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 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...

![Pie chart showing mining efficiency]

Mining efficiency affects the science exploitation efficiency
The ALMA Science Archive from a miner perspective

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

THE ASA IS A GOLD MINE

- 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
**ALMA FITS Archive : Target Info**

Target Name : NGC5135

- **Filter by Frequency**

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

Number of data per page : 20

Ordered by dataset_id (desc)

Total number : 4

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

<table>
<thead>
<tr>
<th>#</th>
<th>dataset id</th>
<th>project code</th>
<th>all</th>
<th>image</th>
<th>spect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ALMA01081334</td>
<td>2013.1.00524.S</td>
<td>☐</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="spectrogram1.png" alt="Spectrogram" /></td>
</tr>
<tr>
<td>2</td>
<td>ALMA01081333</td>
<td>2013.1.00524.S</td>
<td>☐</td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="spectrogram2.png" alt="Spectrogram" /></td>
</tr>
<tr>
<td>3</td>
<td>ALMA01081334</td>
<td>2013.1.00524.S</td>
<td>☐</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="spectrogram3.png" alt="Spectrogram" /></td>
</tr>
</tbody>
</table>
Image analysis tools

(Burkutean et al. 2018)

Please write to kafe@ira.inaf.it for access information to the web interface.
Statistical analysis requiring many datasets are impossibly time consuming (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

Science needs images!
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

- Selecting
  - Add to Processing Queue
  - Read processing queue
  - Download from ASA
  - Parse the info needed for processing
  - Run the scriptForPI
  - Add additional flagging
  - Run imaging pipeline
  - Generate README
  - Generate Product and MS folders

- Sorting

- ARI-Processing
  - Add to ARIQA interface
  - ARIQA

- QA

- Ingestion
  - Deliver images to ESO
  - Store MS to INAF-IA2

- Clean-up

Preparation ARI-Processing QA Ingestion Clean-up

Quarantined
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).
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/ari-l. 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.

### BASIC PROJECT INFO

**Notes of the ARI-L process**

ARI-L processing CASA version: 5.6.1-8

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

<table>
<thead>
<tr>
<th>EB</th>
<th>TimeRange [MJD]</th>
<th>Array</th>
<th>No. of Antennas</th>
<th>SPW Range [GHz]</th>
<th>ch_res [MHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u10__A087_Xbb154h_X947</td>
<td>57717.9588234-57717.9551974</td>
<td>$42$</td>
<td></td>
<td>[108-1-104.99, 102.05-103.93, 90.05-91.93, 91.55-93.45]</td>
<td>[7.813, 7.813, 7.813, 7.813]</td>
</tr>
</tbody>
</table>

### SOURCE PROPERTIES

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.18086</td>
<td>-5.39315</td>
<td>1</td>
<td>1</td>
<td>23.61333</td>
<td>0.58641</td>
</tr>
<tr>
<td></td>
<td>0.30066-0.623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

| RA, DEC, INTENT, ID, LAS, ANGULAR RES. |
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.
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
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
ARI-L WEBSITE

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

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

https://sites.google.com/inaf.it/ari-l
Which ALMA dataset may have an ARI-L product?

All the ALMA Cycle 2, Cycle 4 NOUS files produced during Cycle 2 (2013.2, 2013.3, 2014.1) that can be processed with the ARA-L pipeline and that have not already been processed during the QA2 stage will be processed with the ARA-L procedures. It is expected that by the end of 2023, more than 70% of them will have an associated ARI-L product folder uploaded into the archive.

Currently, the latest version of the imaging pipeline cannot be applied to full Science, VLBA, Rapid, and Early observation data. NOUS belonging to any of those classes will not have an ARI-L product folder associated. We also do not work with Tera Reas data.

Only datasets that have been processed and accessible for public download (E, not in QA2) or under conditional restrictions during the duration of the project with the QA2 discussed in Art. 7 are released. Images are convolved and are included in the product data.

How are the products generated?

ARIA-L products are derived from the measurement data generated from NOUS containing the relevant data belonging to QA2.

The data is delivered to the measurement data for every NOUS containing the relevant data belonging to QA2. Those data are extracted and applied to the ARA-L processing pipeline after the calibration and before imaging.

The final product consists of a series of images that are applied to the measurement data to generate images for each NOUS containing the executed block belonging to QA2. AQA2/RAQL products are automatically generated to include convolved products.

What is the quality of the ARI-L products?

The ARA-L products have a quality that is similarly good to the currently accepted products where the ARA-L products include complete data for all the images in the database, unlike the QA2, where they are generated on a smaller portion of the images, or the represented target.

All the ARA-L products will be visually checked before being released. The ARA-L products consist of three layers:
- ARA-L products that have been visually checked.
- ARA-L products that have not been visually checked.
- ARA-L products that have been rejected.

The quality of the ARA-L products will be visually checked to verify that the data that has been visually checked is complete. The ARA-L products will be visually checked on a larger portion of the images, or the represented target.

The products of the dataset requiring in situ evaluation are ARA-L products with significant differences in image resolution, and the thermal, radiative, and dynamic properties of the source to the products. Differences should be evaluated on the basis of the combined results of the QA2 and RIA-L products.
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. Stochi$^2$, K. L. J. Rygl$^3$, F. Guglielmetti$^4$, 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$^5$, 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.
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.
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

PKS 0521-365

3C273

Massardi et al. subm.
mom1 of CO(3-2) on SDP9 as published 3 yrs ago compared with ARI-L momentum

Massardi et al. (2017 vs subm.)
- 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