12. TYCHO PROCESSING SUMMARY

An overview of the actual data processing in the Tycho project and the size of the various data streams is given. Compared to the overview of the Tycho processing given in Chapter 2, this chapter puts the emphasis on the technical structuring of the data reduction chain, rather than on its scientific logic.

12.1. Introduction

As already described in Chapter 2, the TDAC processing scheme was a very complex data reduction pipeline with large amounts of data flowing through the different stages. Figure 12.1 shows an overview of the complete processing scheme of the main processing (left) and the reprocessing (right). The main purpose of this figure is to provide a different overview of the Tycho data reduction than was given in the Figures 1.2 and 10.1. The figure consists of two principal sorts of objects:

- boxes with embedded texts, indicating the institutes and the types of processes carried out;
- arrows with associated texts, indicating the type and volume of data transferred between the processing steps.

All the processes internal to TDAC are marked by boxes having the same size; smaller boxes indicate auxiliary data provided by institutions from outside TDAC, e.g. the Standard Star Catalogue (SSC) from Geneva. The arrows indicate the direction of data delivery, their width give a hint of the relative amount of data. Approximate numbers for the data volume are given with most of the arrows. The abbreviations present in the figure are those used throughout this volume. Some of those considered most important, sparsely used or particularly confusing because of the existence of a number of similar ones are briefly explained in the lower part of the figure.

Figure 12.1 emphasizes the actual data flow and the technical structuring of the data reduction chain, while Figures 1.2 and 10.1 put more emphasis on the scientific logic. The figure also shows more clearly the dependencies between the different stages and major blocks, like the main processing, reprocessing and the catalogue production.

Since most of the arrows in Figure 12.1 denote physical data interfaces (i.e. mostly 160-Megabyte half-inch magnetic tapes delivered by surface mail), it is obvious that the definition and proper application of these interfaces played a key role within TDAC. In fact, the consumption of time per tape and disk input/output processes, rather than

the actual processing of the data, has been the main bottleneck within TDAC. The process with the highest data turnover from and to other processes was the detection and estimation step in the main processing. More than half of the computing time and power in this step was used to maintain the input/output from and to tapes.

12.2. Raw Data and Pre-Reduction

The main entry point of the reduction scheme was the delivery of the telemetry tapes from the European Space Operations Centre to TDAC. Two identical copies of these tapes were sent in batches of about one month of real satellite time to Astronomisches Rechen-Institut, Heidelberg, and Astronomisches Institut, Tübingen. The attitude, satellite status and orbit data were used at Heidelberg, the housekeeping and raw photon counts (Tycho data stream) were used at Tübingen.

Figure 12.1 is divided into two big data reduction chains, namely the main processing and reprocessing. Each of the two chains is also divided into two major blocks, although less obviously. They are separated by the arrows labelled 'ITDB'. The two reduction blocks above the 'ITDB' arrows (i.e. the upper left and upper right portions of the figure) form the pre-reduction of the raw data. Their results are the two identified transit data bases (ITDB), one for the main processing and one for the reprocessing, respectively.

The Pre-Reduction Block for the Main Processing

The pre-reduction block for the main processing (upper left) consisted of eight different processing steps distributed between the participating institutes. Much of the complexity of this first block stems from the need to distribute the huge amount of data and processing requirements of the Tycho project over several institutes (this statement has to be understood in view of the computer technology affordable at the time of project initiation, i.e. about 1985). There were also scientific reasons for the complex structure, signified by the intermediate processes recognition, prediction updating and transit identification, which were described in Chapter 2.

Some software tools could be used for several processing steps. This is indicated by using a single box in the case of prediction updating where the independent processes 'updating-2' and 'updating-3' have been carried out by essentially the same software. Some of the logical processing steps were integrated into a single technical process or at least lined up to build a processing pipeline. This is indicated in the figure by the boxes containing two or three different processes.

The Pre-Reduction Block for the Reprocessing

Most of the resources used in reprocessing were imported from the main processing and adopted or expanded to the special needs. The reprocessing scheme looks much more straight-forward, because there was no need to carry out the recognition and updating steps, as explained in Chapters 2 and 10.

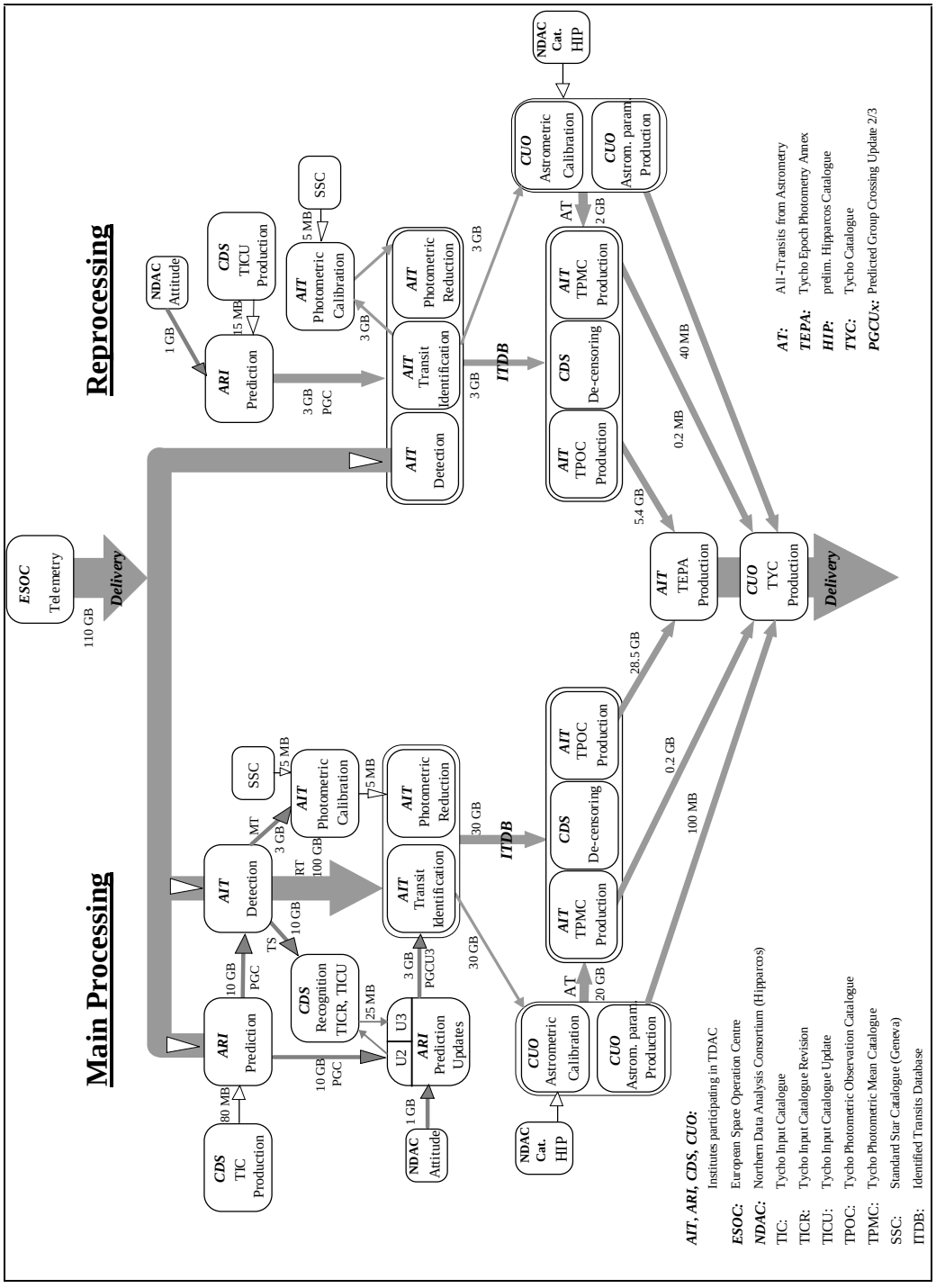

Figure 12.1. *Complete TDAC processing scheme. For a detailed explanation see text.*

The box containing the three processes: detection, transit identification, and photometric reduction indicates the processing pipeline established at Tübingen in order to reduce input/output from and to tape.

12.3. Sorting and Catalogue Preparation

There is another, somewhat hidden, division in Figure 12.1, namely between the transit (time) and the star (coordinate) domain of the TDAC data flow. This division is not particularly relevant in terms of astronomy, but it implied a very demanding data reorganisation, consisting mainly of a large degree of re-sorting. This reorganisation is indicated by the small gap between the upper and the lower part of Figure 12.1, connected only by the arrows labelled ITDB.

The thin diagonal arrows running towards the astrometric calibration box represent a reformatted version of the ITDB, also of 30 Gigabytes, delivered from Tübingen to Copenhagen. Due to the rather different output requirements of the astrometric and photometric branches of the Tycho data reduction, the sorting took place at very different stages of the processing. Astrometry was an iterative process where the astrometric parameters of the stars (and their errors) were consecutively updated as new observations were treated. Thus the individual observations were used during the first processing steps only. Photometry, on the other hand, was supposed to produce the Tycho Photometric Annexes, containing individual observations for selected groups of stars. Thus the photometry process had to keep on-line most of the information belonging to the individual transits. The ITDB was a strictly time-sorted data base of transit records, but it already contained a pointer structure for every star, tying together all its transits. The physical sorting of the records took place within the box labelled 'TPOC Production'. This box actually represents a complex processing scheme, which is detailed in Chapter 11 (see Figure 11.5).

All-Transits Updating

The arrows labelled 'AT' (All-Transits), leading from the boxes 'Astrometric Calibration' to the photometry blocks should be discussed here because they are mentioned only briefly in Sections 7.4 and 11.4.

Astrometric residuals are the most reliable means to decide whether a transit is true or false or parasitic. The de-censored magnitudes and all transits contained in the Tycho Epoch Photometry Annexes should thus be based on a transit data base selected by astrometric criteria. Moreover, the separately derived photometric and astrometric values in the final catalogue should be based on a common set of transits (the transits actually used by Tycho photometry and astrometry are subsets of such a common set, but not necessarily exactly identical). Since astrometry was carried out at Copenhagen and photometry at Tübingen, a special data stream was created to tell photometry which transits had been used by astrometry for the derivation of the astrometric parameters for each star. This data stream was called AT ('All-Transits') since it contained almost all transits, not only the accepted ones. Each transit was accompanied by flags and quantities indicating the acceptance level, and the reasons for rejection, if any.

This data stream was quite different from all the others defined within TDAC, because it was designed more than five years after the other ones, using hardware and software tools which had become available in the meantime. It contained floating point values and flags originating in a certain hardware architecture. Moreover the data was compressed and archived using the standard ZIP tool, version 2.0.1, available on a wide variety of computing platforms. The data were sent on the small EXA-byte tapes, each containing more than ten times as much data as the 160-Megabytes magnetic tapes used in all other TDAC data exchange. The advantages of this procedure were obvious:

- the compression resulted in a very small number of tapes, and thus facilitated the possibility of running a huge stretch of data from a single tape without human interaction;
- a single ZIP archive file contained a large number of data files, thus the data handling was much easier;
- there was no need to write dedicated input/output routines in order to code/decode the data (the transition from the IEEE format to the VAX format was carried out using available standard routines).

The all-transits updating was a big processing step since it was necessary to identify every record contained in the all-transits files with exactly one record in the Tycho Photometric Observation Catalogue, and to rewrite the identified transit records in the Tycho Photometric Observation Catalogue, updated with some of the all-transits information.

12.4. Catalogue Production

The actual production of the catalogue was separated from the other steps, because the data contained in the final catalogue were merged from a number of different sources and the format of the catalogues is quite different from what was used inside TDAC. In Figure 12.1 the complex processes of Tycho Catalogue and Photometry Annex production are merged into a single box each. The individual processing steps belonging to them are detailed in Chapter 11.

A. Wicenec