January 25, 1992

UV Data System and Processing Requirements - An Initial Distillation

AIPS++ UV Data System, Calibration and Imaging Group[1]

1 Introduction

This document is an inital attempt to integrate the *data system, calibration and imaging* requirements for AIPS++, based on the requirements from the individual consortium members. However, this is by no means a complete requirements specification for the data system; decisions by the *image handling* and *user interface* groups are likely to have implications for data system requirements. In addition there will be some overlap, in that code which is developed by the *image handling* and *fundamental library* groups will be required by the *data system, calibration and imaging* group.

This document may have to be revised to incorporate GMRT requirements.

2 General Nature of the Data

This section includes requirements which will affect the data system explicitly, or implicitly because of the nature of the observed data.

2.1 Instrumental Requirements

A number of requirements are generally applicable to single dish and interferometer data:

- A full measurement of the electromagnetic field requires four complex (polarisation) parameters, which can be converted to Stokes parameters; support for this should be fundamental.
- Polarisation measurements may be time switched if all four polarisation parameters are not correlated simultaneously.
- Multiple frequency bands may be simultaneously observed (*e.g.* for observing multiple lines simultaneously or multi-frequency synthesis), with variable numbers of channels within each band.
- The frequency axis may be non-linear (*e.g.* as produced by acousto-optical spectrometers) and time variable.
- Data for combination from different observations may have different (but overlapping) numbers of spectral channels and channel widths. These may need to be accomodated within the one dataset.
- Focal plane arrays with arbitrary geometry (*e.g.* field rotation through the observation) characteristics must be supported.
- Mosaiced observations may have many (∼1000) pointing centres. This must be supported at both the *uv* dataset and image dataset level.

[1]Current membership: C.Flatters, M.Holdaway, R.Sault, D.L.Shone

- The measured polarisations, frequencies and pointing centres may be rapidly time switched (e.g. these may change every integration).
- Time-series data of profiles and visibilities (e.g. pulsar data with bin number as a data axis) needs to be supported.
- Error measures or estimates (*e.g.* weights) should be regarded as standard in an observation.
- Correlation data may be 16-bit integers or 32-bit floating point.

2.1.1 Requirements specific to single dish observations

- Total power time series.
- Total power measurements, irregulary spaced in space and/or time for imaging.
- Series of spectra, including the case where the size of the spectrum is variable.
- Bit-field data for raw pulsar data.
- Interferometer autocorrelation data.

2.1.2 Requirements specific to interferometry

Interferometer arrays should be regarded as being generally inhomogenous (although this should not preclude taking advantage of homogeneity where possible) in a number of ways which should be handled transparently whenever possible:

- Polarisation characteristics (*e.g.* linear or circular feeds and equatorial or alt-azimuth mounts) differ among antennas or arrays of antennas.
- Antenna sizes and system temperatures may differ widely.
- The procedures and input data for *a priori* antenna calibration may vary from antenna to antenna.
- Space VLBI will require support for variable antenna positions.
- Integrations times may vary from baseline to baseline.

2.1.3 VLBI

VLBI support is implicit in many requirements, but in addition, support must be provided for the many correlator formats which will prevail, including MkII, S-2 and K-4, as well as the MkIII and VLBA modes.

The data system should support the merging of correlation data and calibration data from different correlators and allow the user to deal with duplicate correlations.

2.1.4 Monitor/history data and site/instrument specific data

There should be support for history associated with data. It should be possible to import and make use of history information from "previous incarnations", such as on-line observing logs.

- Monitor and meterological data need to go directly into dataset.
- Easy access to observe-time information which cannot be put into the dataset at the time of observation.
- Flexible addition to the dataset of observatory dependent monitor data.

2.1.5 Other observed data types

It is important that the data system be as extensible as possible to support new data data types in future:

- Triple correlation data, including the case where one of the visibilities has a different frequency.
- Optical interferometer data.
- Images; at some level, the data system should support the requirements of image handling, rather than having one data system for images and another for all other data. In general, "vector images" should be supported, allowing scope for complex images, with associated errors *etc.*, as well as double precision images.

2.2 General (Non-Instrumental) Data-System Requirements

- The data system file format(s) should be accessible from all supported machines without conversion.
- There should be no distinction between "multi-source" and "single-source" files.
- Applications should function on data and data-cubes in arbitrary sort order; all sorting should be hidden to the user.
- Coordinate handling must be very general:
  - Support for data with non-regular increments in coordinate axes (such as an optical velocity axis).
  - Coordinate systems and ephemeris information required for astrometry, and near field imaging will be supported.

  This will also assist in the merging of data sets.
- Support for data errors should be fundamental to the data system, providing a basis for support of data error handling in a variety of applications:
  - Easy generation of error models;
  - Error propagation through a series of tasks;
  - Error images associated with astronomical images;

  - Plotting error bars on spectra and profiles;
  - Automatic warning if contour levels are below noise;
  - Error-based blanking in display of results;
  - Properly formatted errors when data are extracted for tabulation.

- Simultaneous processing of "associated" datasets, such as different (*e.g.* by calibration, integration time, fringe fitting *etc.*) versions of the same observation, so that the best can be selected later.

- A processing history should be maintained for each data set that can be reviewed by the astronomer.

- The user should have access to both data values and the 'header' values that govern the interpretation of the data.

3 Data Display and Editing

Users must be able to display and interact with data in ways which are much more flexible and effective than in AIPS. This has long been regarded as vital to system with few baslines, such as VLBI, but is also likely to prove useful in processing other forms of data. Data display and editing should be seen as generic tools, applicable to single dish and other data as well as interferometer visibilities. To this end, some degree of data visualisation should perhaps be seen as an integral, or at least closely coupled part of the data system.

Many "viewing strategies" are desirable for display and interactive flagging of data:

- Baseline by baseline display and editing (including multiple, simultaneous baselines).

- Display of data aggregated in various ways (*e.g.* averaged over a number of spectral channels).

- Interactive selection of data to be displayed:
  - taking arbitrary cuts (*e.g.* circular, radial or a user-defined locus) through the *uv*-plane;
  - setting windows in space and/or time;
  - expanding aggregates (*e.g.* pointing and clicking on an averaged multi-channel visibility to show the component spectrum).

- Display of generic model data (*i.e.* from CLEAN components, gaussian fitting *etc.*) with observed and/or processed data, and display of data with model subtracted or divided.

- Flagging should be reversible, with the ability to store flagging information and apply this on the fly.

- Flagging on the basis of monitor/observing log data.

- Flagging from "consistency check" information, in particular redundancy, where possible, or crossing-points in the *uv*-plane.

- "Intelligent" automated flagging for large datasets.

4 Calibration and Associated Processing

- Calibration, like flagging, should be reversible, with the ability to store calibration information and apply it "on-the-fly". However, it should also be possible to apply calibration information "once-and-for-all", creating a new, calibrated data-set.
- Calibration should be made as generic as possible; site-specific/instrument dependent code should be kept to a minimum.
- Calibration of data should be possible from derived tables of instrumental parameters such as system temperature and gain *vs.* elevation. It should be possible to derive such tables from calibration observations.
- The calibration process should include flexible interpolation and averaging of calibration data under the control of the user.
- Redundancy (possibly including crossing points) should be used whenever possible, as an additional constraint on calibration and self-calibration.
- Model fitting should be possible in both the image and *uv*-planes, and it should be possible to use the resultant model in the same way as a CLEAN component model in calibration and self-calibration.
- Fringe fitting (normally only for VLBI data) should be possible by baseline, as well as globally.
- Cross-calibration of different instruments should be possible, (*e.g.* flux scale, pointing) particularly where data from different arrays are to be combined.

5 Synthesis Imaging

The title of this section might be construed as covering simply the act of transforming edited, calibrated data and deconvolving the images. However, we wish to emphasise that the process should be viewed as a whole, and whilst data editing, for example, is discussed separately (because it is a generic operation, applicable beyond interferometry), it must be possible to integrate (self-)calibration, data inspection/editing, transformation and deconvolution more closely than is possible in AIPS.

It should be possible to easily "mix-and-match" self-calibration, transformation, and deconvolution tools, for example, using CLEAN to deconvolve in the early stages, and Maximum Entropy later on when CLEAN would begin to break-down. This also demonstrates the need to make self-calibration use a generic model, which may be CLEAN-components, an image, or a gaussian model.

6 User Access to UV Data

There need to be flexible and simple ways for the users to manipulate their uv datasets. For example:

- The user should be capable of using the host systems capabilities and utilities to manage his datasets. The datasets should use a normal file name, and the user should be able to use the host directory hierarchy to best organise the data.

- Tasks should generally take multiple input uv datasets where this makes sense. For example, the map making program should be capable of taking multiple input datasets, all of which contribute to the output images.
- There needs to be a flexible way for the user to select the particular subset of data, in a dataset, to be processed. As well as selection based on time, antenna number, frequency, etc, it should be possible to select based on the values of other data (including monitor data).

There need to be methods for general users to perform unusual or different processing and analysis on their uv data. In particular:

- It should be possible to extract a subset of data from a dataset, manipulate it in some powerful (e.g. IDL-like) command language. This would include displaying the data and optionally replacing it in the dataset.
- It should be possible to extract a subset of data from a dataset in a variety of formats (eg. FITS or plain text) in order to transfer the data to other programs or packages. It should also be possible to read the modified data back into AIPS++ using the same formats.
- It should be possible to import general astronomical data (eg. catalogues) into AIPS++.
- For applications where the built-in tasks and command language features are insufficient, there needs to a program interface to allow the casual programmer reasonable access to the data. Some flexibility and efficiency can be sacrificed in making this interface comparatively simple. FORTRAN programmers should be supported.