

On-demand radio imaging: access to calibrated data for all astronomers

**R. Beswick^a, S.T. Garrington^a, E. Gonzalez-Solarez^b, P.A. Harrison^c, M. Kettenis^d,
T.W.B. Muxlow^a, A.M.S. Richards^{*a}, H.J. van Langevelde^d, N.A. Walton^b,
N. Winstanley^a**

^aJodrell Bank Observatory, University of Manchester, SK11 9DL, UK.

^bInstitute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.

^cEuropean Southern Observatory, D-85748 Garching-bei-München, Germany

^dJoint Institute for VLBI in Europe, Postbus 2, 7990 AA Dwingeloo, The Netherlands

We present the first publicly available service which allows astronomers to obtain customised radio interferometry images remotely, using the AstroGrid Virtual Observatory interface. This uses parseITongue, a python-based scripting language developed by RadioNet, to extract customised images from calibrated MERLIN visibility data stored at Jodrell Bank Observatory. Similar services are being developed by other archives. We discuss this in the context of older, more limited, data access methods, and future developments.

We thank all members of the AstroGrid (<http://www.astrogrid.org>), MERLIN (<http://www.merlin.ac.uk>) and RadioNet (<http://www.radionet-eu.org>) teams who have contributed to developing the software and services described here.

*The 8th European VLBI Network Symposium on New Developments in VLBI Science and Technology and EVN Users Meeting
September 26-29 2006
Torun, Poland*

*Speaker.

1. Introduction

Radio interferometers typically produce GB to TB of raw correlated data per experiment (including calibration sources and ancillary data such as antenna tables), which need complicated specialised software for data reduction. Recent advances (e.g. the EVN and MERLIN pipelines) allow much of the data reduction to be performed at the correlator. In some cases human judgement is still essential (e.g. flagging) but even if every stage was automated, could all data be reduced to final images as soon as it is observed, obviating the need for further processing either by the proposer or later archive users? No, for several reasons:

1. Interferometry data are multi-dimensional. Possible products include spectra, 3-D data cubes, time series, visibility amplitudes and polarization images, as well as total intensity ‘flat’ images. The data processing paths mostly diverge during visibility data handling.
2. The maximum field of view can be 10^9 pixels and it is usually impractical to make a single image (which has to be mosaiced from facets correctly processed to allow for sky curvature anyway). Instead, the region(s) of interest is (are) usually extracted separately as images of a manageable size.
3. Even within a limited region, the image characteristics can be modified by data weighting and averaging to optimise resolution or sensitivity – but not both simultaneously. Similar considerations apply to other products.

In addition, although a single set of calibration suffices for most present datasets, this is not adequate for very wide-field high-sensitivity data where techniques such as peeling[5] are becoming necessary.

All this makes it desirable to prepare methods which evaluate suitable variables (e.g. data averaging time, phase centre shift) in response to specific requirements. Data can be reduced as far as possible through common stages and then processed as appropriate when requested.

The data products (whatever they are) should be accompanied by layered documentation/history so that a novice can find basic details easily but an expert can get the full history back to raw visibility data. Experts should also be able to revert to the earliest possible processing stage (e.g. a visibility data set with optional extension tables). In our experience, most users (even the original proposers) want to know that this is available but they are happy to start by examining an advanced product (e.g. an image, a calibrated split visibility data set) and only go backwards if there are problems. The data description or metadata should also be provided in a form which complies with VO (Virtual Observatory) standards and models.

2. Traditional approaches

Access to radio data has improved greatly this century. Observatory archives such as MERLIN, the EVN, NRAO, ATCA and the GMRT provide web forms which let you find data by a variety of search terms including position or using the SIMBAD name resolver (rather than obscure experiment codes) and let you download data in a variety of forms. These range from raw or partially calibrated visibilities plus user-friendly pipelines, to quick-look images and other data

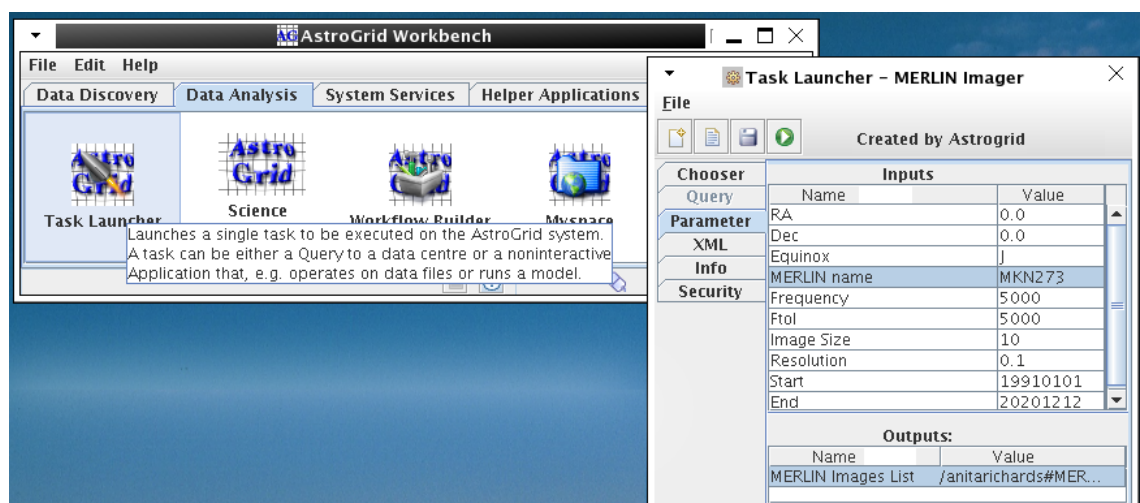


Figure 1: Use of the MERLIN imager. Only the name or position are obligatory; other values default to search the whole archive and produce images at a suitable size and resolution.

products. VO-accessible services have provided specially published data for some years, such as static field-centre images from the MERLIN archive or radio source catalogues cross-matched using SpecFind[10]. Major surveys like WENSS, SUMSS and NVSS are widely used; the latter is the most flexible and allows you to specify the size and resolution (achieved by image-plane smoothing) for cut-out images. However, you still need to use AIPS or a similar suitable package to do much as change the font, let alone the weighting of an image. If you want to compare radio, X-ray and IR images, how do you achieve aperture matching? The first user-triggered pipeline was developed using a VO data model for ATCA, but only as a prototype.

3. The MERLINImager

The MERLINImager is a service available via the AstroGrid workbench (Fig. 1). You fill in a simple form specifying an object or sky region and, optionally, the resolution, size, frequency and date – all parameters devoid of interferometry jargon. The AstroGrid Common Execution Architecture[1] provides the software environment to send the request to the MERLIN archive server at Jodrell Bank. Here, visibility data with calibration tables is stored inside the AIPS environment, on a RAID array.

The link between the AstroGrid web services (which can control any suitable resource) and the local specialised service is provided by *parseITongue*[2], a python-based scripting language developed by RadioNet. Python scripts control a database query to identify which visibility data could produce images meeting your criteria and rank the data sets by overall quality and by proximity to the centre of the field of view. If the data cannot support the resolution or image size requested the nearest approximation is used. The *parseITongue* pipeline processes calibrated visibility data to produce up to 5 images per request. The field centre is always imaged to avoid confusion. The images are written as FITS to the MERLIN server and their web locations (URLs) and basic metadata (position, pixel size, size etc.) are recorded in a VOTable[6] (an XML-based format) which complies with the Simple Image Access Protocol[9]. The table can be downloaded and viewed in

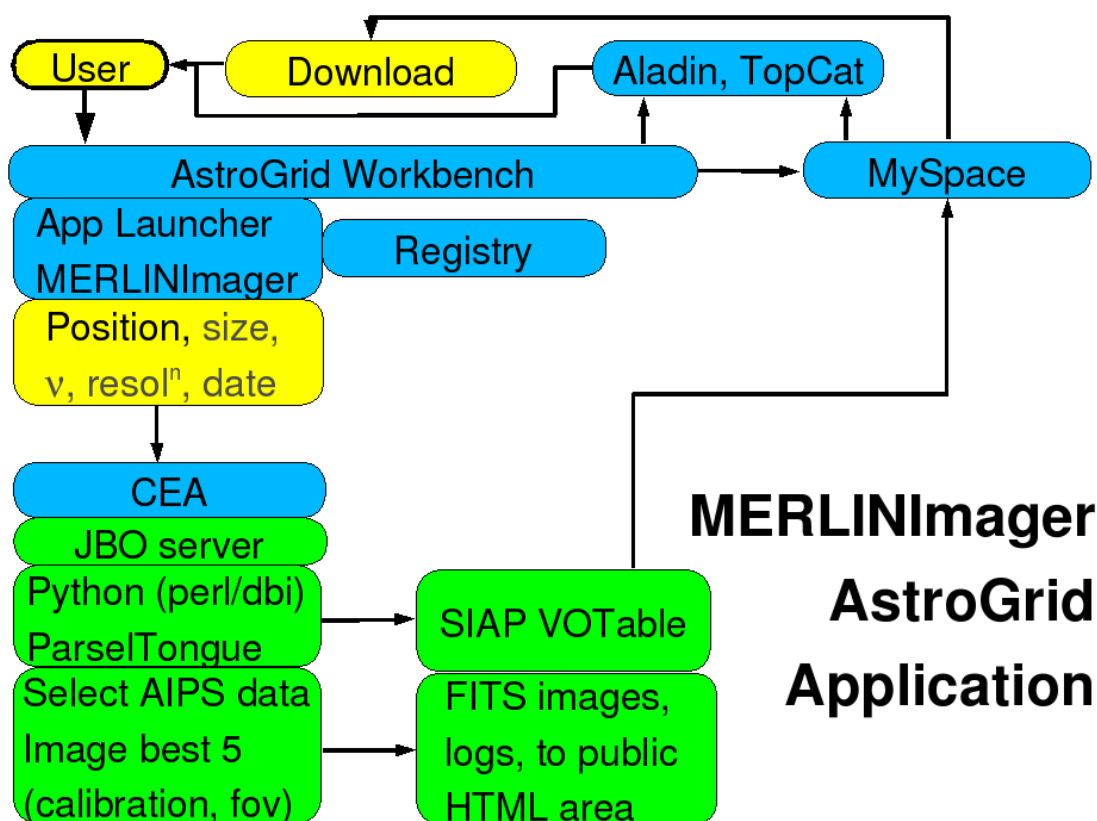


Figure 2: MERLINImager flow chart. User inputs are shown in yellow; steps shown in blue are handled by the AstroGrid system and stages in green take place at the MERLIN archive.

any suitable tool (e.g. TopCat[8]), allowing the images to be saved to your desktop, but in addition, VO-compatible visualisation tools like Aladin can recognise the URLs and display the images directly. This interoperability is possible thanks to the PLASTIC[7] protocol.

Fig. 2 shows the various stages and Fig 3 shows the final images compared with an HST image. All data processing takes place at the archive where the final products are stored until requested. Using the AstroGrid infrastructure means that there are no browser time-out issues even for very long-running requests, in fact you can log out and switch off your personal computer and retrieve the results any time at your convenience. You still need to understand interferometry at the level of being wary of missing spatial frequencies, but not the jargon of AIPS etc.

These VO services give access to radio interferometry images without having to transport many-GB visibility data sets. The output images can be passed automatically to further stages in a workflow (e.g. a source extractor). A similar service is being developed to provide cutouts (at any suitable resolution) of MERLIN+VLA HDF(N) images.

4. Next-generation interferometry requirements

We developed the MERLINImager first because many users regard images as the most basic representation of the sky and there are many other VO services to manipulate them. The service

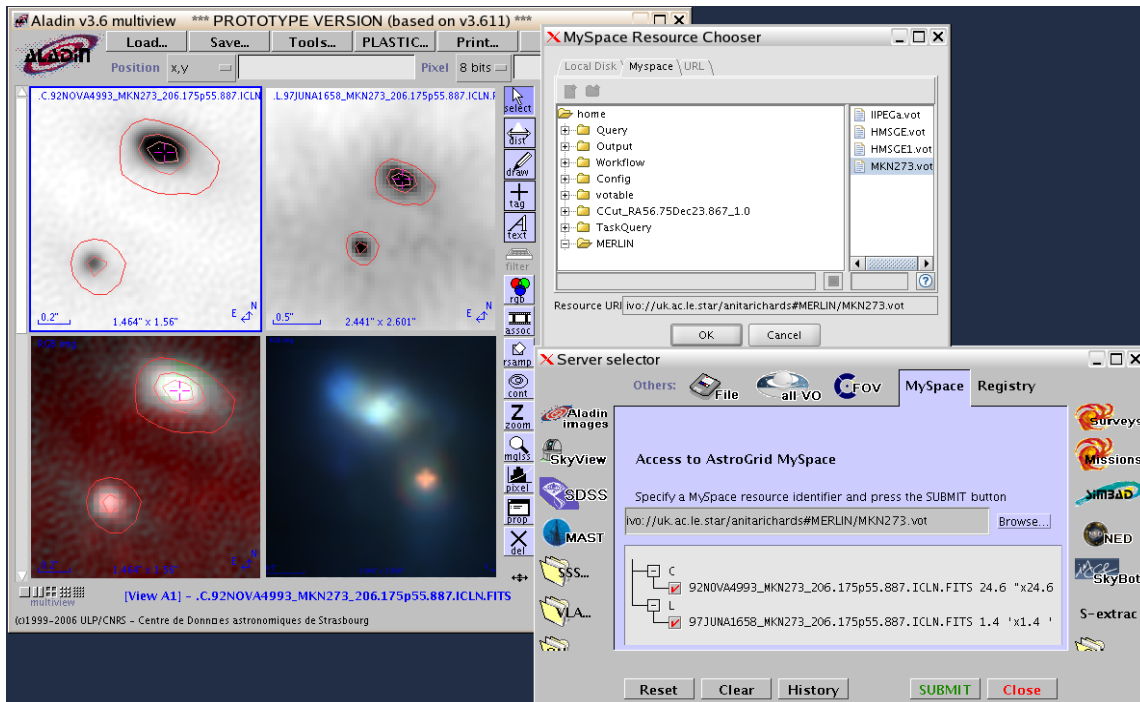


Figure 3: A single simple query gives two images of Markarian 273 at different frequencies but matching resolution (upper row; 6 cm on left, 18 cm on right) which Aladin can turn into a spectral index map (lower left, redder shade shows longer wavelength dominates), eliminating the SE component as an AGN candidate. Images from any other VO-aware archive can be obtained for comparison and easily aligned (HST, lower right).

can be extended to provide (reversibly) calibrated visibility data, (which is already available via the MERLIN web site) and other data derived from visibilities such as a ‘light’ curve. Some limitations need to be overcome for heavy use, such as the maximum number of simultaneous AIPS processes. In future, spectral line data will be added to the MERLIN archive so that spectra and data cubes will become available. Spectra can be manipulated using VO tools such as VOSpec and SPLAT and Aladin can display cubes as a movie (see links from PLASTIC[7]). Other possibilities include VO access to source extraction, constructing moment and spectral index maps and so on using specialised radio astronomy software.

Similar developments are taking place at most other existing radio archives such as JIVE, the VLA and single dish facilities. In many cases these also use *parseTongue* or other python-based scripting languages (e.g. CASA[4], being developed at NRAO) suggesting a good prospect of close interoperability at the user interface whether the raw data is reduced using AIPS, *aips++*, or any other package. One possibility which would be very popular with users is a service for combining visibility data from different arrays to improve image (and other) sensitivity over a wide range of spatial scales. It would also allow domain-specific software to be used for each data-set e.g. in aperture-matching for constructing a radio-to-X-ray spectral energy distribution.

Next-generation interferometers such as ALMA, *e*-MERLIN, the EVLA, LOFAR and SKA are committed to implementing Virtual Observatory access. This is not simply a means of serendipi-

dous access to ‘old’ data but an integral part of ‘beginning to end’ data processing; at each stage from proposing and scheduling to archiving and retrieving metadata are recorded to facilitate the next step. VO components can be used for internal as well as public data management including (in the near future) authentication of users with rights to restricted access data.

There are outstanding issues for both radio and VO communities. At present (2006) one of the biggest obstacles to efficient use of VOs is that although a very large amount of data are becoming available, it is rarely fully described even at a coarse level of spatial, spectral and temporal coverage. This hampers direct searches for data at a given position and frequency without first specifying the archive or catalogue to be queried. As well as agreeing data models, the International Virtual Observatory Alliance[3] needs to develop and promote easy ways for data curators to provide at least basic metadata.

There are not yet adequate standards for quantities like polarization, but these will emerge once there are VO-enabled tools to do science as well as the data. More fundamentally, most tools assume that the fundamental quantum of an image is the pixel and may produce incorrect results if flux densities are in Jy/beam. These are solvable problems; all domains face similar issues (e.g. conversion of optical magnitudes to physical units) and experience shows that they are overcome if there are strong science cases driving the technological development.

References

- [1] P. Harrison, *A Proposal for a Common Execution Architecture 1.20*, 2005,
<http://www.ivoa.net/Documents/latest/CEA.html>
- [2] M. Kettenis, *parselTongue*, <http://www.radionet-eu.org/rnwiki/ParselTongue>
- [3] IVOA Documents <http://www.ivoa.net/Documents/>
- [4] NRAO, ATNF & ASTRON groups, *Common Astronomy Software Applications*
<http://aips2.nrao.edu/daily/docs/casa.html>
- [5] T. Oosterloo, *Peeling*, 2005
http://www.skatelescope.org/pages/news/SKA_WFI2005/oosterloo.ppt
- [6] F. Ochsenbein et al., *VOTable Format Definition 1.1*, 2004,
<http://www.ivoa.net/Documents/latest/VOT.html>
- [7] J. Taylor et al., *PLASTIC (PLatform for Astronomical Tool InterConnection)*, 2006,
<http://plastic.sourceforge.net/>
- [8] M. Taylor, *TOPCAT (Tool for Operations on Catalogues And Tables 2.2)*, 2006,
<http://www.starlink.ac.uk/topcat>
- [9] D. Tody & R. Plante, *Simple Image Access Protocol 1.0*, 2004,
<http://www.ivoa.net/Documents/latest/SIA.html>
- [10] B. Vollmer, E. Davoust, P. Dubois, F. Genova, F. Ochsenbein, W. van Driel, *SPECFIND: A method for determining radio continuum spectra and its application to large surveys.*, 2005, *A&A*, 432, 1177