DOI: 10.18727/0722-6691/5277

Overview of the Additional Representative Images for Legacy (ARI-L) Development Project for the ALMA Science Archive

Marcella Massardi[1,2]
Felix Stoehr[3]
George J. Bendo[4]
Matteo Bonato[1]
Jan Brand[1]
Vincenzo Galluzzi[5]
Fabrizia Guglielmetti[3]
Cristina Knapic[5]
Elisabetta Liuzzo[1]
Nicola Marchili[1]
Anita M. S. Richards[4]
Kazi L. J. Rygl[1]

[1] INAF–Institute of Radio Astronomy, ARC node, Bologna, Italy
[2] International School for Advanced Studies (SISSA), Trieste, Italy
[3] ESO
[4] Jodrell Bank Centre for Astrophysics, University of Manchester, UK
[5] INAF–Astronomical Observatory of Trieste, Italy

The Additional Representative Images for Legacy (ARI-L) project is a European Development project for ALMA Upgrade approved by the Joint ALMA Observatory and ESO in 2019. It aims to increase the legacy value of the ALMA Science Archive by bringing the reduction level of ALMA data from Cycles 2 to 4 close to that of data from more recent cycles processed for imaging with the ALMA Pipeline. To date, ARI-L has produced, assessed the quality of, and delivered more than 150 000 images. These represent more than 85% of the science datasets from Cycles 2 to 4 processable with the ALMA Pipeline but lacking pipeline-generated images, and accordingly the project accomplished all its goals during its official runtime.

ARI-L project rationale

ALMA was the first radio astronomy facility to offer calibrated, deconvolved images and data cubes as fundamental data products of the Joint ALMA Observatory (JAO).

These data products are not unique because of the relatively large freedom in parameter choices during the interferometric imaging process, but even in their generic form they provide a quick way for users to assess the data quality, the content of the data products and the interesting spatial and spectral regions. Depending on the science case, the pipeline-generated data products may also be used for scientific analysis. For these reasons, the image products are delivered to ALMA users through the ALMA Science Archive (ASA).

From ALMA's Early Science period up to Cycle 3 (i.e., up to projects observed in late 2015), the part of the ALMA pipeline dedicated to imaging was not available. The staff at the observatory and at the ALMA Regional Centres (ARCs) manually performed the quality assurance (QA2) of the data before they were delivered to the principal investigators (Petry et al., 2020). This manual procedure was carried out for over 1800 ALMA projects. The manual imaging of each full data set is very time consuming, so the analysis often focused on only a small subset of a project's calibrated data that was just large enough to assess the quality, mostly for the defined purposes of the project proposers rather than for potentially broader goals of users of the archival data. As a result, only a very small fraction of all raw data (< 10%) was converted into images and image cubes, and typically they do not include images of the calibrators. This fraction increased dramatically from Cycle 5 (i.e., from 2017) when the ALMA Science Pipeline was used almost exclusively for QA2 data reduction.

The availability of deconvolved images and data cubes vastly speeds up researchers' data analysis process. Archive researchers can not only download the images for local analysis, but also use the ALMA archive remote visualisation tools.

The use of a well-established pipeline makes the data analysis process more efficient and the products more homogeneous, even across projects, arrays and epochs, so the products can be compared or combined accurately. This aids investigation of variability, spectral and/or spatial behaviour, or the use of statistical techniques on samples taken from multiple observations.

The main goal of the ARI-L project was to use the ALMA Imaging Pipeline (as introduced in 2017) on the data from early observing Cycles (2–4) to create, where

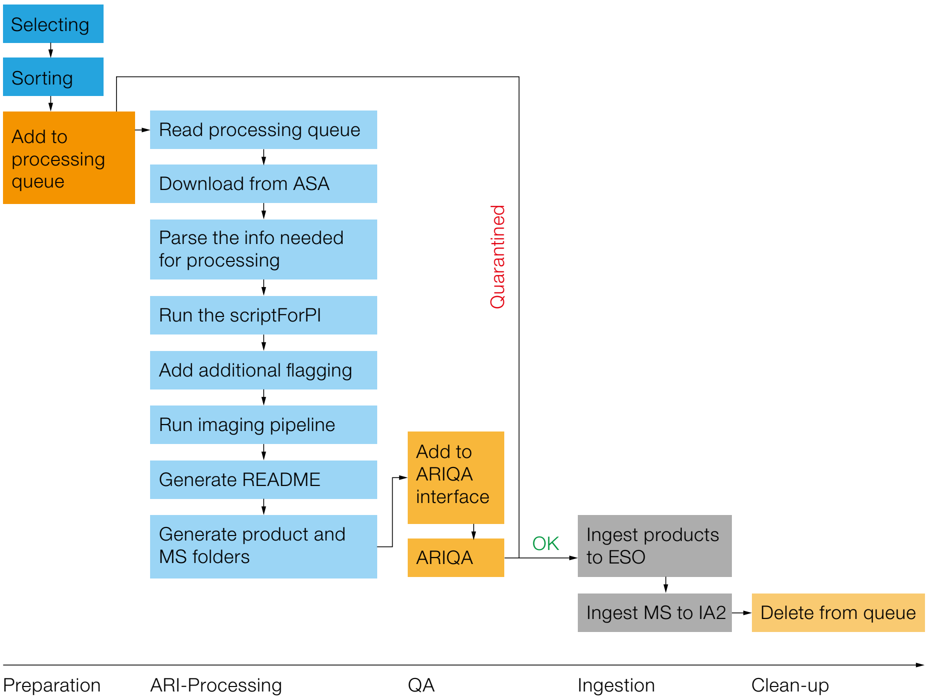


Figure 1. The decision tree and the different processes of the ARI-L project as applied to each MOUS in Cycles 2–4.

	Cycle 2	Cycle 3	Cycle 4	Total
Processable MOUS	973	1603	605	3181
Delivered	701	1495	518	2714
Processing and QA issues	272	108	87	46
Delivered fraction (%)	72.5	93.3	85.6	85.3

Table 1. Numbers and properties of datasets, grouped, as in the ASA, in Member Observing Unit Sets (MOUS) for ALMA Cycles 2–4 processed in the ARI-L project. Statistics are also reported for MOUS that failed the quality assurance procedure, either because of issues in the processing or identified in the QA stage. Fraction of successfully delivered MOUS are also reported. For comparison, the ARI-L main goal was to reach the 70% of the processable MOUS with a best effort goal of 80%.

possible, data products of the same completeness and quality as ALMA is now creating for new observations and to ingest them into the ASA.

This objective required the production of a uniform set of full data cubes and continuum images, covering at least 70% of the data from Cycles 2–4 which can be processed with the ALMA Pipeline, with a best efforts goal of 80%.

The project began processing in June 2019 and its first products were ingested into the ASA in November 2019. After the three years of its official runtime (and despite operating under pandemic conditions) the ARI-L project delivered more than 85% of the data from Cycles 2–4 that can be processed with the ALMA Pipeline, reaching and surpassing all its goals.

The ARI-L cubes and images complement the much more limited number of archival image products generated during the data quality assurance stages (QA2), which cover only a small fraction of the available data for those cycles.

The ARI-L project workflow

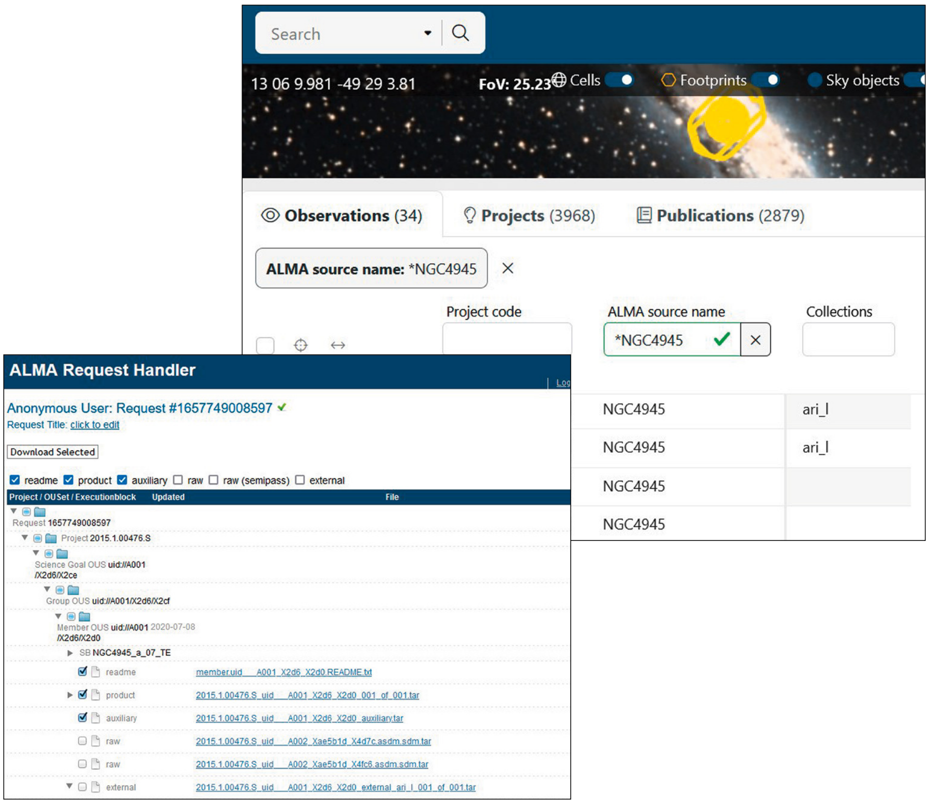
When requesting observations, investigators specify their observational requirements as a series of science goals. In terms of observational operations, these correspond to one or more "Group Observing Unit Sets", each of which may be split into various "Member Observing Unit Sets" (MOUS) that include the instrumental settings needed to reach the investigator's goals. MOUS constitute the selection and analysis level for ARI-L datasets.

To be selected for processing by ARI-L, MOUS must be accessible for public download, with calibration scripts available in the ASA. They must have been observed in modes that could be handled by the imaging tasks of the ALMA Pipeline at the time of the project definition (i.e., excluding solar, full-Stokes, very-long-baseline interferometry, and total power observations).

The ARI-L project uses the Imaging Pipeline outside the range of goals for which it has been commissioned, as it is the best tool currently available to create homogeneous images of data from past cycles.

The ARI-L products are generated by a Python-based workflow engine. For each processable MOUS this starts by downloading from the ASA the data packages (including raw data and calibration scripts and tables and existing products). The workflow then restores the calibration, generates the data products with the Imaging dedicated part of the ALMA Pipeline, extending the processing to include datacubes for observations of calibrators. Finally, the ARI-L image products and calibrated measurement sets are packaged for permanent storage.

All ARI-L images are primary-beam corrected. The corrected images cover a region of diameter equal to the full width half maximum (FWHM) sensitivity of the primary beam centred at the pointing position. The approximate geometric



ALMA source name: *NGC4945 ×

Project code	ALMA source name	Collections
	*NGC4945 ✓ ×	
	NGC4945	ari_l
	NGC4945	ari_l
	NGC4945	
	NGC4945	

ALMA Request Handler

Anonymous User: Request #1657749008597 ✓

Request Title: click to edit

Download Selected

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

Project / OUSet / Executionblock	Updated	File
▼ Request 1657749008597		
▼ Project 2015.1.00476.S		
▼ Science Goal OUS uid://A001/X2d6/X2ce		
▼ Group OUS uid://A001/X2d6/X2cf		
▼ Member OUS uid://A001/X2d6/X2d0	2020-07-08	
▶ SB NGC4945_a_07_TE		
☑ readme		member.uid___A001_X2d6_X2d0.README.txt
▶ ☑ product		2015.1.00476.S_uid___A001_X2d6_X2d0_001_of_001.tar
☑ auxiliary		2015.1.00476.S_uid___A001_X2d6_X2d0_auxiliary.tar
☐ raw		2015.1.00476.S_uid___A002_Xae5b1d_X4d7c.asdm.sdm.tar
☐ raw		2015.1.00476.S_uid___A002_Xae5b1d_X4fc6.asdm.sdm.tar
▼ ☐ external		2015.1.00476.S_uid___A001_X2d6_X2d0_external_ari_l_001_of_001.tar

Figure 2. Snapshots of the ASA query interface showing (top) a search for projects in Cycle 4 (i.e., project code 2016*), with the "ari_l" flags in the "Collections" column, and (bottom) the download interface listing the ARI-L products as "External". Note that single ARI-L images can be accessed with the remote CARTA viewer available for all the ASA images.

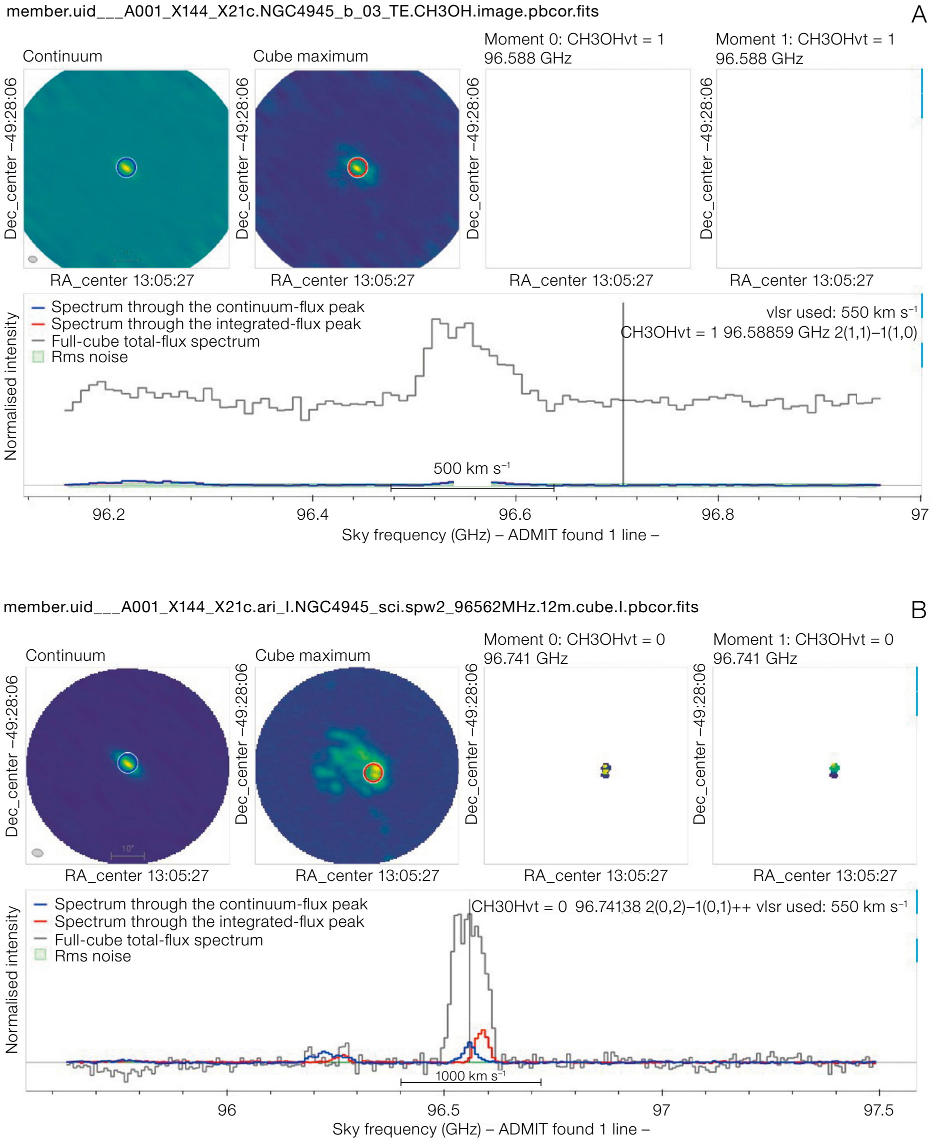

member.uid___A001_X144_X21c.NGC4945_b_03_TE.CH3OH.image.pbcor.fits
A
Continuum
Cube maximum
Moment 0: CH3OHvt = 1
96.588 GHz
Moment 1: CH3OHvt = 1
96.588 GHz
Dec_center −49:28:06
RA_center 13:05:27
Spectrum through the continuum-flux peak
Spectrum through the integrated-flux peak
Full-cube total-flux spectrum
Rms noise
vlsr used: 550 km s^{-1}
CH3OHvt = 1 96.58859 GHz 2(1,1)–1(1,0)
Normalised intensity
500 km s^{-1}
96.2
96.4
96.6
96.8
97
Sky frequency (GHz) – ADMIT found 1 line –
member.uid___A001_X144_X21c.ari_l.NGC4945_sci.spw2_96562MHz.12m.cube.I.pbcor.fits
B
Continuum
Cube maximum
Moment 0: CH3OHvt = 0
96.741 GHz
Moment 1: CH3OHvt = 0
96.741 GHz
Dec_center −49:28:06
RA_center 13:05:27
Spectrum through the continuum-flux peak
Spectrum through the integrated-flux peak
Full-cube total-flux spectrum
Rms noise
CH3OHvt = 0 96.74138 2(0,2)–1(0,1)++ vlsr used: 550 km s^{-1}
Normalised intensity
1000 km s^{-1}
96
96.5
97
97.5
Sky frequency (GHz) – ADMIT found 1 line –

Figure 3. Snapshots of the ASA query interface showing the interactive previews for an MOUS in the project 2015.1.01151.S comparing the CH3OH line in the galaxy NGC 4945 in the QA2 image (panel A) and in the ARI-L image (panel B) and for the CN line in NGC 1068 in the QA2 image (panel C, following page) and in the ARI-L image (panel D, following page).

centre is usually used for mosaics, out to the FWHM of the most distant fields. The Imaging Pipeline by default attempts automasking of the images. Primary-beam and mask maps are delivered together with the product image. For cubes, the channel resolution is defined by the native resolution of the observations but also takes into account that the nominal channel size has been reduced by a factor of two by Hanning smoothing. The Briggs robust parameter is set to 0.5 for all data unless the Imaging Pipeline indicates a different approach.

The ARI-L imaging products undergo a quality assurance step before they are delivered to ESO and then to the JAO for ingestion into the ASA. A README file is enclosed in the ARI-L product folder; its purpose is to trace the history and hierarchy of the dataset used to generate all the product images.

For all the successfully processed MOUS that pass quality control, the final ARI-L calibrated measurement sets are stored in a dedicated storage system outside the ASA that is hosted and maintained by the INAF-IA2[1] facility. The calibrated sets are available to the user community via the ARI-L webpage[2].

A visual representation of the project workflow is shown in Figure 1. Further details of the ARI-L processing and quality procedures are described by Massardi et al. (2021).

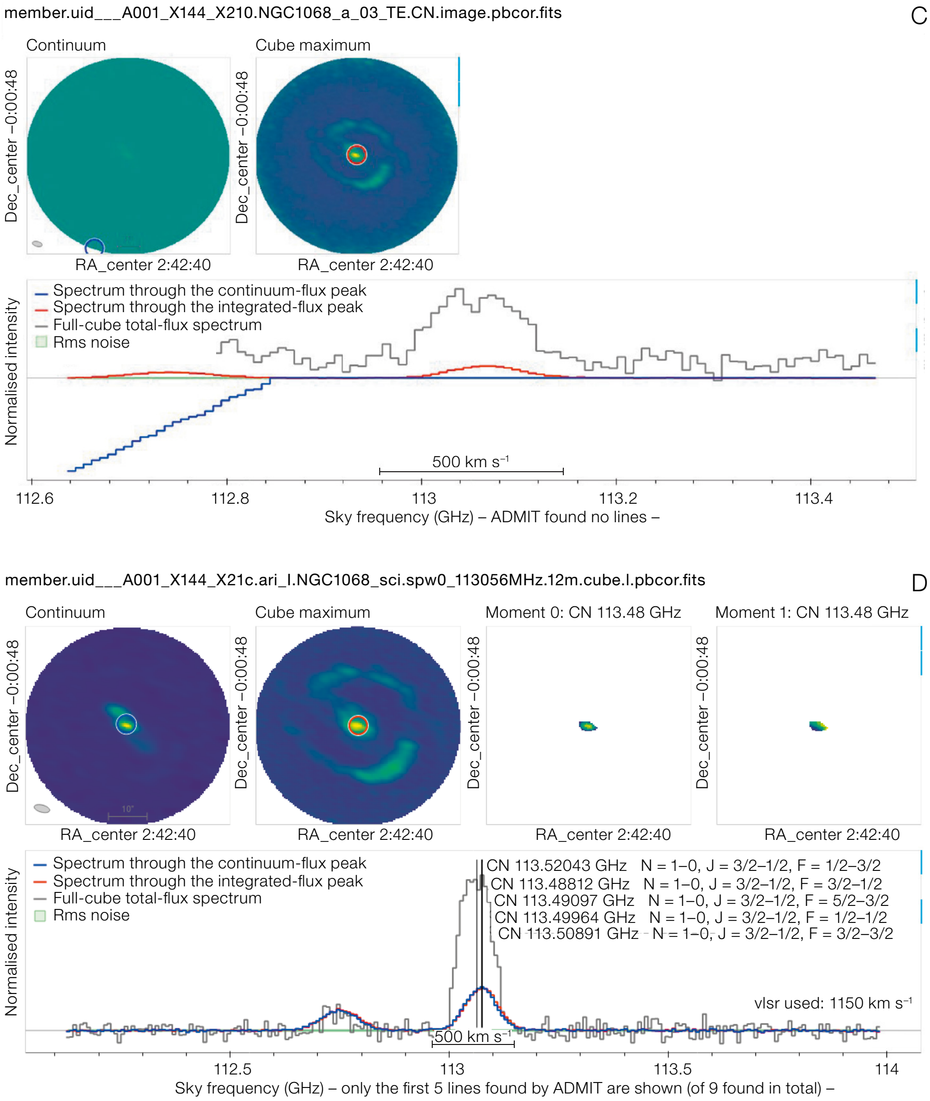
ARI-L successes

Table 1 summarises the statistics of MOUS processed in the ARI-L project. ARI-L products are currently available for 2714 MOUS stored in the ASA. They amount to more than 410 000 files (including images, masks, readme files). 150 247 of the ARI-L archived files are continuum and cube images, of which 84 626 are for science targets and 65 621 are for calibrators. The total number of channels for which there is an ARI-L map in the ASA is 126 707 258.

Images underwent a quality control before being delivered to ASA. Outcomes and logs of the calibration and imaging processing were verified, and completeness and sensitivity and resolution achievement of the imaging products certified. When a failure was encountered, causes were investigated, and, when available, reprocessing solutions were applied to attempt recovery. Only when reprocessing required changes in the calibration scripts available from the archive, or the pipeline application was not suitable for processing, was the MOUS quarantined. An extension of six months has been allocated to the project, starting in June 2022, to continue the effort on such unfortunate instances.

So far 703 331 ARI-L files have been downloaded, meaning that on average each ARI-L file has been downloaded 1.7 times.

Sources belonging to MOUS that have been successfully imaged by ARI-L are flagged “ari_l” in the “Collections” column of the ASA query interface (see Figure 2). It is therefore possible to apply filters to queries. Once a line is, or lines are, selected for download, ARI-L products can be accessed and downloaded as “External products”.



A total of 24 793 preview files of ARI-L fits images/cubes have been looked at in the ASA interface. The ARI-L images made possible the inclusion of a preview in the archive for hundreds of MOUS for which the limited spectral coverage of the QA2 images did not allow it. In particular, ARI-L delivered images of the calibrators — mostly radio loud quasars, local active galactic nucleus cores or supernova remnants — for each of the delivered MOUS; in usual QA2 processing images of the calibrators are not produced.

The use of ALMA data elaborated by ARI-L data should be acknowledged using the standard ALMA acknowledgement statement and it is suggested, and would be appreciated, that Massardi et al. (2021) be cited when ARI-L data are used (see, for example, Di Mascolo et al., 2021; Pantoni et al., 2021; and Van't Hoff et al., 2022).

The ASA legacy enhancement with ARI-L products

ARI-L imaging products are highly relevant for many science cases and significantly enhance the possibilities for exploiting archival data.

Use of the archive has increased over the years, and the ASA is now considered an important, widely-used resource; currently about 28% of all ALMA publications are either based purely on archival data or use archival data in addition to PI data (Stoehr et al., 2022).

ARI-L products facilitate archive access and data usage for science purposes even for non-expert data miners. They provide a homogeneous view of all data for better dataset comparisons and download selections, make the archive more accessible to visualisation and analysis tools, and enable the generation of preview images and plots similar to those possible for subsequent cycles. Furthermore, archive exploitation is valuable for researchers in countries which are not (yet) ALMA partners. This is especially useful for less well-off countries where computing facilities are more restricted. Archive data are also in demand for teaching as they constitute an excellent testbench for data processing and visualisation tools.

Rather than having to download the raw data, identify the Common Astronomy Software Applications (CASA) versions to use, run the calibration or restoration script, and modify and run the imaging script for hours just to determine whether objects or spectral lines have been detected, researchers can use previously-created products to make these types of assessments in a few minutes.

The availability of ARI-L data in the archive allows the visualisation of products using the remote visualisation tool CARTA[3], including the creation of preview images, which can be displayed directly via the ASA query interface. Data products from the ASA are also accessible for automated post-analysis, for example, with the ALMA Data mining ToolkIT (ADMIT[4]; Teuben et al., 2015), or using the Keywords of Astronomical FITS-Images Explorer (KAFE; Burkutean et al., 2018). Figure 3 clearly shows that the completeness of the ARI-L spectral coverage allows the production of more useful previews with better identification of detections and spectral lines. ASA products can also be accessed directly through virtual observatory services that allow for spectral multi-band comparisons, source identifications, and catalogue reconstruction; the ASA offers ALMA data via Table Access Protocol[5], Simple Imaging Protocol[6] and DataLink[7] services.

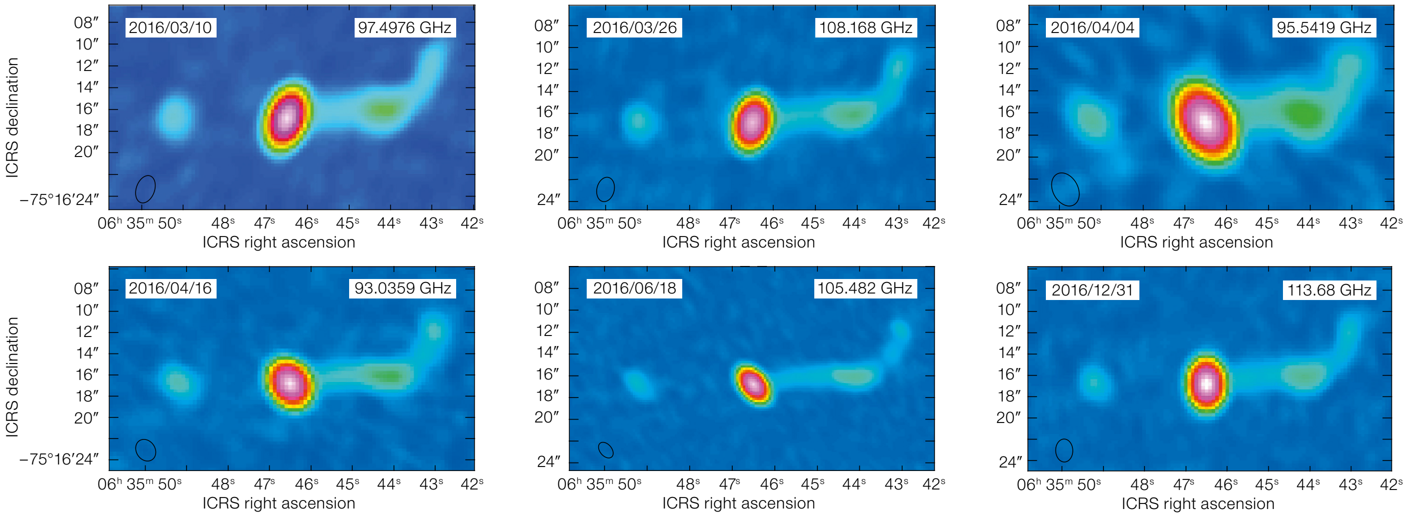


Figure 4. Example of the reconstruction of timelines for calibrator sources. Six epochs of continuum emission are shown, imaged by ARI-L for different archival projects for the calibrator PKS 0635-752. A flare in all the source components is clearly visible in early April, fading over the following months, while a new increase of flux density affects the core component, in December.

ARI-L data strongly enhance the possibilities for comparing archive products in a statistically meaningful way, and for combining archival products, even across different cycles. This is enabled by the enhanced homogeneity given to the products by the use of the imaging pipeline for all of the cycles.

Products can be used to reconstruct timelines to analyse variability (see, for example, Figure 4), to build spectral energy distributions to investigate frequency dependence of emission, or to stack images and statistically enhance signal to noise ratios to obtain average detections. Care is needed to compare products of compatible resolutions and sensitivities, and the scientific significance of initial outcomes may vary from case to case, but they at least provide preliminary insights into what could be reasonably expected (or even improved) with a more detailed analysis of downloaded data.

Finally, the possibility of requesting visibility data sets with the ARI-L calibration applied through the IA2 service makes these measurement sets available without the user's having to install a possibly obsolete CASA version.

Acknowledgements

This paper made use of ALMA Science Archive data. ALMA is a partnership of ESO (representing its Member States), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST 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.

MM acknowledges support from grant PRIN MIUR 2017 - 20173ML3WW 001.

References

Burkutean, S. et al. 2018, JATIS, 4, 028001
Di Mascolo, L. et al. 2021, A&A, 650, A153
Massardi, M. et al. 2021, PASP, 133, 085001
Pantoni, L. et al. 2021, MNRAS, 507, 3998
Petry, D. et al. 2020, The Messenger, 181, 16
Stoehr, F. 2022, The Messenger, 187, 25
Teuben, P. et al. 2015, ASP Conf. Ser., 495, 305
Van't Hoff, M. L. R. et al. 2022, ApJ, 924, 5

Links

[1] INAF-IA2 facility: www.ia2.inaf.it
[2] ARI-L webpage: https://almascience.eso.org/alma-data/aril
[3] CARTA visualisation tool: https://cartavis.github.io
[4] ALMA data mining toolkit (ADMIT): http://admit.astro.umd.edu
[5] ASA Table Access Protocol: https://almascience.eso.org/tap/
[6] ASA Simple Imaging Protocol: https://almascience.eso.org/sia2/
[7] ASA DataLink: https://almascience.eso.org/datalink/

ESO/B. Tafreshi (twanight.org)

In this panoramic image the ALMA Observatory's antennas appear to take in the sight of the Milky Way, arching like a galactic rainbow of dust and stars over the Chajnantor Plateau in the Chilean Andes.