

Virtual Observing for the SKA

A.M.S. Richards and S.T. Garrington

*AVO/MERLIN/VLBI National Facility, Jodrell Bank Observatory,
University of Manchester, Cheshire SK11 9DL, UK.*

C. Reynolds

JIVE, Postbus 2, 7990 AA Dwingeloo, the Netherlands.

M.G. Allen

CDS, Observatoire de Strasbourg, Strasbourg 67000, France.

Abstract. The Square Kilometre Array will produce Tb – Pb of raw data per experiment. Some projects will involve large collaborations; many will contain serendipitous sources of interest to astronomers other than the proposers. By the time the SKA is in operation, virtual observatories will be well established under the coordination of the International Virtual Observatory Alliance (IVOA) and could offer access to SKA data reduction and results without onerous travel or massive data transfers, especially if VO compatibility is built in to the SKA from the start.

1. Access to SKA data

A typical SKA image will involve 100 Tb of data. Continuum observations are likely to be made in multi-channel mode to allow the subtraction of confusing sources. Short integration times are needed for wide-field imaging. D’Addario (2002) estimated the output of the correlator at 215 Gb per dump to archive for an array of many small dishes. The data rates from various SKA designs were compared by Carilli (2002), showing that the number of correlations (adjusted for 4096 channels, per integration) could be reduced by phasing-up groups of antennas, giving ~ 4 Gb; using large dishes and focal plane arrays (e.g. KARST) would give ~ 0.5 Gb. This implies a minimum of 50 Tb of data per hour for relatively long integrations, or a factor of 1000 higher for some array designs, spectral observations or complex (e.g. multi-target) modes. The data load could be reduced by extensive post-correlation processing, only storing target images or other products specified by the PI. However in most cases it is essential to retain the visibility data as there are likely to be serendipitous sources of interest in many individual fields of view. There are also many incompatible ways to image the same field, to optimise sensitivity, to optimise resolution, to extract spectral index information and so on. The archive – including fields containing calibration sources – will build up to an all-sky survey.

SKA data reduction will be far more complex than any current operation, especially if multiple experiments are performed simultaneously. Calibration and confusing sources may be extracted or extrapolated on-the-fly from recent SKA data, to match the relevant frequency and resolution. SKA analysts will want diagnostic results during the run. The first requirement of most SKA observers will be preliminary images (or equivalent), delivered to their desktop, without having to wrestle with unfamiliar software. They may want little else except to tweak plots for publication – or they may want to re-reduce the data from an early stage. At present most radio observers want control over at least the imaging process, for example to produce a series of comparable results at different epochs. Someone accessing archive data may want different parameters (for example a displaced phase centre or different time-averaging intervals) from the original observer.

Having got their first SKA results, astronomers want to make comparisons across epochs, classes of sources, the entire electro-magnetic spectrum, using models derived from anything from chemical databases to cosmological theories. If they cannot find what they want, they schedule a new SKA experiment; access to the best existing data on the region of interest will enable them to estimate the observing time and other constraints. This implies an integrated set of tools for proposal generation, scheduling, on-the-fly and off-line calibration and conversion to storage format, and flexible final products. Some levels will be used by SKA operators whilst others will be accessible to any user, with the option of a simplified interface providing the most popular functions. Every stage should be self-documenting and reversible where possible.

We might encourage new users to attend the SKA or a specialised centre and become familiar with its facilities. However the goal of linking VO technology with observatory archives is to enable astronomers to spend less time travelling and reducing data and more time interpreting it. In general full-time operators will schedule the SKA. Observers may occasionally monitor observations, such as exceptionally high data-rate PSR observations, where more processing is performed on-line and less data are archived, but real-time control can be supervised remotely using the internet (as with current Arecibo observing).

Data storage is becoming cheaper and cheaper (although durability needs investigating) and processor speed is increasing, but the timescale for doubling in speed of internet links is twice as long as the doubling timescale for localised performance (Quinn 2000), hence the VO slogan ‘move the results, not the people, not the data’. The archive and the specialised software are held in one (or a few) places, expertly maintained, and accessed via web interfaces.

2. Remote access to radio interferometry

The major international radio interferometry facilities are all developing electronically accessible archives. The basic service is an on-line searchable list of observed targets and calibration sources. A variety of additional features are being developed by the major international arrays (ATCA, EVN, MERLIN, VL(B)A, WSRT). The NRAO e2e project is well-documented (e.g. Benson *et al.* 2002). This is based on AIPS++, used as a storage format (measurement tables) as well as providing data processing tools.

MERLIN+VLA visibility data

Observation and Processing Details

Source Name	HE
RA (J2000.0)	12
Dec (J2000.0)	62
Proposal Code	97A16
PI Name	T.Muxlow
PI Email	tmuxlow@jodrellbank.ac.uk
PI Institute	Jodrell Bank Observatory UK
Proposal Title	MERLIN & VLA Observations Of The HST Deep Field
No. visibilities	5672150
Obs. type	Target
Source comment	MERLIN Era
Associated Phase ref. source	12334688
Processing Block	87HDFN1420
Observations Between	19980203 and 19970427
Frequency	1412.0 MHz
Channels	32 x 500 kHz
Processing Block Comments	8 antennas (De Ca Kr Wa Da Mi Lo Te). Some data with no Lo nor Wa.
Data processing script	RUNFILE VERSION.PL
Additional notes	L BAND NOTES PLOT NOTES POLARISATION NOTES

These data are pending release

Remote Imaging

Use the sector below to generate postage stamp images from the visibility data directly (you'll browse cache if you run this more than once in one session).

Offset field 1 (J2000 hh:mm:ss.sss dd:mm:ss.s)

RA position:

Dec position:

If this box is checked the central bias will also be mapped (necessary if there is a source there; Central Fields (J2000 hh:mm:ss.sss dd:mm:ss.s))

RA center:

Dec center:

Imaging center:

Sub-image size (arcsec):

Resolution (arcsec):

MERLIN Remote Imaging: HST-FIELD

Offset field at 12 36 56.000 +62 12 07.000 (J2000, ICRF)

Simple inputs

Returns selected image

Compare: best map over HST image

Figure 1. The MERLIN+VLA observations of the Hubble Deep and Flanking Fields are several Gb of uv data, producing $> 10^8$ image pixels. A small region has been automatically remotely imaged on demand from the calibrated archive data. The inset shows the hand-made image of this source (Muxlow et al. 2003, in prep).

The MERLIN archive now contains uv data with basic calibration applied, preliminary images and other plots. This can also be accessed via the Vizier and Aladin services provided by CDS. The next stage (currently a prototype) is to allow the user to re-image calibrated uv data at the position and resolution they desire (within the data constraints). This is performed by simple inputs to a web form (Fig. 1), which are interpreted as AIPS commands and return a plot and FITS image data. This process could also be used to produce a radio light curve or any other product. The uv data storage and all processing occur at Jodrell Bank.

After EVN data are correlated at JIVE they are edited and calibrated using a pipeline (Reynolds et al. 2002), with some human quality control. This involves retrieving electronic data such as system temperatures from the individual stations, which are archived in Bologna, and generating email information for the PI; data processing uses AIPS++ and AIPS. At present, the PI receives tapes of FITS data by post, but the processing history, calibration tables and image plots are available on the web. The data and calibration tables are archived and calibration source results are used for network monitoring. Information management is seen as highly important, with the aim of providing details as diverse as correlator parameters and publications arising from an experiment in compatible formats in the same (virtual) place. Soon the FITS data will be available from the same server, involving different time-dependent restrictions on access to target and calibrator data from the same experiment. In parallel to this, the information content in the archive of previous EVN observations is being expanded.

Experiences in creating the MERLIN and EVN archives suggest that it is vital to integrate data documentation, processing and retrieval into same architecture as scheduling and monitoring observations. We have also learnt that even the PI, let alone people retrieving archive data, need unpredictable products – this will be even more true of SKA. It is necessary to store data in a form which can be flexibly reprocessed, e.g. calibrated, multi-channel *uv* data.

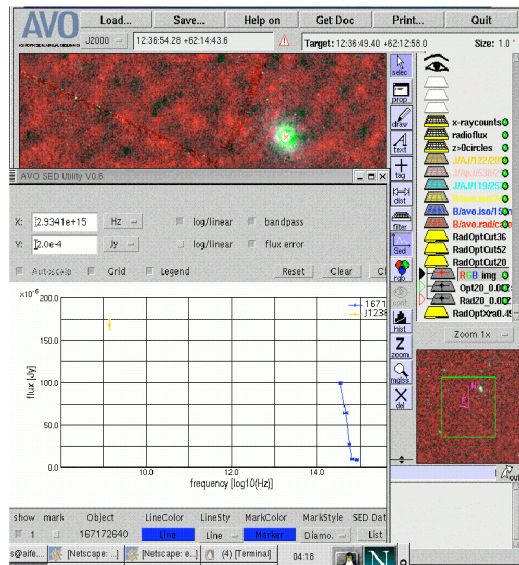


Figure 2. Screenshot of the AVO Aladin prototype, showing a false-colour composite (radio, optical, x-ray) image and spectral energy distribution plot, both created on-the-fly using the Strasbourg CDS server and data and catalogues anywhere in the world.

3. Nascent virtual observatories

The International Virtual Observatory Alliance was launched in June 2002 at a conference in Garching attended by about 200 representatives of national and international VO projects. AstroGrid will provide access to UK data centres and is pioneering distributed computing and user-painless authentication. The European VO (AVO) is building on the strengths of CDS giving access to very heterogeneous data. The ‘First Light’ held at Jodrell Bank Observatory (Clarke 2003) used GOODS data (Hubble Deep Field North and *CHANDRA* Deep Field South) to demonstrate real-time combination of data servers, catalogues and images located in the UK, France and Germany. Images from the *HST*, ESO telescopes, MERLIN, the VLA and *CHANDRA* were used, on a range of pixel scales, in heterogeneous units. The capabilities include colour composition, spectral energy distribution plotting and even re-extraction of sources from FITS data according to user-defined criteria. Enhanced versions of the familiar Aladin and Vizier tools are employed, see Fig. 2.

4. Planning the SKA

Data delivery and reduction needs to be integrated at the earliest stages of planning. The correlator load is greater for some designs such as many small dishes. These are already recognised as fundamental design constraints, to be solved by dedicated technology (Carlson 2000). However, current developments suggest not all (nor most) astronomers can be at the end of the fastest possible links with all known software (and expertise) on their desktop or even those of their students. We do not want effective use of SKA to be limited to the existing radio community. This means a few data reduction centres around the world should have high-speed links to acquire SKA data, and supply reduced data of manageable size to users. Each centre should hold – and be capable of maintaining – the necessary software, although whether these operate as spokes of a hub which performs all initial processing, or as nodes sharing the workload, is a detail for future determination. Many users will be happy with pipelined data with small changes which can be carried out locally or by remote access to a processing centre. They should always have the options, however, of accessing data at any stage of processing, and re-reducing it. This implies that each stage of pipeline data processing should be reversible (such as by the use of calibration tables) and generate a full history.

For example, you might have a GMRT image and want to zoom in by combining it with SKA data; a linked network of VOs would provide:

- Linked Registries which store descriptions of array archive contents (sky, frequency and temporal coverage, range of resolutions provided etc.) and access methods, in globally compatible formats;
- One-step authentication to decide if you have access to data which are not completely public domain;
- A choice of retrieving (reversibly) calibrated and edited visibility data from each array, or of uploading the data to temporary personal scratch space at a remote processing centre;
- Wrappers to any special remote software/data format conversion packages, so that you can request appropriate data using universally understood parameters like position, resolution frequency, sensitivity, rather than esoteric array-specific details, and also (optionally) have calibration, averaging, imaging etc. performed before retrieval;
- Access to software maintained at a few central locations, via an interface which allows a range of user intervention in data processing, from fine control for experts, to sensible defaults for speedy processing to produce an image at the requested place and resolution.

In every case the data should have complete astrometric and photometric accountability; and at the earliest practical stage the data products should be in a universally recognised format. Products may be single images, mosaics, spectral index or polarization-related images, spectra and spectral cubes, radio ‘light’-curves at selected positions, visibility data for model fitting, pulsar timing data, and much else from some of the more imaginative designs.

We can help make sure this happens by not only influencing the development of the SKA, but also of VOs. As an international collaboration, the SKA is likely to push for and use global standards in interferometry data reduction software and formats. The VLBI link suggests it will set the standards for astrometry. We do need to involve non-radio astronomers to make sure we are aware of the needs of those who do not get a kick out of Fourier transforms.

We also need to make sure VOs can give access to Fourier data, handle products other than images, and make truly multi-wavelength comparisons with data from sources as diverse as X-ray satellites, chemical reaction-rate databases or cosmological simulations. We can ensure this by maintaining links with the existing projects and starting to develop data models for SKA products. A data model (McDowell et al. 2002; Hanisch et al. 2003) is a description of the product and all the essential background, in a common language intellegible to the VO. For example, in order to describe the resolution range of the SKA you need to know the antenna positions, frequency range, integration time etc. using universally understood *metadata* such as the CDS system of UCDs (Unified Content Descriptors). A standard xml-based data format, VOTable, has already been agreed and an XML wrapper will be developed for FITS data. Allen (2003) outlines immediate goals to enable archive interoperability.

Acknowledgments. This style file is based on one provided by PASP.

References

Useful links:

AVO www.euro-vo.org; documents www.euro-vo.org/pub/articles/index.html;
www.euro-vo.org/twiki/bin/view/Avo/AvoPubArts
 EVN archive www.ira.cnr.it/tventuri/cata.html
 IVOA www.ivoa.net and discussion forums e.g. www.ivoa.net/forum/radiovo/
 MERLIN Archive www.merlin.ac.uk/archive
 Metadata and UCDs cdsweb.u-strasbg.fr/interoperability/Standards.htx
 and IVOA forums
 VOTable cdsweb.u-strasbg.fr/doc/VOTable

Allen, M.G. et al., 2003 in *Toward an International Virtual Observatory, 2002*, Garching, Germany

Benson, J., Waters, B., Cornwell, T., 2002, www.nrao.edu/e2e

Carilli, C.L., 2002, SKA Memo #24

Carlson, B.R., 2000, www.drao-ofr.hia-ih.nrc-cnrc.gc.ca/science/ska/#documents

Clarke, T., 2003, *Nature*, www.nature.com/nsu/030113/030113-11.html

D'Addario, L.R., 2002, SKA Memo #25

Hanisch, R. et al. 2003, www.us-vo.org/publications.html

McDowell, J. et al., 2002, www.us-vo.org/publications.html

Quinn, P., 2000, ESO, www.euro-vo.org/pub/articles/AVO-IT-paper2.html

Reynolds, C., Paragi, Z., Garrett, M.A., 2002, URSI, www.evbi.org/pipeline/user_expts.html