# THE USE OF STANDARDIZED DATA FORMATS WITH THE WESTERBORK RADIO TELESCOPE

R.H.Harten and T.A.Th.Spoelstra

Netherlands Foundation for Radioastronomy,
Sterrewacht,
Leiden - 2405, The Netherlands

ABSTRACT: Since 1970 the Westerbork Synthesis Radio Telescope has been collecting data on galactic and extragalactic fields. The data is stored and cataloged on magnetic tape using a standard format. This data is routinely reduced and analyzed at different observatories in Europe. A reduction program package allows manipulation and display of the data. After the initial reduction all tape output from the reduction programs uses the same tape format regardless of the program. Transportation of data between observatories is done via magnetic tape with the same standard format. For data transportation to observatories not normally using Westerbork information a common tape format has been developed. Suggestions to alleviate problems with this data transfer are discussed.

## 1. THE WSRT AS A SYSTEM

Since 1970 the Westerbork Synthesis Radio Telescope (WSRT) has been collecting data on a large number of galactic and extragalactic fields. It consists of an array of 12 equatorially mounted 25 m paraboloidal dishes on an east-west line of which the two eastern telescopes can be placed at any position on a 300 m rail track. The other ten telescopes are fixed and arranged at equally spaced intervals of 144 m. In the second half of 1976 this system is enlarged with two additional telescopes, identical to the existing ones, which will provisionally be added as moveable telescopes to the existing rail track (Baars and Hooghoudt, 1974).

At present correlations between each of the orthogonal di-

*C. Jaschek and G. A. Wilkins (eds.), Compilation, Critical Evaluation, and Distribution of Stellar Data.* 69-75.

poles of a movable telescope and each of the orthogonal dipoles of a fixed telescope form 80 interferometers at 20 different baselines. When the two new telescopes come in operation, a new 5000 (frequency/polarization) channel auto-correlation receiver will also be used. The observing wavelengths include 6, 21 and 49 cm. The 6 and 21 cm observations are done both in the continuum and line mode (Casse and Muller, 1974; Allen, Hamaker and Wellington, 1974).

At present we have approximately $2\times10^{10}$ words ($64\times10^{10}$ bits) of data stored on magnetic tape. This data is divided into units called maps ($\sim 2.5\times10^{5}$ words) which contain the brightness distribution, at one of the three observing wavelengths, for a region of the sky (typically $1^{\circ}\times1^{\circ}$). Clearly there must exist a well defined system to handle this volume of data. The basis for WSRT reductions consist of a standard software package and a fixed set of tape formats. Details of the reduction system have been given by Van Someren Grêve (1974). The tape formats are discussed below.

The processing of WSRT data can roughly be broken into four stages: (1) observation and preliminary correction; (2) calibration; (3) Fourier transformation; (4) manipulation of map data (display, etc.). Between each stage, data is written on magnetic tape. The corrections in step 1 include instrumental corrections (e.g. for delay, baseline, frequency, clock, gain, phase errors and shadowing of the telescopes) and corrections for sky effects (e.g. spherical refraction, precession, nutation, aberration and ionospheric Faraday rotation). Other operations possible in this stage are: shifting the field centre, corrections for objects moving with respect to the fixed $\alpha$, $\delta$ coordinates, subtraction of point sources and additional gain corrections.

Reductions are considered to begin after the Fourier transform of the data when a map is made from the calibrated and corrected interferometer output. Generally the maps are produced in one of the four Stokes parameters. The output tape format for this stage is used in all further reductions. All programs use the same tape format for input and output. This holds for maps in the v,l or v,m as well as l,m space (here v stands for velocity in km/s; $l = \sin(\alpha - \alpha_o)\cos\delta$, $m = -\cos(\alpha - \alpha_o)\cos\delta$ and $\alpha - \alpha_o$ is the right ascension with respect to the field centre). The expanded WSRT system will require a 3-dimensional Fourier transform (l,m,v) but the data will remain stored in a 2-dimensional form (see below), but with a 3-dimensional cross-reference index. The reduction programs will accept both the old and new formats.

This data is routinely reduced and analyzed at at least four different observatories in Europe. There exists a standard reduction program package which also allows manipulation and display of the data. The manipulation includes map combination, projections, polarization parameter determination, cleaning of a map (Högbom, 1974), etc. The display modes include line plots, contour plots and optical intensity modulated photos. Also interactive programs have been developed for manipulation of these data at various observatories.

## 2. TAPE FORMATS

There are four tape formats used in data processing for the WSRT: (a) a format for Telescope tapes on which the observations are transported from the telescope to the computing centre where off-line corrections and calibrations are done, (b) a format for Makeobs tapes containing output from these corrections and calibrations, (c) a format for the reduction stage (Makemap tapes, i.e. Linemap tapes for the new system), (d) a format for tapes with which the results are transported to other institutes (Transport tapes, see Sections 3 and 4). The first is used solely by the reduction group and the users generally work only with the last three tape formats. The formats are summarized in Table 1.

## 3. TRANSFER OF DATA VIA TAPE TO OTHER INSTITUTIONS

Several other observatories use Westerbork data on a regular basis. Previously, data had been transfered via punched cards. This method has proved to be unsatisfactory. The chance for punch errors is too high and for large volumes of data it is impractical. To solve this problem, we developed a special tape format which was readable on almost any computer. This allows us to send data without special programs for each computer model. Also, users could use our tapes even if they switched computers. We write the data and catalogs in BCD card image format in records of 120 bytes long. Several of these records can be packed into a block to save space on the tape. Generally we write 7-track even parity tapes. However we also use 9-track odd parity if the institution is very modern and doesn't have the older 7-track units. To prevent problems of label reading, we do not write standard tape labels, only data and tape marks. The record size of 120 bytes is quite important since it is divisable by 4, 6, 8, 12 or 60. These are the most common word sizes in various computers presently in use. The block size should be an even number of 120 byte records. For many users the 240 byte length is quite handy since it represents 3 punched card images. Clearly, this format is not designed for maximum efficient use of the space available on a magnetic tape, but it reduces the amount of word needed to read and write transport tapes.

https://doi.org/10.1017/S0252921100052891 Published online by Cambridge University Press

TABLE 1

| Tape | Area | Format [1]) | Length [2]) | Contents | Organization |
|---|---|---|---|---|---|
| A Telescope | Start Catalog | Mixed: 2-4 ƀ I,<br>4-8 ƀ R,<br>1-4 ƀ L,<br>1 ƀ A | 116 | Parameters for object identification, receiver parameters and corrections added to the observation | 1 dataset per label (each dataset consists of 1 start catalog, ⩾1 data blocks and 1 end catalog) |
| B | Data | (as for entry A) | 636 | Interferometer output per baseline per time interval | |
| C | End Catalog | (as for entry A) | 116 | (as for entry A) | |
| D Makeobs | Catalog | Mixed: 2-4 ƀ I,<br>4-8 ƀ R,<br>1 ƀ L,<br>1 ƀ A | 160 | Parameters for object identification, observational identification | ⩾1 dataset per label (each dataset consists of 80 sets of 1 catalog and 1 or 2 data blocks) |
| E | Data | 2 ƀ I | ⩽4096 | Corrected observations recorded per time interval per baseline | |
| F Makemap | Catalog | Mixed: 2-4 ƀ F,<br>1,2,4 ƀ I | 4096 | Information to define the map and its parameters; description for each baseline whose data was used in making the map | 1 dataset per label (the number of blocks is determined by the number of points in a map. Any unfilled part of the last block contains meaningless values). The basic unit is a map or antenna pattern and its catalog |

https://doi.org/10.1017/S0252921100052891 Published online by Cambridge University Press

TABLE 1 (continued)

| | Tape | Area | Format | Length | Contents | Organization |
|---|---|---|---|---|---|---|
| G | | **Data** 3) | 4 Ƃ F | 4096 | Values of Brightness/ Velocity/Polarization values on an evenly spaced grid | |
| H | Linemap | Master Catalog | (as for entry F) | 3840 4) | Information common to an entire set of maps: i.e. coordinates, map size & cross-reference of all maps in the set | |
| I | | Map Catalog | (as for entry F) | 1920 4) | Information about a single map: i.e. frequency, bandwidth, etc. | The number of blocks is adjustable. It is possible that a second block is needed |
| J | | Data | Mixed: 2-4 Ƃ I | 30720 or 3840 5) | See entry G. The data in this section are the mantissa of a floating point number. All numbers have some exponent which is stored in the catalog. The first bit is a sign bit, the remaining are the mantissa. | Data and catalogs are organized to allow for a 3-D array of data. For each set of 2-D maps which make up the 3-D array, there is a master catalog. Each map consists of a map catalog and a data section. Map catalog and master catalog are also written at the end of the label. |

Notes to Table 1:

1) 1 Ƃ = 1 byte = 8 bits
I = integer; R = real;
L = logical; A = alphabetic;
F = floating point number

2) in bytes

3) Map and Antenna Pattern have the same format

4) 120 Ƃ records

5) 3840 Ƃ for CDC machines

https://doi.org/10.1017/S0252921100052891 Published online by Cambridge University Press

## 4. STANDARD INDENTIFICATION BLOCK

At present when data is transfered between observatories, either a copy of an existing tape is made or a special tape is created. Usually, the contents of the tape are described (in varying degrees of detail) in a separate description. In many cases the description is lost or not available and the user has to go to some effort to find out what is on the tape and how it is written. Even if the tape is well documented, some of the technical details such as parity, block length, etc. are not described. To help alleviate this problem, we suggest that each tape begin with an identification area. This area would contain a description of the tape format and its contents. The area would have a fixed, standard block length but the number of blocks would be adjustable. The number of blocks in the area would be written in the first block. The information written in the area would be in BCD card image format. Thus the contents of the area could be read by a simple FORTRAN, ALGOL or PL/1 instruction. To minimize problems we suggest that the tape be written without standard labels. Thus, the tape would consist of the identification area, tape mark and as many data areas, followed by tape marks, as needed. The information contained in the area would be written in a form that is

TABLE 2

Identification Area

A. Number of records/blocks of 240 bytes.

B. General Format of Tape:
   Number of labels, labeling conventions used, parity, BCD, EBDIC or binary. Machine used to create the tape.

C. General Description of the data on the tape and its purpose including:
   Type of data on the tape i.e. radio, infrared, spectral data, etc., format of data (grid of some quantity on a grid of some coordinates, including description of coordinates), intensity units, and any general notes pertaining to all data on the tape.

For each label there should be an entry describing:

D. Technical Parameters of data in the label:
   Block length of the data in the label; number of blocks in the label; format of data within the label.

E. Description of contents or type of information written on tape:
   Description of field coordinates, grid size and astronomical source. Calibration procedure and sources used; telescope, beamsize (or grating), wavelength, filters used, bandwidth, integration time, noise level, person with knowledge of observing procedure and calibration.

self defined, thus no description is necessary. A suggestion for the contents of such an area is outlined in Table 2. While the proposed identification area may not be the optimum solution to the problem, we suggest that some form of standard identification be written on all tapes used to transport information between observatories. The value of the system is that a tape's contents can be identified without a description and by someone not familiar with the tape.

The record/block length for the identification area would be 240 bytes per record/block. This would equal three card images and would be handleable on most computers. The first record would contain the number of records/blocks in the area. This would be a 4 digit number at the start of the record/block. The rest of the information would be in BCD character form. The identification area would be followed by a tape mark. Unfortunately, one must know the parity and number of tracks (7 or 9) of the tape before it can be read. If a standard is adopted 7 track, even parity or 9 track, odd parity are reasonable possibilities.

## REFERENCES

Allen, R.J., Hamaker, J.P., Wellington, K.J., 1974, Astron.& Astrophys. 31, 71

Baars, J.W.M., Hooghoudt, B.G., 1974, Astron.& Astrophys. 31, 323

Casse, J.L., Muller, C.A., 1974, Astron.& Astrophys. 31, 333

Högbom, J.A., 1974, Astron.& Astrophys.Suppl.Series 15, 417

Van Someren Grêve, H.W., 1974, Astron.& Astrophys.Suppl.Series 15, 343

https://doi.org/10.1017/S0252921100052891 Published online by Cambridge University Press