AT/25.1.1/003

A SOURCE CATALOGUE FOR THE AT

Ray Norris

17 Sept. 1984

SUMMARY

The AT's requirements for a source catalogue are listed and considered, and a system of catalogues is proposed. This system consists of an astrophysical source catalogue (a file of ASCII records containing data on astrophysical sources) and a separate calibration catalogue (a set of binary files). Comments and criticisms of this draft are invited from all quarters. In particular, it would be helpful to know now if any additional information needs to be stored in the observation file.

1.0 INTRODUCTION

1.1 Purpose And Scope Of This Report

This report describes a draft specification for a source catalogue for the AT. This catalogue is composed of a system of several files which will be of use for (a) cataloguing information about sources observed with the AT, and (b) maintaining records of calibrations. At the moment it is not clear exactly how calibration will be done on the AT, and in particular whether it will be necessary to calibrate every hour or so (à la VLA), or every 12 hours (à la MERLIN), or, more probably, some combination of these. We probably won't know exactly what we want to do with our calibrators until we actually get some hands-on experience on the AT. Any calibration and catalogue sytem developed now will therefore almost certainly be modified in the early days of the AT. However, it is necessary to get some prototype sytem working now.

This draft specification is being circulated prior to the writing of the software so that pundits from all fields can check that the catalogue system will do what they want it to do. In particular, astronomers might like to consider the tasks listed in Section 2, the form of the source catalogue described in Section 3, and the types of calibration information stored in the observation file of the calibration catalogue. Those concerned with the maintenance of the AT might like to see if the receiver monitoring capabilities are adequate, and software experts might like to comment on the efficiency, or otherwise, of the file structures. Criticism from all quarters is, perhaps rashly, invited.

1.2 Modes Of Calibration And Observation

It is assumed that the compact array (CA) and long baseline array (LBA) will be used as separate instruments. Thus, although calibration information from both arrays may be stored in the catalogue described here, there will be no need to calibrate both arrays simultaneously. In particular, each observation record (desribed below) can contain calibration observations on only 6 telescopes at a time.

For the purposes of this report, a calibration observation is defined as a short (few minutes maximum) observation of a calibrator source, resulting in one complex visibility per baseline per channel. Each telescope is assumed to have up to 8 frequency bands available, and up to 4 simultaneous data streams transmitted to the correlator. It is assumed that calibration observations will be limited to 8 cross-correlation products per baseline. Thus, standard polarisation corrections will usually have been done before the data is written to the calibration catalogue, and all four Stokes' parameters (or uncorrected polarisation parameters) can be recorded on two frequency bands simultaneously.

Other types of calibration, not included above, will also be needed occasionally. These will include:

1. Long (upto 12 hours) observations of calibrators for initial baseline determination, and

2. Calibration observations over many (e.g. 1024) frequency channels per baseline, to calibrate filter responses.

These types of calibration are so specialised that it is expected that special software will be developed for these tasks. They are therefore not catered for in the catalogue system. However, it is important to ensure that the data archive system will be capable of handling, and disgorging upon request, these data.

1.3 Definitions

A catalogue is defined to include several files together with the software necessary to maintain them.

A pointer is

1. an INTEGER*2 which gives a relative address (i.e. record number) of a datum in another file. Thus files are not expected to exceed 32768 records.

2. a small number (e.g. 4) of bits within a BYTE field which point to an element of an array elsewhere in the same record.

A calibration source is a source which is observed to calibrate either the amplitude or phase response (or both) of an array.

An astrophysical source is a source which is observed because of astrophysical interest in the source itself. An astrophysical source may also be a calibration source.

2.0 THE NEED FOR A CATALOGUE

When the AT commences operation, there will be an immediate need for a computer-readable catalogue of calibration sources. In addition, as observations progress there will arise a need for a more general source catalogue, and a set of software for handling these catalogues. The requirements for such a system may be summarised as follows:

1. When asked for a flux or position calibrator, the system must be able to supply the position and other parameters of a suitable source.

2. For such calibration observations, the sytem must be able to find all previous observations of that source, so that changes in receiver gain, for example, may be monitored.

3. When observing an astrophysical source, the system must be able to find the most recent (or optimum in some other way) calibration observations, so that the astrophysical data may be calibrated on-line.

4. When observing any source that has been previously observed with the AT, it should not be necessary to supply the co-ordinates: they should be stored and readily accessible. This should not, of course, preclude the optional use of new or improved source co-ordinates.

5. The history of AT observations of any astrophysical source should be readily accessible in human-readable form. Thus it will be possible to see if a source has already been observed with the AT, and, if so, in what mode, and by whom. This could be elaborated to include a reference to the published results.

6. This AT history must include the location of all sections of the data, giving the numbers of the archive and transport disks and tape. If the LBA ends up using a tape-recording system, the history could include

the identification numbers of the tapes, although this section could be deleted once the tapes have been processed and erased.

7. The option should exist for an astronomer to include brief notes on the source. At its simplest, this might be a velocity, transition frequency, or flux density. At its most elaborate, this could include a classification (e.g. quasar, HII region), or references to the literature.

The list above indicates that there are two distinct groups of needs. One is for a calibration catalogue which is rapidly accesible and updatable by computer, but need not display its guts to humans. The other is for a source catalogue which is human-readable, but need only be machine-readable to the extent of obtaining a source position.

These needs seem best satisfied by the creation of two separate catalogues, which will now be descrbed in detail.

3.0 THE SOURCE CATALOGUE

3.1 Overview

This must be human-readable,and readily updatable, but with a small amount of information on each source which is machine readable. These requirements are most easily satisfied by a single file of 80-character card images. Each source will have one entry consisting of one or more fixed format cards and a number (which can be zero) of free format cards.

3.2 The Fixed Format Cards

3.2.1 The Source Card -

Each entry starts with a fixed format card containing upto 4 alternative source names, the R.A. and dec in both B1950 and J2000 co-ordinates, and a flag to say if it's a calibrator source (i.e. in the calibration file too). A suggested format for this card is:

```
$ 1234+567A NAME00001 NAME00002 NAME00003 hhmmss.s ddmmss.s hhmmss.s ddmmss.s C
```

The meaning of each datum is:

$	Identification of a source card
1234+567A	Identification of the source. Each source must have its first name in this form. The optional suffix A allows for the possibility of upto 26 sources in a 6 arcmin sq. It will have to be decided whether this name is J2000 or B1950. I suggest J2000.
NAME0000n	Three alternative names of up to 9 characters each The names must not contain spaces, and can if necessary contain more than 9 characters by occupying more than one field. Programs can test for this by looking for a dividing space between the fields.
hhmmss.s ddmmss.s	RA and Dec in B1950 co-ordinates
hhmmss.s ddmmss.s	RA and Dec in J2000 co-ordinates
C	A flag to show that it's a calibration source

Note that the spaces, and provision of both B1950 and J2000 co-ordinates appear to waste space but will make the file more accessible to users.

3.2.2 The Alias Card -

An alias card may be included if required for sources with more than 3 alternative names. This alias card is identical to the source card except that the $ in column 1 is replaced by an @ .

3.2.3 The Observation Card -

Astrophysical sources that have been observed with the AT will have a number of fixed format cards describing the AT observations. Each card will describe one observation, which is defined as a continuous (or nearly continuous) period of observation of the source with the array in one configuration. Thus there will usually be one card for each 12 hours of observation. It is expected that these cards will be written automatically by the synchronous computer at Culgoora. Each card will contain the start and stop times, the identification of the array and configuration used, and the location of the data. It will not contain precise details about calibration, etc., which are stored along with the data. A suggested format of this is:

```
  start start end array   freq    vel  lsr pol  bw chan  arch. transport
  date    st   st   id                          id       tape    tape

> ddmmyy hhmm hhmm aaaa ffff.fff nnnnn aa aaaa aa nnnn nnnnnnn nnnnnnn
```

Comments:

>	This is the observation card identification
array id	This could be e.g. L003 for an LBA configuration, or 3001 for configuration No.1 of the 3km array, etc.
freq	For freq > 10000MHz, the dec. pt. can be shifted to the right
vel	in integer km/s, as this is simply a guide to the user. (An exact value will be stored with the data)
lsr	e.g. LR for lsr radio defn. HO for heliocentric optical.
pol	e.g. IQUV , RR , etc.
chan	Number of channels per baseline

Archive and transport tapes: Lots of characters are available here to cope with anything we devise!

3.3 User Cards

Each data card may be followed by an arbitrary number of card images, in free-format, containing user-supplied information. The only constraints are that they must not start with $, @, or >. These cards may be written by anyone at Culgoora or Epping.

3.4 Ordering The Catalogue

The insistence of a PKS-type name for the first source name allows the catalogue to be sorted and ordered. The order will be of increasing RA, with increasing Dec being used to sort sources with the same RA.

3.5 Updating The Source Catalogue

Because of the size and vulnerability of this catalogue, most users will not have write access to the catalogue (although they will have read access). Instead, a program will be provided which will extract an entry from the catalogue, allow the user to alter it or add to it, and then append it to the end of the catalogue. At intervals (weekly, say) a privileged user will inspect the alterations, and, if they are OK, will re-order the catalogue so that the new entries replace the old ones. Software will be provided for checking and re-ordering.

It is expected that a copy of the catalogue will be available at both Epping and Culgoora. If the 'master' catalogue is assumed to reside at Culgoora, then it will be necessary to transport to Culgoora amendments made at Epping. This may be done weekly using, for example, a floppy disk, since only the amendments, rather than the entire catalogue, need be transported. On the other hand, it would probably be desirable to transport the entire catalogue from Culgoora to Epping, probably also at weekly intervals, and this would require a mag. tape which could be transported with the data tapes.

3.6 Example

Here is an example of an entry in the source catalogue. It should be noted that it is fairly understandable in its computer-stored form. The only non-obvious features are things like remembering which RA and Dec are B1950 and which are J2000, and which code means which bandwidth.

```
$ 0532-054  ORION-A   ORION-KL  3C999      053246.7-052423 051234.5-054321 C
@ 0532-054  ORION(OH) ORION-IRC2           053246.7-052423 051234.5-054321 C
> 010484 0900 1300 L004 1665.401     8 LR IQUV  2 1024 RPN0001 T000042
> 301288 0200 1400 3001 1612.231    20 HR IQUV  3 8096 RPN1041 T999998
> 311288 0200 1400 3002 1612.231    20 HR IQUV  3 8096 RPN1042 T999999
See Bloggs et al. Ap.J. 123,456. for velocities to be used.
Data from 010484 published in MNRAS 64898327, 33.
Data from 30-311288 being analysed by RPN
```

3.7 The Size Of The Source Catalogue

In one year the AT might make about 1000 12hour observations, and observe about 1000 sources including calibrators. If the verbosity of users extended to 2 lines per source, this would give a catalogue of 4000 lines, or 625 blocks on the VAX. After 10 years of this prolific activity, the file would still be only 6250 blocks. In practice the file would grow much more slowly after the first year or two of activity, since many sources will be re-observed many times, adding only one line for each observation. Calibrator sources will not have one line for each observation, as their observations are stored in the separate calibrator catalogue.

4.0 THE CALIBRATION CATALOGUE

4.1 Overview

The calibration catalogue is essentially a machine-readable catalogue, which does not need to be easily human-accessible. Several systems were considered, including a system based on Mike Batty's source catalogue for the

the VAX. After 10 years of this prolific activity, the file would still be only 6250 blocks. In practice the file would grow much more slowly after the first year or two of activity, since many sources will be re-observed many times, adding only one line for each observation. Calibrator sources will not have one line for each observation, as their observations are stored in the separate calibrator catalogue.

4.0 THE CALIBRATION CATALOGUE

4.1 Overview

The calibration catalogue is essentially a machine-readable catalogue, which does not need to be easily human-accessible. Several systems were considered, including a system based on Mike Batty's source catalogue for the Tidbinbilla interferometer, but in the end it was decided that a relatively simple sytem of indexed files was best. The catalogue actually consists of several files:

1. The observation file: This is the main file, which serves as an index file to all the other files. It consists of a number of 1024 byte fixed length binary records, each of which refers to one calibration observation. A record is added to this file each time a calibration source is observed.

2. The source file: This is a file of fixed length binary records. Each record contains details (RA, Dec, flux, etc.) of one source, together with a pointer to the record in the observation file which contains the most recent observation of that source. Each source identifier is accompanied by a generation number, so that source coordinates may be replaced by more accurate ones when they become available, without destroying the old coordinates assumed for previous calibrations.

3. The station file: This is a file of binary fixed length records, which list the positions of each telescope station. Each station identifier is accompanied by a generation number, so that station coordinates may be replaced by more accurate ones when they become available without destroying the old coordinates assumed for previous calibrations.

4. The receiver file: This is a file of 80 character ASCII card images which describe the history of individual receivers.

The format of each of these files will now be described in detail. Various identifiers which are common to the files are listed in Appendix B.

4.2 The Receiver File

Each record of the receiver file contains 80 ASCII characters. The first 13 characters of each record are in fixed format, and the remaining 67 characters can contain any details of changes to the receiver. The idea of this is so that the calibration program can alert the user of any receiver changes that occurred within a calibration data set.

The format is as follows:

tbiddmmyyhhmm...

where:

t is a single character telescope identifier (e.g. 1 ,2, or P)

b is a single character frequency band identifier (e.g. L, C, X)

i is a single character id to distinguish the receivers on any one telescope

ddmmyy is the date of the modification

hhmm is the AEST of the modification

.... is the description of the change (e.g. "Cryogenics exploded")

An example for a change to a C-band receiver on telescope 2 might be:

2C11009841123 Replaced first stage FET

A section of this file might look like:

```
2C11009841125 Replaced first stage FET again after misconnecting supply
PXA1109841234 Brought the receiver from Culgoora (6X2) and installed it
TK11909840930 Changed K band feed to linear polarisation
TK21909840930 Changed K band feed to linear polarisation
```

4.3 The Station File

Each record of the station file is 32 bytes long, and contains binary words written by:

```
CHARACTER*2 STATION_ID
REAL*8 X,Y,Z
INTEGER*2 GENERATION, DATE, TIME
WRITE n, STATION_ID, GENERATION, X,Y,Z, DATE, TIME
```

These parameters have the following meanings:

STATION_ID contains two characters which identify the station. Examples might be the numbers 01 to 36 for the 6km array, P for Parkes, T6 for the Tidbinbilla 64m, etc.

GENERATION is an integer with the value 1 for the initial (assumed) station coordinates, and incrementing by 1 for each re-determination.

X, Y, Z are geocentric coordinates of the stations in microseconds.

DATE and TIME are integers giving the date when the coordinates were measured, and are calculated by:

DATE=(year-1900)*1000 + (day number)

TIME=(hours of AEST)*100 + (minutes of AEST)

4.4 The Source File

Each record of the source file is 80 bytes long, and contains binary words written by:

```
      CHARACTER*8 NAME
      REAL*4 FLUX(8)
      LOGICAL*1 FLAG1(8),FLAG2(8)
      REAL*8 RA, DEC
      INTEGER*2 POINTER, GENERATION, DATE, TIME

      WRITE n, NAME, GENERATION, POINTER, RA, DEC, DATE, TIME,
    1 (FLUX(I), FLAG1(I), FLAG2(I), I=1,8)
```

These parameters have the following meanings:

NAME is a 8 character source name

GENERATION is an integer with the value 1 for the initial (assumed)

	source coordinates, and incrementing by 1 for each re-determination.
POINTER	gives the address in the observing file of the last observation of this source.
RA	is the J2000 Right Ascension of the source in radians
DEC	is the J2000 Declination of the source in radians
DATE and TIME	are integers giving the date when the coordinates were measured, and are calculated by: DATE=(year-1900)*1000 + (day number) TIME=(hours of AEST)*100 + (minutes of AEST)
FLUX(i)	The flux density (in Jy) of the source at frequency band i Set negative if not suitable as flux calibrator.
FLAG1(i)	Set TRUE if the source is unresolved for the CA at band i so that it can be used as a position (or phase) calibrator.
FLAG2(i)	Set TRUE if the source is unresolved for the LBA at band i

Note that the fluxes and flags allow for 8 frequency bands. At present, only 7 are planned for the initial operation of the AT.

4.5 The Observation File

The exact structure of the observation file depends pivotally on the selection of quantities that are to be stored. Since the preliminary draft of this report, the absence of one such quantity has already been noted. There will probably be other such additions, and the main function of the distribution of this document is to invite such additions.

Each record is 1024 bytes long (only 941 of which are currently allocated), and consists of a BYTE array EQUIVALENCEd to various quantities. Its structure may be illustrated by the following lines of FORTRAN which would read one such record.

```
      BYTE BUFFER(1024)

      INTEGER*2 SNUM, DATE, TIME, PTR, STNID(6), RXID(4,6)
      REAL*4 SCALE,FREQ1,FREQ2
      BYTE TELID(6), BASID(15), DATA(7,8,15)

      COMMON /BUFF/ SNUM, DATE, TIME, PTR, STNID, TELID, RXID, BASID, SCALE,
     1 FREQ1, FREQ2, DATA

      EQUIVALENCE (BUFFER,SNUM)

      READ n, BUFFER
```

A full description of the data items is given as Appendix A.

4.6 The Size Of The Calibration Catalogues

The calibration catalogue is dominated by the observation file. Assuming that a calibration observation is made every hour, 24 hours a day, 365 days a year, then the file will amount to 16k blocks after one year. This will therefore be small in terms of VAX storage for several years, after which earlier records may be deleted.

APPENDIX A

THE STRUCTURE OF THE OBSERVATION FILE

A.1 THE RECORD STRUCTURE

Each record contains data from all baselines at a particular time. It can accomodate upto 15 baselines from upto 6 telescopes. Each telescope may have up to 4 channels, at two frequencies, to yield 8 correlation products per baseline (2 frequencies * 4 polarisation parameters). The data items in each record are illustrated by the following code to read one record.

```
      BYTE BUFFER(1024)

      INTEGER*2 SNUM, DATE, TIME, PTR, STNID(6), RXID(4,6)
      REAL*4 SCALE,FREQ1,FREQ2
      BYTE TELID(6), BASID(15), DATA(7,8,15)

      COMMON /BUFF/ SNUM, DATE, TIME, PTR, STNID, TELID, RXID, BASID, SCALE,
     1 FREQ1, FREQ2, DATA

      EQUIVALENCE (BUFFER,SNUM)

      READ n, BUFFER
```

The meaning~~s~~ of each of these items is now listed.

SNUM — (INTEGER*2) is the source number, and is a pointer to the source file

DATE and TIME — (INTEGER*2) give the date when the ~~coordinates~~ observations were ~~measured~~ made, and are calculated by:

DATE=(year-1900)*1000 + (day number)

TIME=(hours of AEST)*100 + (minutes of AEST)

PTR — (INTEGER*2) is a pointer to the previous observation of this source

SCALE — (REAL*4) is a scaling factor for the flux densities:
True flux=(value in channel subfield) * SCALE

STNID(6) — (INTEGER*2) is a station identifier, and is a pointer to the station coordinate file

TELID(6) — (BYTE) is a telescope identifier, as used in the receiver file.

RXID(4,6)	(INTEGER*2) describes the four receivers in use on each telescope. It has a complex bit pattern which is described below.
BASID(15)	(BYTE) describes which of the possible combinations of receivers are used to form each correlation product. It has a complex bit pattern which is described below.
SCALE	(REAL*4) is a scaling factor for the amplitude: True amplitude=stored amplitude*SCALE
FREQ1	(REAL*4) is the frequency in MHz of frequency channel 1
FREQ2	(REAL*4) is the frequency in MHz of freqyency channel 2
DATA(7,8,15)	(BYTE) contains the data on each correlation product. It has a complex bit pattern which is described below.

A.2 THE RECEIVER IDENTIFIER (RXID)

The INTEGER*2 variable RXID is encoded as follows:

Bit 1	determines the frequency (0 => FREQ1 ; 1 => FREQ2)
Bits 2-7	determine the polarisation, and correspond to a character ASCII variable less the top bit.
Bits 8-15	contain a character which identifies the receiver.

A.3 THE BASELINE IDENTIFIER (BASID)

Each element of BASID consists of one byte with the following structure:

Bits 1-3 contain the pointer of telescope A on this baseline. This pointer points to an element of the array TELID

Bits 4-6 contain the pointer of telescope B on this baseline.

Bit 7 is a flag which is set if the data has been calibrated for atmosphere.

Bit 8 is a flag which is set if the data has been converted to Stokes' parameters.

A.4 THE DATA ARRAY (DATA)

The arguments of the array DATA(I,J,K) are as follows:

K is the baseline number, corresponding to an element of the array BASID

J is a channel number on this baseline, and cannot exceed 8.

I is the byte number, and is described in the structure below.

```
<byte1><byte2><byte3><byte4><byte5><byte6><byte7>
<RX2ID><  amplitude ><   phase    ><   slope    >
```

where:

RX2ID contains two 4-bit pointers to show which of the receivers have been correlated. Each pointer corresponds to the argument I of the array RXID(I,J).

Amplitude and phase are each INTEGER*2 (amplitude is scaled by the factor SCALE).

Slope (INTEGER*2) is the phase gradient (assumed linear) across the band, in units of 0.001 degree/MHz. Note that its maximum value is about 32.8 degree/MHz, or about 14 turns across 160MHz. This quantity may be obtained from any observation using more than one frequency channel (e.g. wideband continuum observations) and is of use in determining closure phase offsets and clock errors.

APPENDIX B

IDENTIFIERS

B.1 TELESCOPE IDENTIFIER

A single character used to identify a telescope regardless of its position. E.g. 1-6 for Culgoora, P for Parkes, etc. Used by the Receiver file and array TELID of the Observation file.

B.2 RECEIVER IDENTIFIER

A single character used to identify a particular receiver on a given telescope. Note that if a receiver is moved from one telescope to another, its Identification will in general change, but the change will be recorded in the Receiver file. Used in the Receiver file and in the array RXID of the observation file.

B.3 FREQUENCY IDENTIFIER

A single character used to identify a freqyency band (e.g. L,C,X, etc). Used by the receiver file. Note that the exact observing frequency (FREQ1, FREQ2) is stored in the Observation file.