ASTRONOMY
FROM LARGE DATABASES
SCIENTIFIC OBJECTIVES AND METHODOLOGICAL APPROACHES

ESO Conference and Workshop Proceedings No. 28

THE INCA DATABASE FOR THE PREPARATION OF THE HIPPARCOS MISSION

Arenou F. & Morin D.
Observatoire de Meudon, F-92195 Meudon Principal Cédex

ABSTRACT : The INCA database, including all stars proposed for Hipparcos observation, has been created as a subset of SIMBAD database. It has now evolved independently and contains about 214,000 stars (with 5,000 stars not yet in SIMBAD). The data contained in INCA are daily improved with the help of specific softwares.

1. WHY THE INCA DATABASE

The Hipparcos project, included in the scientific programme of the European Space Agency (E.S.A) in 1980, aims at the very accurate measurement of the positions, parallaxes, and proper motions of about 110,000 preselected stars, with a precision of about 0.002" on positions and parallaxes, and of 0.002" per year for proper motions.

The INCA Consortium, leaded by C.Turon, is carrying out the preparation of the INput CAtalogue, which will contain all the stars to be observed by the Hipparcos satellite.

From the very beginning of this work, in early 1982, the support of the Centre de Données de Strasbourg (C.D.S.), and the intensive use of the SIMBAD database turned out to become essential in order to sort out redundancies among the 600,000 stars submitted by more than 200 proposers of the international astronomical community.

At that time, SIMBAD clearly was the only set of identifications and stellar data which could be interrogated using any identification, thus allowing to clarify most of the redundancies (see Fig 1). Cross-identifications were done using the Strasbourg-Cronenbourg Univac computer from Meudon Observatory.

The completeness of SIMBAD for all the stars up to about 9th magnitude allowed the construction of a basic list of bright stars (survey) accounting for half of the Input Catalogue (see Turon, Gómez & Crifo, this volume).

The first release of the catalogue (1983) contained 160,000 stars in a sequential file. Including the survey stars and stars newly identified, and eliminating redundancies, the INCA database now contains about 214,000 stars. The necessity of interrogating, of making statistics, and of updating very frequently the catalogue naturally implied, early 1985, the creation of a specific database, named INCA (acronym for INput CAtalogue), suited to our task, with the same structure and basic data as SIMBAD, thus allowing the use of all the softwares developed for SIMBAD, which have proved to be very efficient.

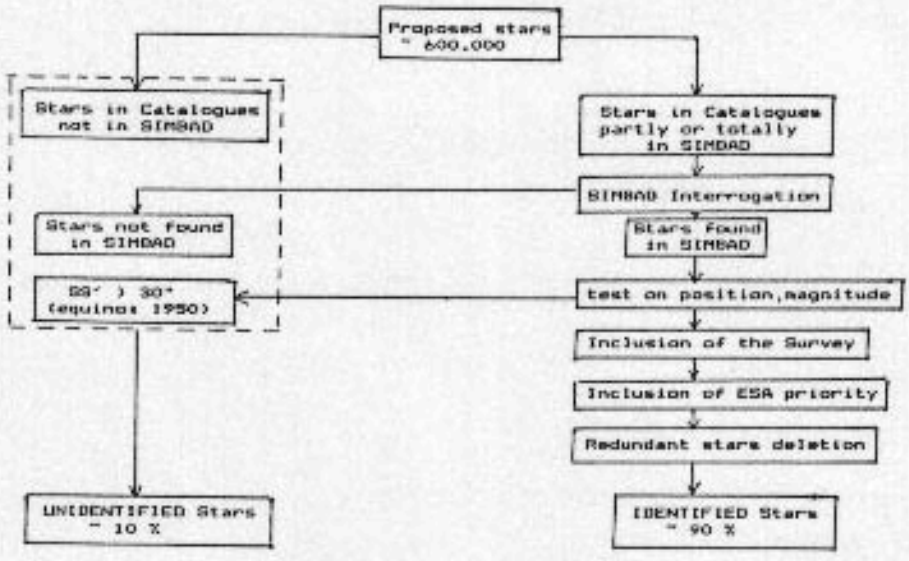


Figure 1

Situation before the INCA Database creation.

2. A SPECIFIC DATABASE

Once created, the INCA database was then developed in an autonomous way, though in close collaboration with the C.D.S. team (see Fig 2). The INCA database includes additional stars, different entries and new identifiers and data :

2.1 New stars

The INCA database now contains about 214,000 stars : 209,000 stars from SIMBAD and 5,000 additional stars which were not yet in SIMBAD.

2.2 Different entries

Due to the specificity of the Hipparcos satellite, stars in double and multiple systems received a special treatment according to the following rule :

1 entry if $\rho < 10''$ (AB) and 2 entries if $\rho \geq 10''$ (A+B), where ρ is the separation between two components A and B.

2.3 Identifiers

In addition to identifiers already in SIMBAD, some other identifiers (see Fig 3) were required to take into account :

- by which proposal each star was demanded ("*I*" identifier)
- the location of a star in 0.81 square degrees cells (size of the field of view of the satellite), to study the sky distribution ("*IBIL*" identifier) ;
- the Lausanne photometric identifier to facilitate the introduction of new photometric data provided by the Lausanne team ("*LID*" identifier) ;
- information about components of double and multiple systems given by the Working Group on Double & Multiple Stars, Observatoire Royal de Belgique ("*CCD2*" & "*CCDM*" identifiers).

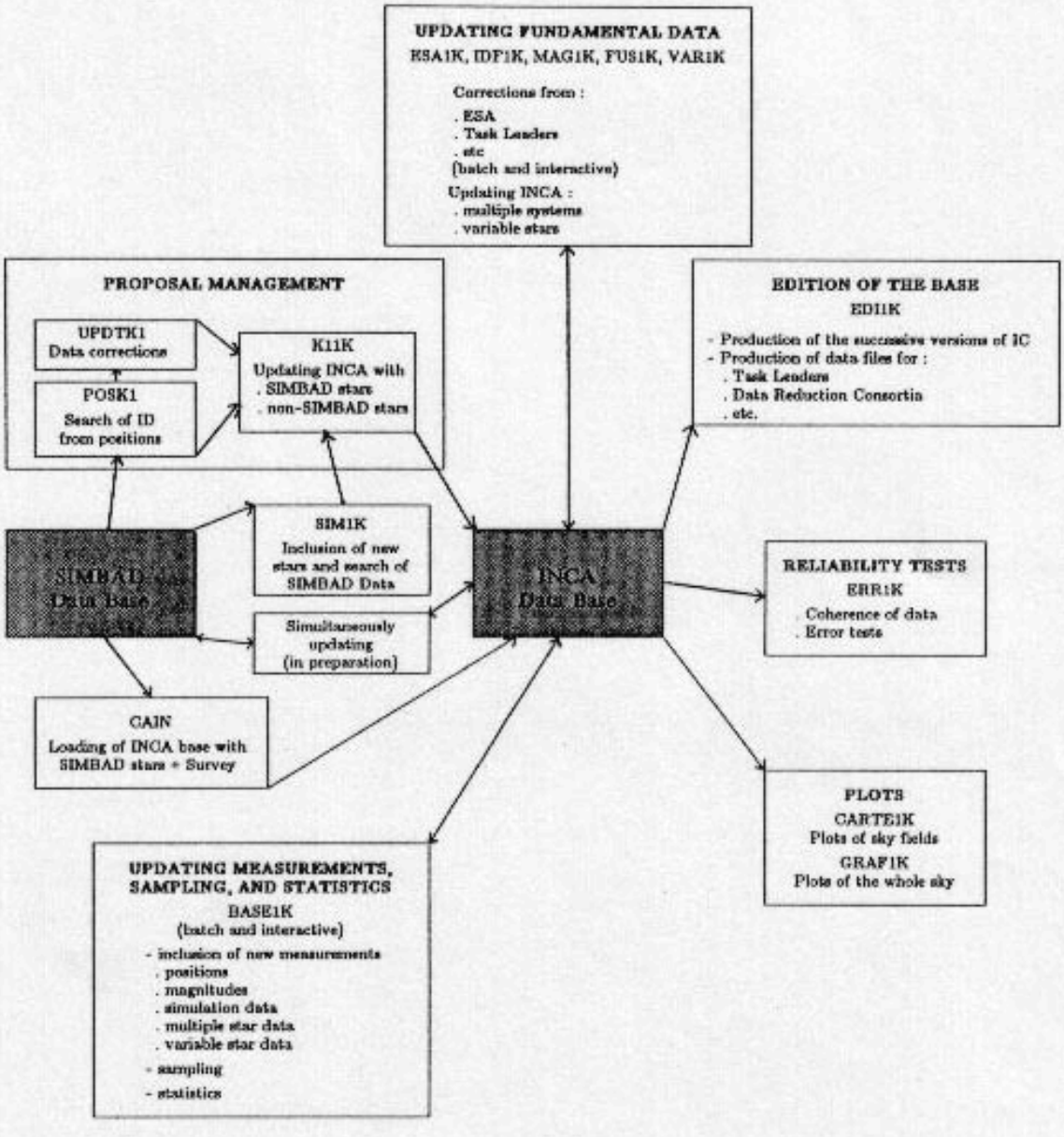


Figure 2
Input Catalogue preparation : New softwares.

INCA Identifier

Identifier name	Proposal number	Running Nr in the proposal	ESA priority	Proposer's priority	error code
I	54	2182	4R	0	0

IBIL Identifier

Identifier name	Cell number in galactic latitude : IB	Cell number in galactic longitude : IL	Discriminating number of stars in the cell
IBIL	-78	134	1

Figure 3
"I" and "IBIL" identifiers.

2.4 Measurements

New measurements, continuously improved, obtained by ground-based observations and measurements or compiled in the literature by other working groups of the INCA Consortium are introduced in several catalogues. The situation in October 1987 is :

- "*pos*" and "*pmt*" containing positions and proper motions : 181,000 and 189,000 measurements respectively ;
- "*pH*" and "*vH*" for photometry about normal and variable stars : 86,000 and 12,000 measurements respectively ;
- "*sH*" for parameters concerning numerical mission simulation : 1,600,000 measurements ;
- "*CCDM*" containing data about double and multiple stars.

3. NEW SOFTWARES

Apart from the softwares written by the C.D.S. staff, many new softwares operating on the INCA database and including a total of 36,000 PL/I statements and 6,000 FORTRAN statements have been written for 3 years, to solve different problems specifically related to the elaboration of the IC, though these softwares are compatible with SIMBAD structure.

3.1 Management of proposals

This software allows the exploration of SIMBAD, within a 10 arcmin radius around each of the 50,000 stars which were proposed using identifiers not recognized by SIMBAD.

Another software introduces the data concerning the newly recognized stars from SIMBAD into INCA .

3.2 Updating the INCA database

New identifiers, fundamental data and measurements are updated following the recommendations of ESA, INCA task-leaders, working groups coordinators, and proposers. This is performed using a C.D.S. SIMBAD software as well as new softwares described in Fig 2. Another software allows corrections of the entries in double & multiple systems.

3.3 Sampling and making statistics

In order to analyse the sky distribution, the effects of the different simultations and to detect redundancies, softwares were designed to perform samples and statistics using different astrometric, photometric, and spectroscopic parameters, ESA priority, etc.

3.4 Plots

Graphic softwares were written, to quickly figure out the local and global distribution of the considered stars. Figure 4 illustrates the aspect of the sky distribution of the stars proposed by proposal #53 ; Figure 5 displays a zone of 4 square degrees around the 502th star of proposal #53.

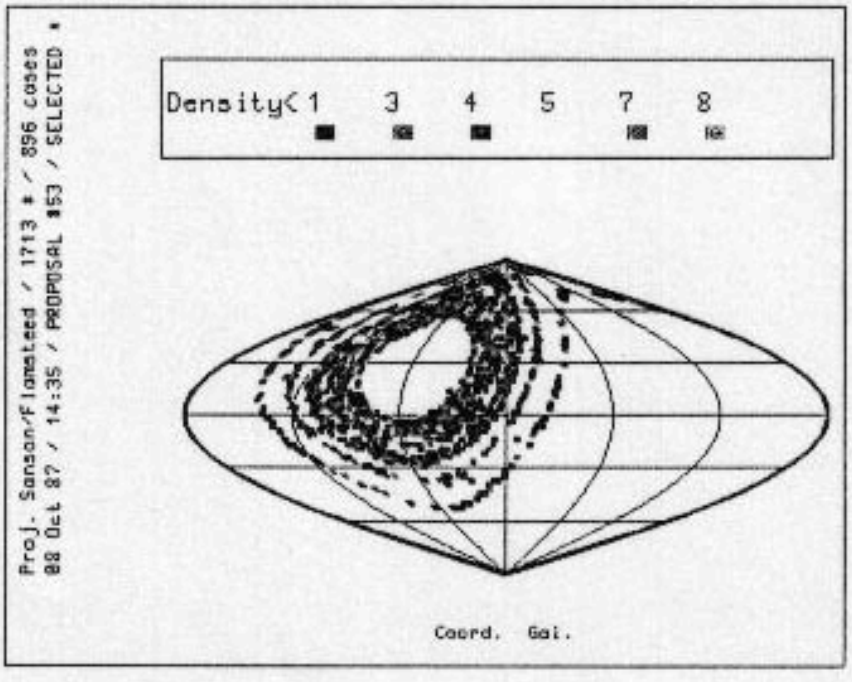
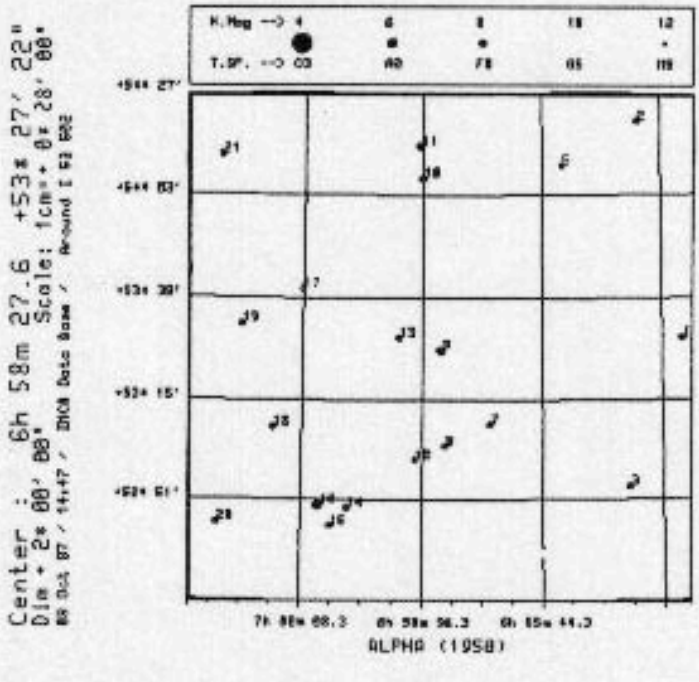


Figure 4

Figure 5



3.5 Edition of the INCA database

One important software allows the edition of the database for the different INCA working groups. Several editions have been necessary, each one containing improved data, for the simulation group, for ground-based observers, for ESA, for Data Reduction Consortia, etc.

3.6 Reliability tests

In order to improve the reliability of the Input Catalogue, a software has been designed to detect possible errors on astrometric, photometric and spectroscopic data, errors on cross-identifications and redundancies ; for each star, the consistency of all the data is tested : within each kind of data, between different types of data (e.g. colour index vs spectral type), and between identifiers and data.

4. TECHNICAL ASPECTS

4.1 Staff

Apart from the staff of the other working groups who provides the newly compiled or observed data for the preparation of the Input Catalogue, 3 scientists, 2 software engineers, and 4 technicians are specifically involved in the updating of the INCA database at the Observatoire de Meudon.

4.2 Internal structure of the database

The structure is the same as that of SIMBAD :

- 1 main file, containing all the data concerning one star : fundamental data, list of identifiers and measurements ; each field be variable in length ;
- 3 identifiers index files in order to allow access to an object from any of its identifiers ;
- 1 coordinate boxes data file in order to allow a quick access to stars when interrogating by their coordinates ;
- the bibliographical texts data file is the SIMBAD one ; it contains the texts of the references of the measurements.

4.3 Content of the database (October 1987)

stars : 214,000 ;

identifiers : 2,125,000 (547,000 "*" identifiers) ;

measurements : 3,166,000 ;

The INCA database contains 153 MBytes :

- 47 MBytes for fundamental data, system information and free bytes reserved for future use ;
- 63 MBytes for identifiers and measurements (90% proper to INCA) ;
- 43 MBytes for the index files ;

On average, each star is read 120 times/year and modified 12 times/year (see Fig 6)

4.4 Performances

The performances of the INCA database are similar to those of SIMBAD (Wenger, 1985) :

- interrogation by identifier : instantaneous answer in dialogue mode, 1h30m for the whole database in batch ;
- interrogation by coordinates : 2 sec in dialogue mode ;
- sampling : less than 5 min for a sample requiring the reading of the whole database.

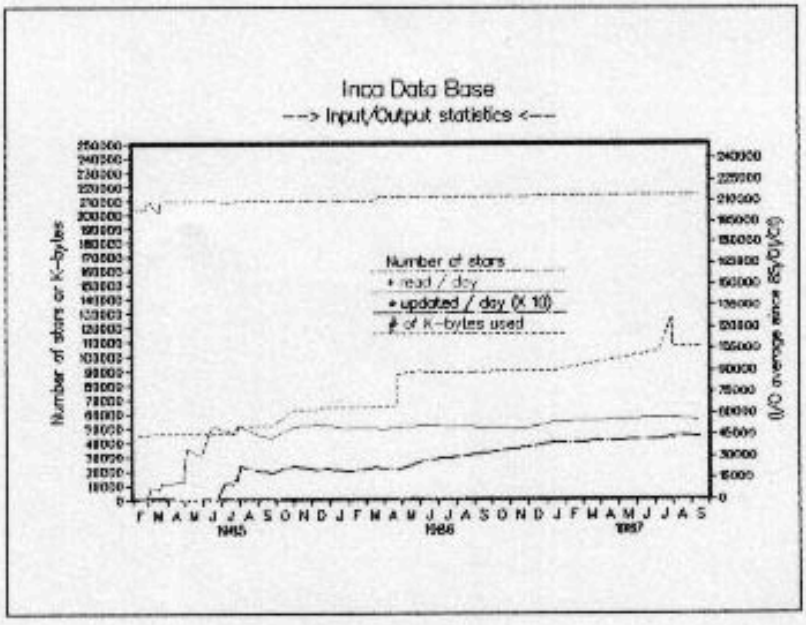


Figure 6

REFERENCES

- "Hipparcos : scientific uses of the INCA database", Turon C., Gómez A., Crifo F., this volume.
- "The INCA database, sub base of Simbad", Morin D., Arenou F., 1985, INCA Colloquium "Scientific aspects of the Input Catalogue preparation", Aussois, pub ESA SP-324, Perryman & Turon eds.
- "Construction of the Input Catalogue", Gómez A., Crifo F., Morin D., Arenou F., 1985, INCA Colloquium, Aussois, op.cit.
- "Presentation of the astronomical database SIMBAD", Wenger M., 1985, INCA Colloquium, Aussois, op.cit.