Section 2.11

Machine-Readable Files and CD-ROMs

2.11. Machine-Readable Files and CD-ROMs

The ASCII CD-ROMs contain the primary products of the mission, as deposited in data centres, and which thus represent the long-term archive of the Hipparcos mission. Users wishing to transfer the Hipparcos and Tycho Catalogues, and their associated annexes, into their own data bases will use these files.

2.11.1. Conventions for the ASCII CD-ROMs

End of records: Records for data and index files in cats/, notes/, and tables/ are terminated by 'carriage return' plus 'line feed' (CR+LF, or \r\n, or decimal ASCII codes [13,10]). Record lengths listed in the file content description tables do not include these two characters.

Retaining CR or LF alone would have satisfied one or more specific systems, but would have left others interpreting the data files as a single record. The penalty for including both is that different systems will interpret the CR+LF as the end of record terminator, plus an additional redundant character (e.g. '^M' appears at the end of line on Unix systems). Nevertheless, the files can be displayed legibly by most Mac word processor applications, by the DOS 'type' command, and by the Unix 'vi' word processor. Files in /src are Unix compatible, and end of record characters must be converted for use on other systems.

Field delimiters: All fields, except for the last one in a record, are followed by the vertical line '|' character, which therefore separates all ASCII file fields. It is indicated in the file content descriptions by the FORTRAN 1X format. The last field in a record is followed by the record terminator.

Alternative choices of spaces, tab characters, or commas, were considered. Avoiding the use of a blank means that blanks can be entered when data are absent, and the files can still be read into editors or spreadsheet-type programmes. The tab character is convenient in some systems, but inconvenient (and invisible) in others.

Integers representing multiple bits: In certain data fields in the annexes a number of 1-bit flags ('binary digits' or 'bits') have been encoded into a single decimal number, for storage efficiency. [A bit-string conversion routine is defined within the file utils.h]

A two-digit decimal number can accommodate up to 6 binary digits (bits), all 6 bits set corresponding to the decimal number 63. The individual bits are to be recovered from the decimal value in the following way: any integer number can uniquely be decomposed into a sum of powers of 2, thus, $45 = 2^0 + 2^2 + 2^3 + 2^5$. The individual bits in such a quantity are numbered from zero (the least-significant bit) upwards. A particular bit is 'set' if the corresponding power of two appears in the decomposition. In this example, bit numbers 0, 2, 3 and 5 are set, while bit numbers 1, 4, and all higher bits are not set.

Missing data: Blanks are used whenever data are absent or when a field is empty. Due caution must be taken to avoid interpreting a blank field as the numerical value 0. [An appropriate interpretation facility is defined within the file utils.h]

2.11.2. Contents and Directory Structure of the ASCII CD-ROMs

The Hipparcos CD-ROMs are formatted for the ISO–9660 Standard, Data Interchange Level 1. Specific details of the data formats are given hereafter.

Directory Structure

Relevant elements of the following directory tree are contained on each CD-ROM:

/			Readme files
	cats/		Catalogue files, index files
	docs/		Documentation (and some figures) in PDF format
	notes/		Notes and references
	tables/		Identification tables (Volume 13)
	charts/		Identification charts
		dss/	DSS identification charts (‘D’ in Field H69)
		gsc/	GSC identification charts (‘G’ in Field H69)
	curves/		Light curves
		a/	Folded light curves (‘A’ in Field H54)
		b/	AAVSO light curves (‘B’ in Field H54)
		c/	Unfolded light curves (‘C’ in Field H54)
	src/		C/Fortran utilities
	fits/		FITS conversion utilities

Data Files and File Names

Details of the index files, file names, formats, and disk organisation are given in Tables 2.11.1–2.11.4. File names follow a DOS-like (ISO–9660 standard) representation: a file name (maximum of 8 characters), followed by an extension of up to 3 letters.

By convention this documentation uses Unix-style directory paths to denote the location of files on the CD-ROMs, with all the paths and file names in lower case. The actual directory path and appearance of the file name on the target computer will depend on the operating system and device driver being used. Thus a UNIX file called temp.ext will appear as follows:

UNIX:	/mycdrom/temp.ext (if the CD-ROM is mounted as mycdrom)
PC:	E:\TEMP.EXT (if the CD-ROM is drive E)
MAC:	DISK1:TEMP.EXT (accessed by double clicking on the DISK1 icon)
VAX/VMS:	MYCDROM:[*]TEMP.EXT;1

The information in each file can be inferred from the extension appended to each file name as follows:

dat:	table data stored as ASCII text
doc:	documentation or other text stored in ASCII format
idx:	index file
eps:	encapsulated PostScript files (for light curves and identification charts)
pdf:	PDF (Portable Document Format) files, viewable with Acrobat/Acroread
f:	Fortran code
c:	C code
h:	C code header

Index Files

Index files are provided for each data file, facilitating the location of records relating to a particular HIP or TYC identifier. As for the machine-readable data files, '|' is used as a field separator, and \r\n at the end of record.

There are four types of 'identifier' which are used to order the relevant data files: HIP identifier; TYC identifier; CCDM identifier; and solar system object identifier. Depending on the specific data file, different structures of the associated index file have been adopted to optimise the search procedure. These search procedures fall into the following categories:

• direct search: hip_main.idx, hip_i.idx, and hip_j.idx contain entries for the complete range of HIP identifier (irrespective of the existence of an associated valid record). The HIP identifier may thus be used as a subscript to directly access the relevant data files, using the corresponding index file;

• binary search: the majority of index files contain entries only for those identifiers existing in the data file. This makes the index file much smaller, but the corresponding data record may only be found indirectly, either by reading the file sequentially, or using a binary search algorithm for more efficient access (see, for example, W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C, 2nd Edition*, Cambridge, 1992);

• mixed search: tyc_main.idx includes direct access for the first component of the TYC identifier (TYC1), and requires a binary search in the data file for the full identifier.

The load and search routines in C described below (and in the readme file) invoke an appropriate search algorithm, based on the index file provided (where appropriate), in a manner fully transparent to the user. Search routines are provided for all files with the exception of curves.idx, charts.idx, and hp_refs.doc for the reasons noted below.

Each index file is summarised hereafter, indicating the applicable search method, and with the detailed structure of each index file given in Table 2.11.1. These will permit users to develop additional access routines, where necessary.

(a) Files ordered by HIP identifier:

hip_main.idx [direct search]: a single-column table provides, as the *n*-th entry, the record number containing the data for HIP *n* (the length of this index table is 120 416). The first record is specified as record 1 (rather than record 0). A value of −1 indicates a non-existing HIP entry.

hip_dm.idx [binary search]: this is a 'combined' index file, which allows the five mutually exclusive parts of the Double and Multiple Systems Annex (C, G, O, V, X) to be accessed using a single index file. The index file has three columns: the HIP identifier; a one-letter column pointing to the relevant annex; and a column providing the record number within the relevant annex. In the case of entries in Part C of the Double and Multiple Systems Annex, the record number corresponds to the *first* record of the solution containing the relevant HIP number: up to eight further records (3 component and 5 correlation records) may follow under the same CCDM identifier. Note that hip_dm_c.dat is also accessible by means of the CCDM identifier (see below).

hip_va.idx [binary search]: this is a 'combined' index file, which allows the two parts of the Variability Annex to be accessed using a single index file. The index file has three columns: the HIP identifier; a one-letter column pointing to the relevant annex; and a column providing the record number within that annex.

hip_ep.idx [direct search]: structured as for hip_main.idx, but with the index giving the *first* record of the relevant HIP identifier.

hip_i.idx [direct search]: structured as for hip_main.idx, but with the index giving the *first* record of the relevant HIP identifier.

hip_j.idx [direct search]: structured as for hip_main.idx, but with the index giving the *first* record of the relevant HIP identifier.

hg_notes.idx [binary search]: the index file has two columns: the HIP identifier; and a column providing the record number of the *first* entry for the given HIP identifier (note that the number of entries for each HIP identifier is stored within each data record).

hd_notes.idx [binary search]: as for hg_notes.idx.

hp_notes.idx [binary search]: as for hg_notes.idx.

hp_auth.idx [binary search]: as for hg_notes.idx. Note that hp_refs.doc provides an alternative (indirect) index associating reference identifiers to a given HIP identifier, particularly convenient for interrogating the printed catalogue. Multiple entries for each HIP identifier are possible. For the machine-readable files, hp_auth.idx circumvents the use of hp_refs.doc (which therefore has no associated index file).

curves.idx: the index file indicates the directory for the corresponding curve. No search routine is provided since the result of the search will be user specific, and since standard UNIX utilities (e.g. grep) are adequate for searching this file.

charts.idx: the index file indicates the directory for the corresponding chart. No search routine is provided since the result of the search will be user specific, and since standard UNIX utilities (e.g. grep) are adequate for searching this file.

Other files (dmsa_o.dat, hip_rgc.dat, ident*.doc) have no associated index files due to their small size, and are interrogated (if required) directly in the relevant search routines.

(b) Files ordered by CCDM identifier:

hip_dm_c.idx [binary search]: this uses a two-column table. The first column gives the CCDM identifier, and the second column gives the record number of the first component record for this system. The lines are sorted on CCDM identifier, in lexical order. hip_dm_c.idx is the CCDM identifier-based index file for Part C of the Double and Multiple Systems Annex only.

Note that all five parts of the Double and Multiple Systems Annex are also indexed on the HIP identifier by means of the combined index file hip_dm.idx. Thus Part C of the Double and Multiple Systems Annex may be accessed either by means of the CCDM or the HIP identifier.

(c) Files ordered by Tycho Catalogue identifier:

Index files for the Tycho Catalogue entries (tyc_main and tyc_ep) are more complex since the TYC identifier comprises three components (e.g. TYC 3512–33–1; see Section 2.2 for details of the numbering scheme). Note that the first two components of the TYC identifier correspond to the GSC identifier, where the first component (TYC1) is in the range 1–9537.

tyc_main.idx [mixed search]: for locating data within the Tycho Catalogue: the Tycho Catalogue comprises one record per entry (i.e. a total of slightly more than 1 million records), ordered by TYC identifier. So, a given entry could be located using a binary-search algorithm in at most 20 steps. In order to speed up this procedure, a one-column table (tyc_main.idx) with 9538 entries provides, as the n-th entry, the record number of the first star in GSC region n. IDX(n+1)–1 gives the last star in the same region. The record related to the relevant entry must be searched for between these limits. The first record is specified as record 1 (rather than record 0).

Example: To find TYC 3512–33–1 within the Tycho Catalogue (on Disk 1), tyc_main.idx is loaded once and held in memory, say in an array I(9538). I(3512) directly gives the record number of the first star in GSC region 3512. I(3512+1)–1 gives the last star in this region. The Tycho Catalogue is then searched for TYC2 = 33, which will fall between records I(3512) and I(3512+1)–1.

tyc_ep.idx [binary search]: for locating data records within the Tycho Epoch Photometry Annex, a table tyc_ep.idx is provided, which contains two columns: the TYC identifier, and the record number of the header record of the corresponding entry in tyc_ep.dat. Note that although the tyc_ep.dat file is larger than tyc_main.dat, a binary search is implemented in the former case because the number of TYC entries (and hence header records) is smaller.

(d) Files ordered by solar system object identifier:

solar.idx [binary search]: this is a combined index file covering the three files containing data on the solar system objects. It has 8 columns. Column 1 gives the solar system object identifier listed in Table 2.7.1 (to be distinguished from a HIP Catalogue number). Column 2 gives a sequential running number (from 1–55) for programming convenience. The remaining 6 columns giving the first record number (–1 if there are no data associated with this object identifier) and the number of records (0 if none) in the three solar system data files, solar_ha.dat, solar_hp.dat, and solar_t.dat.

Data Formats

The formats of data files are included at the end of the section describing the corresponding contents of each relevant catalogue or annex (the relevant section number describing the data is given in Tables 2.11.4/1–6).

Note that 'hip_main' includes the 'extended' Fields H71–77, which are not given in the printed catalogue. Certain data in the Double and Multiple Systems Annex are also contained only in the machine-readable files (as described in Section 2.3).

The formats of the files containing the notes and references are given in Table 2.11.2.

The formats of the 'cross-identification' tables (Volume 13) are given in Table 2.11.3.

Load and Search Routines

Search programs: C code search programs are provided for all data (*.dat) files. Each search program contains a main function which scans the command line for identifiers and applicable search options, then uses the routines summarised below to search for (and print) the relevant data records. The search options are:

- h help
- a apply b, c, s together
- b bit field decoding (applicable to HEPA and TEPA)
- c correlation coefficient extraction (applicable to DMSA/C, G, O, V)
- s sub-record print
- t tabular print

These options may be combined in any order, thus –a ≡ –bcs. The default option prints the main record (verbose), sub-record with no decoding (tabular). Multiple identifiers, and requests for following records, may be given. For certain files (notes/*.doc and solar*.dat) only the tabular print option is likely to be of use.

Naming convention: The C file search programs generally have the same name as the corresponding index file, but with the letter 's' prepended to the name, underscores (_) removed, and with the extension .c (for example, the search routine for hip_main.dat, which makes use of hip_main.idx, is shipmain.c). In the case of hip_ep_e.dat, which utilises the same index file as hip_ep.dat, the search program is named shipepe (i.e. following the name of data file rather than that of the index file). This results in the following search programs: shipmain, stycmain, shipdm, shipdmc, shipva, ssolar, shipep, shipepe, stycep, shipi, shipj.

Additional ('redundant') search programs are provided based on the data file name as follows: shipva1, shipva2, shipdmg, shipdmo, shipdmv, shipdmx, shgnotes, shdnotes, shpnotes, shpauth, sdmsao (indexed on the reference number in Field DO16), shiprgc.

Index code: The C code provided in the src directory loads the index file and locates an entry given an identifier of the appropriate type. It implements a binary or direct search for each index file, as appropriate. The C file has the same name as the corresponding index file, but with the letter 'i' prepended to the name, underscores (_) removed, and with the extention .c (for example, the index interrogation routine accompanying hip_main.idx is ihipmain.c). A header file (.h) accompanies each .c file.

Data code: The src directory contains .c and .h files for every data file (plus some auxilary files in some cases). The source code contains information on the record structure of the corresponding data file (i.e. field names and record length, etc.), as well as routines to perform the following functions:

• search: for a given identifier, the appropriate find routine is used to derive the record number of the requested identifier, and the following routine is used to read it;

• read: for a given record number (not an identifier), the routine reads the record from the data file, including any related data and sub-records (in normal usage, the read routine is invoked by the search request);

• verbose print: for a given record, print the record 'vertically', with a summary description of each field;

• tabular print: for a given record, print the record 'horizontally' (as it would appear in the data file). Field identifiers are given at the top of the record, truncated where necessary to preserve the appropriate layout.

Further details of the load and search routines are given in the readme file in the src/ directory.

Position Propagation Routines

Disk 1 includes Fortran and C implementations of the equations given in Section 1.5.5 for the rigorous epoch transformation of the astrometric parameters and their covariances. The subroutine takes as input the original epoch (in Julian years), the five astrometric parameters (in the units used by the Hipparcos and Tycho Catalogues), the radial velocity parameter ζ (defined by Equation 1.5.24, in mas/yr), and the covariance matrix of the six parameters, all referred to the original epoch, and the required new epoch; the output consists of the corresponding six astrometric parameters and covariance matrix referred to the new epoch. The relevant files are:

- for the Fortran 77 version:
  – src/pos_prop.f (source code with documentation)
- for the C version:
  – src/pos_prop.h (header file containing prototypes and documentation)
  – src/pos_prop.c (source code of the procedure).

It has been verified that the two versions give identical results to within the rounding errors of the double precision arithmetics. No test program is supplied with the routines. The following numerical example should help in validating their application.

Example: Given the Hipparcos astrometric data for Barnard's Star (HIP 87937), propagate its astrometric parameters and covariance matrix to the epoch J1900.0. The five astrometric parameters at epoch J1991.25 are taken from Fields H8–9 and H11–13. Table 1.2.3 gives the radial velocity as −111.0 km/s. The initial values for the parameter vector to be propagated are therefore:

$$\left.\begin{aligned} \alpha_0 &= 269.45402305 \ \text{deg} \\ \delta_0 &= +4.66828815 \ \text{deg} \\ \pi_0 &= 549.01 \ \text{mas} \\ \mu_{\alpha*0} &= -797.84 \ \text{mas/yr} \\ \mu_{\delta 0} &= +10326.93 \ \text{mas/yr} \\ \zeta_0 &= -12855.29 \ \text{mas/yr} \end{aligned}\right\} \ \text{J1991.25}$$

where ζ_0 is calculated from Equation 1.5.24. Using the standard errors in Fields H14–18 and the correlations in Fields H19–28, the initial covariance matrix is found to be:

$$\mathbf{C}_0 = \begin{pmatrix} +1.768900 & +0.143640 & -0.336224 & -0.620977 & -0.017157 & +7.872819 \\ +0.143640 & +1.166400 & -0.324216 & -0.086940 & -0.362232 & +7.591647 \\ -0.336224 & -0.324216 & +2.496400 & +0.381570 & -0.203820 & -58.45420 \\ -0.620977 & -0.086940 & +0.381570 & +2.592100 & +0.436149 & -8.934613 \\ -0.017157 & -0.362232 & -0.203820 & +0.436149 & +1.664100 & +4.772526 \\ +7.872819 & +7.591647 & -58.45420 & -8.934613 & +4.772526 & +4721.914 \end{pmatrix}$$

where Equation 1.5.69 was used to compute the last row and column, assuming a standard error $\sigma_{V_{R0}} = 0.5$ km/s for the radial velocity. Units are mas^2, $\text{mas}^2\text{yr}^{-1}$ and $\text{mas}^2\text{yr}^{-2}$ as appropriate for the different elements.

Propagation of these parameters and covariances from T_0 = J1991.25 to T = J1900.0 by means of pos_prop results in the following parameters (note that the propagation was based on the parameters and covariances calculated in double precision from the catalogue values, not on the rounded values given above):

$$\left.\begin{aligned} \alpha &= 269.47419117 \ \text{deg} \\ \delta &= +4.40801091 \ \text{deg} \\ \pi &= 545.90 \ \text{mas} \\ \mu_{\alpha*} &= -788.54 \ \text{mas/yr} \\ \mu_{\delta} &= +10210.27 \ \text{mas/yr} \\ \zeta &= -12829.25 \ \text{mas/yr} \end{aligned}\right\} \ \text{J1900.0}$$

with covariance matrix:

$$\mathbf{C} = \begin{pmatrix} +21510.98 & +3187.087 & -36.37617 & -234.0989 & -30.28683 & +958.0380 \\ +3187.087 & +14919.00 & +41.45412 & -33.92317 & -172.8802 & -2343.445 \\ -36.37617 & +41.45412 & +2.440553 & +0.412525 & -0.711158 & -56.34087 \\ -234.0990 & -33.92317 & +0.412525 & +2.548358 & +0.314266 & -11.97391 \\ -30.28683 & -172.8802 & -0.711158 & +0.314266 & +2.092324 & +46.58443 \\ +958.0380 & -2343.445 & -56.34087 & -11.97391 & +46.58443 & +4615.337 \end{pmatrix}$$

From ζ and π the radial velocity at J1900.0 is found to be $V_R = -111.406330$ km/s, with a standard error of 0.499986 km/s. Propagation of these parameters and covariance matrix from J1990.0 to J1991.25 should return the original values to within the rounding errors of the arithmetic.

PDF Files

A series of PDF (Portable Document Format) files, viewable with Acrobat/Acroread, are also included on the ASCII CD-ROMs as follows:

- disk 1 includes, in the docs/ directory, four sub-directories, vol1–4, which contain the pdf forms of Volumes 1–4. In each directory the file tocvol*.pdf provides a 'clickable' table of contents. Some figures are low-resolution scans of the printed originals; and some colour images (in particular those from Section 3 of Volume 1, and Appendix F of Volume 2) are not included on disk 1 for reasons of space (see below);

- disk 5 includes, in the docs/ directory, volume*.pdf, which are the pdf forms of Volumes 5–10. The contents of these files follows the sequence of the corresponding printed volumes, as summarised on page xi. The file tocvol.pdf provides a 'clickable' table of contents for these six volumes;

- disk 6 includes, in the docs/ directory, volume11.pdf, which is the pdf form of Volume 11. The contents of this file follows the sequence of the corresponding printed volume, as summarised on page xi. The file tocvol11.pdf provides a 'clickable' table of contents for this volume;

- disk 6 also includes, in the sub-directory docs/figs/, the separate figures from Section 3 of Volume 1. These are included as f3_1_001–007.pdf, f3_2_001–127.pdf, f3_3_001–047.pdf, f3_4_001–042.pdf, and f3_5_001–026.pdf.

The *.pdf files may be viewed (and magnified), searched, printed, and converted to *.ps files, from within Acrobat/Acroread.

Table 2.11.1. Format of the index files.

File	Column	Format	Description of the *n*-th entry (−1 if no information)
hip_main.idx			
	1	I6	Record number of HIP *n* in hip_main.dat
tyc_main.idx			
	1	I7	Record number of first star in GSC region *n*
hip_dm_c.idx			
	1	A10, X	CCDM identifier of HIP *n*
	2	I5	Record number of *first* record for this CCDM identifier in hip_dm_c.dat
hip_dm.idx			
	1	I6, X	HIP identifier
	2	A1, X	Annex identifier (C, G, O, V, X) indicating hip_dm_c/g/o/v/x.dat
	3	I5	Record number of HIP *n* in the given annex (if Field 2 = C, gives the *first* record of the solution)
hip_va.idx			
	1	I6, X	HIP identifier
	2	A1, X	Annex identifier (1, 2) indicating hip_va_1/2.dat
	3	I4	Record number of HIP *n* in the given annex
solar.idx			
	1	I3, X	Solar system object identifier (see Table 2.7.1)
	2	I3, X	Sequential running number of object (from 1–55)
	3	I5, X	Record number of object in solar_ha.dat
	4	I4, X	Corresponding number of records in solar_ha.dat
	5	I5, X	Record number of object in solar_hp.dat
	6	I4, X	Corresponding number of records in solar_hp.dat
	7	I5, X	Record number of object in solar_t.dat
	8	I4	Corresponding number of records in solar_t.dat
hg_notes.idx			
	1	I6, X	HIP identifier
	2	I4	Record number of *first* record for this HIP identifier
hd_notes.idx			
	1	I6, X	HIP identifier
	2	I4	Record number of *first* record for this HIP identifier
hp_notes.idx			
	1	I6, X	HIP identifier
	2	I4	Record number of *first* record for this HIP identifier
hp_auth.idx			
	1	I6, X	HIP identifier
	2	I4	Record number of reference in hp_auth.doc for this HIP identifier Records 1 and 2 may be repeated for a given HIP identifier

...cont

Table 2.11.1. Format of the index files (cont.).

File	Column	Format	Description of the *n*-th entry (−1 if no information)
curves.idx			
	1	I6, X	HIP identifier
	2	A1	A, B, C or J indicating curve for HIP *n* in a/ b/ c/ (J = A and B)
hip_ep.idx			
	1	I8	Record number of *first* record of HIP *n* in hip_ep.dat
charts.idx			
	1	I6, X	HIP identifier
	2	A1	'D' or 'G' indicating chart for HIP *n* in dss/ or gsc/
tyc_ep.idx			
	1	I4, I6, I2, X	TYC identifier
	2	I8	Record number of *first* record of this TYC identifier
hip_i.idx			
	1	I7	Record number of *first* record of HIP *n* in hip_i.dat
hip_j.idx			
	1	I7	Record number of *first* record of HIP *n* in hip_j.dat

Table 2.11.2. Format of the notes and references.

File	Field	Bytes	Format	Description
hg_notes.doc				
	GN1	1–7	I6,X	Hipparcos Catalogue (HIP) identifier
	GN2	8–9	A1,X	If 'D', double and multiple systems note also given
	GN3	10–11	A1,X	If 'P', photometric note also given
	GN4	12–14	I2,X	Number of records *N* for this HIP identifier
	GN5	15–17	I2,X	Sequential number (1 to *N*) of this record
	GN6	18–97	A80	Text of note
hd_notes.doc				
	DN1	1–7	I6,X	Hipparcos Catalogue (HIP) identifier
	DN2	8–9	A1,X	If 'G', general note also given
	DN3	10–11	A1,X	If 'P', photometric note also given
	DN4	12–14	I2,X	Number of records *N* for this HIP identifier
	DN5	15–17	I2,X	Sequential number (1 to *N*) of this record
	DN6	18–97	A80	Text of note
hp_notes.doc				
	PN1	1–7	I6,X	Hipparcos Catalogue (HIP) identifier
	PN2	8–9	A1,X	If 'G', general note also given
	PN3	10–11	A1,X	If 'D', double and multiple systems note also given
	PN4	12–14	I2,X	Number of records *N* for this HIP identifier
	PN5	15–17	I2,X	Sequential number (1 to *N*) of this record
	PN6	18–97	A80	Text of note
hp_auth.doc				
	PA1	1–7	A6	Identifier of this photometric reference (yy.nnn)
	PA2	8–80	A73	Text
hp_refs.doc				
	PR1	1–7	I6,X	Hipparcos Catalogue (HIP) identifier
	PR2	8–10	I2,X	Number of records *N* for this HIP identifier
	PR3	11–13	I2,X	Sequential number (1 to *N*) of this record
	PR4	14–19	A6	Reference identifier (yy.nnn)

(The format of dmsa_o.doc is given in Section 2.3)

Table 2.11.3. Format of Identification Tables

File	Column	Format	Description
ident1.doc			
	1	I6	Hipparcos Catalogue (HIP) identifier
ident2.doc			
	1	I6,X	HD Catalogue identifier
	2	I6	Hipparcos Catalogue (HIP) identifier
ident3.doc			
	1	I6,X	HR (Bright Star Catalogue) identifier
	2	I6	Hipparcos Catalogue (HIP) identifier
ident4.doc			
	1	A11,X	Bayer/Flamsteed name
	2	I6	Hipparcos Catalogue (HIP) identifier
ident5.doc			
	1	A11,X	Variable star name
	2	I6	Hipparcos Catalogue (HIP) identifier
ident6.doc			
	1	A16,X	Common star name
	2	I6	Hipparcos Catalogue (HIP) identifier

Table 2.11.4/1. Directory contents for Disk 1

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	hip_main.dat	Hipparcos Catalogue [Section 2.1]	53 434 536
	hip_main.idx	Index on HIP	963 328
	tyc_main.dat	Tycho Catalogue [Section 2.2]	372 532 864
	tyc_main.idx	Index on TYC	85 842
	hip_dm_c.dat	DMSA Part C [Section 2.3]	8 922 960
	hip_dm_g.dat	DMSA Part G [Section 2.3]	516 534
	hip_dm_o.dat	DMSA Part O [Section 2.3]	79 665
	hip_dm_v.dat	DMSA Part V [Section 2.3]	42 048
	hip_dm_x.dat	DMSA Part X [Section 2.3]	37 464
	hip_dm_c.idx	Index on CCDM for DMSA Part C	219 510
	hip_dm.idx	Combined index on HIP for DMSA Parts C-X	286 672
	hip_va_1.dat	Variability annex (periodic) [Section 2.4]	390 528
	hip_va_2.dat	Variability annex (unsolved) [Section 2.4]	798 048
	hip_va.idx	Combined index on HIP for hip_va_1/2	123 810
	solar_ha.dat	Solar system objects: HIP astrometry [Section 2.7]	370 194
	solar_hp.dat	Solar system objects: HIP photometry [Section 2.7]	171 535
	solar_t.dat	Solar system objects: TYC data [Section 2.7]	28 227
	solar.idx	Combined index for solar system objects	2310
docs/			
	vol*	Volumes 1–4 (pdf)	~ 70 Mb
	tocvol*.pdf	Table of contents, Volumes 1–4	
notes/			
	hg_notes.doc	Notes (HIP) ordered by HIP	385 902
	hd_notes.doc	Notes (DMSA) ordered by HIP	259 182
	hp_notes.doc	Notes (variability annex) ordered by HIP	241 956
	hp_auth.doc	Author list (variability annex)	355 470
	hp_refs.doc	References (variability annex)	709 149
	dmsa_o.doc	References for DMSA Part O [Section 2.3]	9676
	*.idx	Index file for the associated *.doc file	494 286
tables/			
	ident1.doc	Inconsistencies between HIC/HIP identifiers	728
	ident2.doc	HD Catalogue number to HIP number	1 486 995
	ident3.doc	HR Catalogue number to HIP number	136 155
	ident4.doc	Bayer and Flamsteed names	88 880
	ident5.doc	Variable star names	127 800
	ident6.doc	Common star names	2400
curves/			
	curves.idx	Index on HIP for light curves	35 810
	a/nnnnnn.eps	Folded light curve for HIPnnnnnn	89 297 691
	b/nnnnnn.eps	AAVSO light curve for HIPnnnnnn	8 495 537
	c/nnnnnn.eps	Unfolded light curve for HIPnnnnnn	22 943 559
src/			~ 1 Mb
	pos_prop.c	Position propagation (C)	
	pos_prop.f	Position propagation (Fortran)	
	pos_prop.h	Position propagation (documentation)	
	readme.doc	Description of the C load and search routines	
	s*.c	Search routine for *.dat (without underscore)	
	i*.[ch]	Index routine for *.dat (without underscore)	
	*.[ch]	Data file structure for *.dat	
	utils.[ch]	Utilities invoked by search and read routines	
fits/			
	tofits.doc	Explanation for FITS conversion	
	tofits.dat	FITS conversion files	

Table 2.11.4/2. Directory contents for Disk 2

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	hip_ep.dat	HIP Epoch Photometry Annex [Section 2.5]	422 367 000
	hip_ep.idx	Index on HIP	1 204 160
charts/			
	charts.idx	Index on HIP for identification charts	116 530
	dss/nnnnnn.eps	DSS identification chart for HIPnnnnnn	109 543 641
	gsc/nnnnnn.eps	GSC identification chart for HIPnnnnnn	98 733 354

Table 2.11.4/3. Directory contents for Disk 3

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	hip_ep_e.dat	HIP Epoch Photometry Annex Extension [Section 2.5]	647 629 400
	hip_ep_c.dat	HEPA Coincidence file [Section 2.5]	1 894 194
	hip_ep.idx	Index on HIP (identical to index for hip_ep)	1 204 160

Table 2.11.4/4. Directory contents for Disk 4

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	tyc_ep.dat	TYC Epoch Photometry Annex [Section 2.6]	584 369 054
	tyc_ep.idx	Index on tyc_ep.dat	792 258

Table 2.11.4/5. Directory contents for Disk 5

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	hip_rgc.dat	Reference great circle data [Section 2.8]	173 234
	hip_i.dat	HIP intermediate astrometry [Section 2.8]	521 438 910
	hip_i.idx	Index on HIP	1 083 744
docs/			
	volume*.pdf	Volumes 5–10 (pdf)	~ 130 Mb
	tocvol.pdf	Table of contents, Volumes 5–10	

Table 2.11.4/6. Directory contents for Disk 6

Directory	File name	Content	File size
	readme.mac	Description file for the CD-ROMs: MAC	
	readme.dos	Description file for the CD-ROMs: PC	
	readme.unx	Description file for the CD-ROMs: Unix	
cats/			
	hip_j.dat	HIP transit data file [Section 2.9]	552 596 812
	hip_j.idx	Index on HIP	1 083 744
docs/			
	volume11.pdf	Volume 11 (pdf)	~ 15 Mb
	tocvol11.pdf	Table of contents, Volume 11	
	figs/f*.pdf	Figures from Section 3 of Volume 1 (pdf)	~ 50 Mb

2.11.3. Checksums for the Printed Catalogue

Astrometric data represents a somewhat specific contribution to astronomical data archives, tending to be only slowly superseded by new observational material. Mindful of its possible historical value, and cautious about the rapid evolution of all forms of machine-readable data archiving, the printed catalogue incorporates a page-by-page checksum in order to facilitate the verification of any form of automated optical character reading of the printed version, should this prove to be necessary or desirable in the future.

For each printed catalogue page, two independent checksums have been computed. All non-blank characters in the table fields (i.e. excluding the headers and page numbers) have been concatenated in normal reading order. For example, for a left-hand page of the main catalogue, the resulting character string starts with the HIP number of the first star (possibly preceded by an asterisk) and ends with the goodness-of-fit of the last star on the page.

At each transition to a new catalogue field, and after the last field on a line, a vertical line symbol, |, is inserted (this field separator, corresponding to the structure of the machine-readable ASCII files, ensures that the correct field association is accounted for). For example, for the first page of the catalogue, the string would begin with:

$$1\,|\,|\,00\,00\,00.22\,|\,+01\,05\,20.4\,|\,9.10\,|\,|\,H\,|\,\ldots$$

and end with:

$$\ldots\,|\,+4\,|\,+16\,|\,+40\,|\,0\,|\,0.26\,|$$

Each character is then replaced by its 8-bit ASCII representation, and a cyclic redundancy check is computed using the CCITT polynomial (see, e.g., W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C, 2nd Edition*, Cambridge, 1992). The resulting checksum is given in decimal representation.

The second checksum has been computed in essentially the same way, but instead of the normal ASCII representation, a bit-reversed representation has been used. For example, the ASCII representation of the character C is 67, or 01000011 in binary representation. For the second checksum, 11000010 is used. The resulting checksum has been bit reversed before being put in decimal form.

The two checksums are appended at the lower right of the tables on each page, in the order: normal checksum, bit-reversed checksum.

The bit pattern corresponding to the character string is considered to represent a polynomial, i.e. the individual bits from the character string are considered as coefficients of a polynomial. If the bits are numbered starting from the least significant bit of the last character, and if the latter bit is numbered 0, then bit n is the coefficient of x^n. This polynomial is then multiplied by x^{16}, and divided in modulo-2 arithmetic by the 'standard CCITT' polynomial $x^{16} + x^{12} + x^5 + 1$. The remainder of this division is a polynomial of degree 15. The (16-)bit pattern representing this polynomial is the checksum.

2.11.4. *Celestia 2000*

The Hipparcos and Tycho Catalogues (main catalogues and some annexes) also appear on one (non-ASCII) CD-ROM which includes software to interrogate the catalogues and annexes, to select samples from the main catalogues, and to draw local and global maps. This product, *Celestia 2000*, includes data from the ASCII disk 1 (see above), including notes and references, identification charts (from disk 2) and variable star light curves, if any. The data related to solar system objects are not included. No other data from disks 2–6 are available on this CD-ROM. The mission results are complemented by additional data taken from an updated version of the Hipparcos Input Catalogue, and by data or cross-identifications taken from the SIMBAD database for Tycho stars.

The data are 'bit-optimised', with an ordering optimised to take account of the requirements of the interrogation software. The CD-ROM follows the ISO-9660 standard. The software offers a wide range of possibilities which are briefly presented below. A complete user's manual and help are included in *Celestia 2000*, as well as technical requirements and instructions for installation and use. At the time of the release of the Hipparcos and Tycho Catalogues the *Celestia 2000* software is only available for IBM-PC compatibles.

Interrogating Facilities

The data are available per star, through eight windows displaying all data available in one catalogue or annex:

(1–5) Satellite data: Hipparcos Catalogue and annexes. These windows provide all data from the main Hipparcos Catalogue, including notes, references, and the identification chart if any; all data from the Double and Multiple Systems Annex, Parts C, G, O, V, and X; tabular data and light curves from Variability Annex, Sections 1 (periodic variables) and 2 (unsolved variable);

(6) Satellite data: Tycho Catalogue. This window provides all data from the main Tycho Catalogue, including notes. These data are complemented by some additional data made available from the SIMBAD data base: cross-identifications and proper motions from the PPM;

(7–8) Ground-based data: updated Hipparcos Input Catalogue and Double and Multiple Star Annex. These windows provide data from the Hipparcos Input Catalogue, with updated radial velocities, spectral types and luminosity classes; and with a few corrections in the cross-identifications. The annex has been updated with the errata provided to the CCDM (J. Dommanget & O. Nys, 1996, *Bull. CDS*, 48, 19).

Sampling Facilities

Samples can be defined using data from the Hipparcos and Tycho Catalogues, or from the Hipparcos Input Catalogue, or using identifiers from the Hipparcos Input Catalogue for HIP stars, or from SIMBAD for TYC stars. Once the source of the data has been selected, three options are available for selecting the subset of stars from which the

sampling will be performed: stars included only in HIP, but not in TYC; stars common to HIP and TYC; and stars included only in TYC, but not in HIP. The chosen subset of stars can also be 'all HIP stars', by selecting the first two options; or 'all TYC stars', by selecting the last two options.

More than 100 parameters can be used to define a criterion. A sample can be selected on the basis of one criterion, or of any logical combination of criteria.

Once a sample has been selected, it is then possible: to list the identifiers of the selected stars (HIP number if the star is in HIP, TYC number *only* if the star is in TYC but not in HIP); to display the HIP or TYC window by clicking on the selected identifier; to switch from any of these windows to any other one, where relevant; to draw a global or a local sky map for the selected sample; to prepare a file for external use, with data and identifiers from HIP and/or TYC and/or HIC.

Mapping Facilities

The global or local sky distribution of HIP and/or TYC stars can be displayed in various ways: draw a local map centred on a given star; draw a local map centred on given coordinates; draw a global sky distribution for a sample previously selected.

The sizes of symbols in local charts are related to the magnitudes of the stars. Different symbols can be used for stars selected by a given criteria, for other HIP stars, or for other TYC stars.