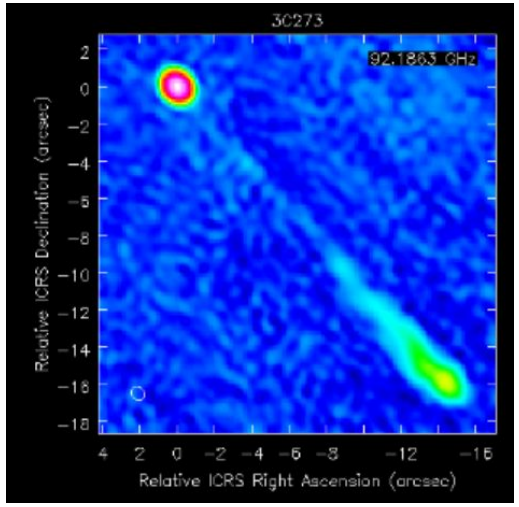


Massardi et al. subm.

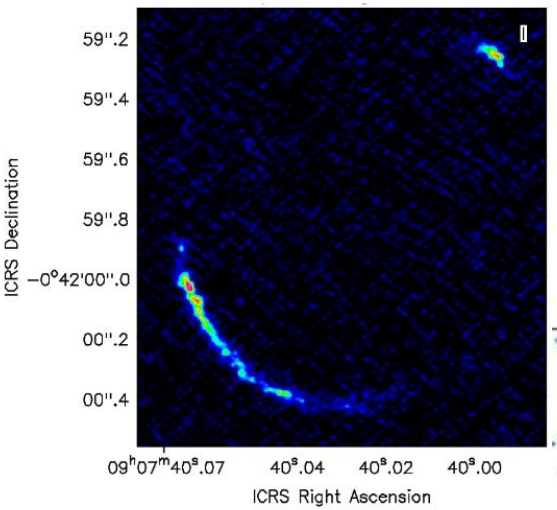
PKS 0521-365



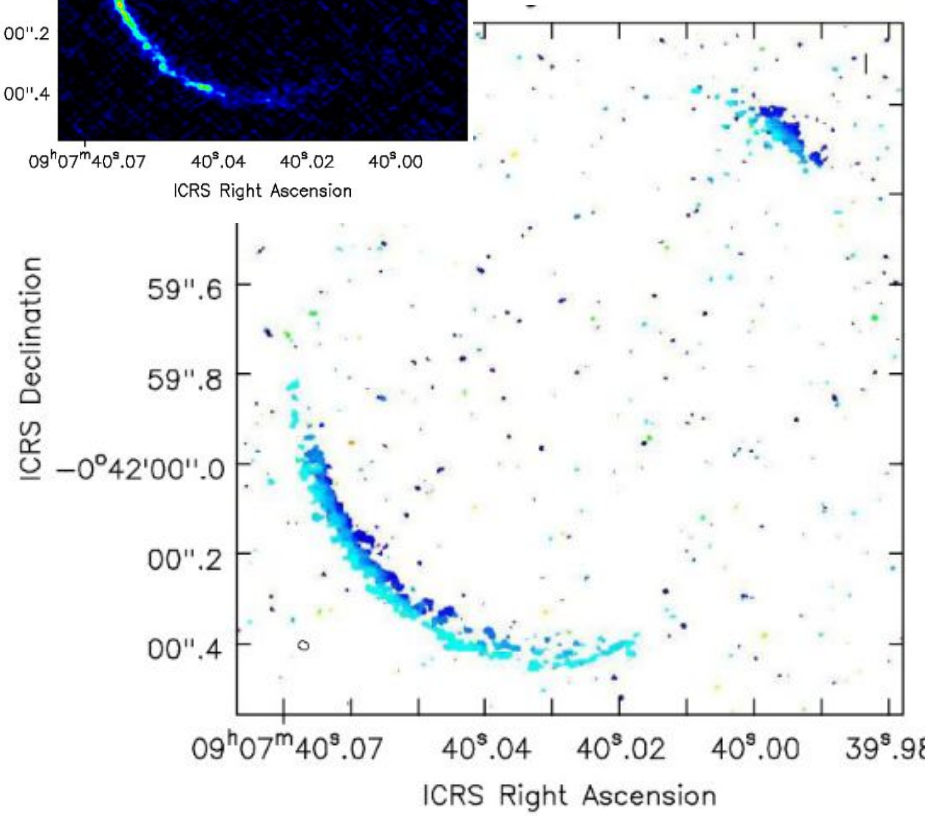
3C273

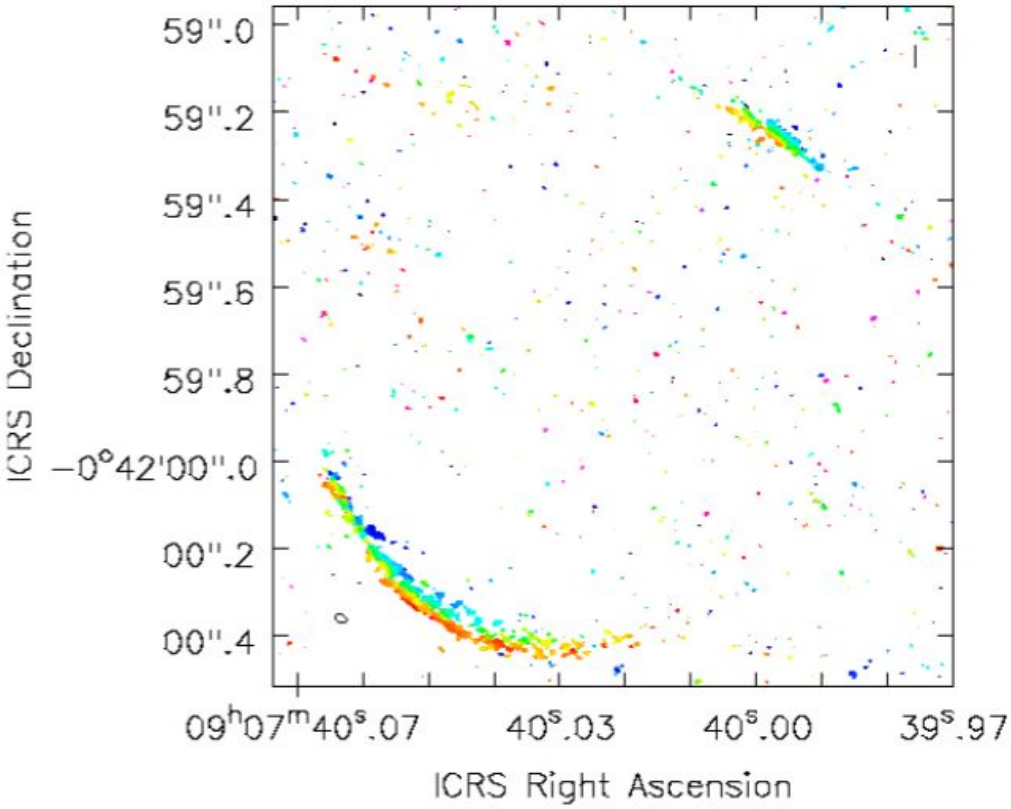


JUST A FLAVOUR OF ARI-L SCIENCE



mom1 of CO(3-2) on SDP9 as published 3 yrs ago
compared with ARI-L momentum





Massardi et al. (2017 vs subm.)

SUMMARY

- The project is in full swing!
- Over 1500 MOUS have already successfully been delivered to the ALMA Science Archive
- All delivery milestones have been met (well ahead of time)
- Overall success-rate above the target of 70%
- A dedicated ALMA Science Portal page, a website, and a user guide have been made publicly available
- An ARI-L paper has just been submitted.

Please visit: https://almascience.eso.org/alma-data/aril