MANAGEMENT OF ASTRONOMICAL DATA AT KANAZAWA DATA CENTER

Yoichi Terashita
Kanazawa Institute of Technology, Kanazawa 921, Japan

ABSTRACT. The methods of controlling and disseminating astronomical data at the Kanazawa Data Center (operated under the Computer Center of the Kanazawa Institute of Technology) are described. Discussions include the coded data descriptions used in conjunction with a file processing language. Problems concerning remote data access are also discussed, and the future plan presented.

1. INTRODUCTION

Since 1978, the Computer Center of the Kanazawa Institute of Technology (KIT) has been engaged in the task of disseminating machine-readable astronomical data to the astronomical community in Japan. The basis of the data service is a collection of data files (star catalogues) supplied by the Strasbourg Stellar Data Center (CDS), but locally produced data files are also being added. The current holdings at KIT is about 200 catalogs (about 300 data files if subfiles are counted). The main task so far has been to supply the user with duplicates (on magnetic tapes) of files, or with their printed versions. A scheme is being developed and experimented, such that the user can access and manupilate the data files through an on-line terminal.

These data files naturally differ in data items involved, in quantitative units, in recording format, and in descriptive symbols. While this type of diversity could be bearable when they are to be handled manually, it brings about a major difficulty in the case of computer processing. It was necessary, therefore, to develop a data management system specialized for astronomical data, by means of which one can handle data files with the least possible degree of complication.

In developing such a system, the emphasis has been in the application

C. Jaschek and W. Heintz (eds.), Automated Data Retrieval in Astronomy, 21–25.

of data management technologies that are currently available. We have been particulaly interested in those languages specialized for data retrieval and file processing. Some of the so-called data base management systems (DBMS) have also been examined and tested, but these are not discussed here.

2. ASTRONOMICAL DATA MANAGEMENT SYSTEM

Figure 1 shows the overall structure of the Astronomical Data Management System (ADMS) developed at the KIT Computer Center. Central to its design are three sets of system files that contain control information on the data files. The data directory (DD) contains the most basic pieces of infomation on the individual data files, such as the file number, file name, record length, number of records, and the storage infomation. The human-readable data description (HDD) contains the information on the contents of a specitic data file, that is, the position of each data item in the record, its description, unit, and other supplementary information. In addition, each data item is assigned a mnemonic code (field name), with which one can compose a file processing job, as described in the below. The HDD is prepared for each data file, and its direct print-out becomes a user reference manual for the file.

The coded data description (CDD) is obtained from the corresponding HDD through a conversion routine. A CDD defines, for each data item in a given file, the field name (the same as that assigned in the HDD), the location in the record, and the type (numeric or character string), in accordance with the syntax of the language that has been selected as the standard file processing language of the system. We can select any language as long as it is so designed that the data definition part of a program can be separately prepared. At present, a language called EASYTRIEVE is used in ADMS.

Functions of ADMS are summarized in Figure 1. When the user wishes to work on a data file(s), he feeds in the file number through an on-line terminal. The system then retrieves the relevant storage information for the file from DD (Step 1) and incorporates the corresponding CDD (Step 2). The user then types in EASYTRIEVE instructions, consulting the printed version of HDD and using the field names defined there (Step 3). Normally this amounts to only a few lines of input, since only procedural specifications are necessary. After the job is activated and processes the designated file (Step 4), the resulting output is returned to the terminal (Step 5). Figure 2 shows a sample session involving the Yale Catalogue of Bright Stars (file number C50021). Lines with lower-case letters are those fed in by the user. In this example, the job is to list the magnitude, position, and name for the six brightest stars with declination greater than +40 degrees.

Although it would certainly be desirable to keep all the data files

https://doi.org/10.1017/S0252921100082610 Published online by Cambridge University Press

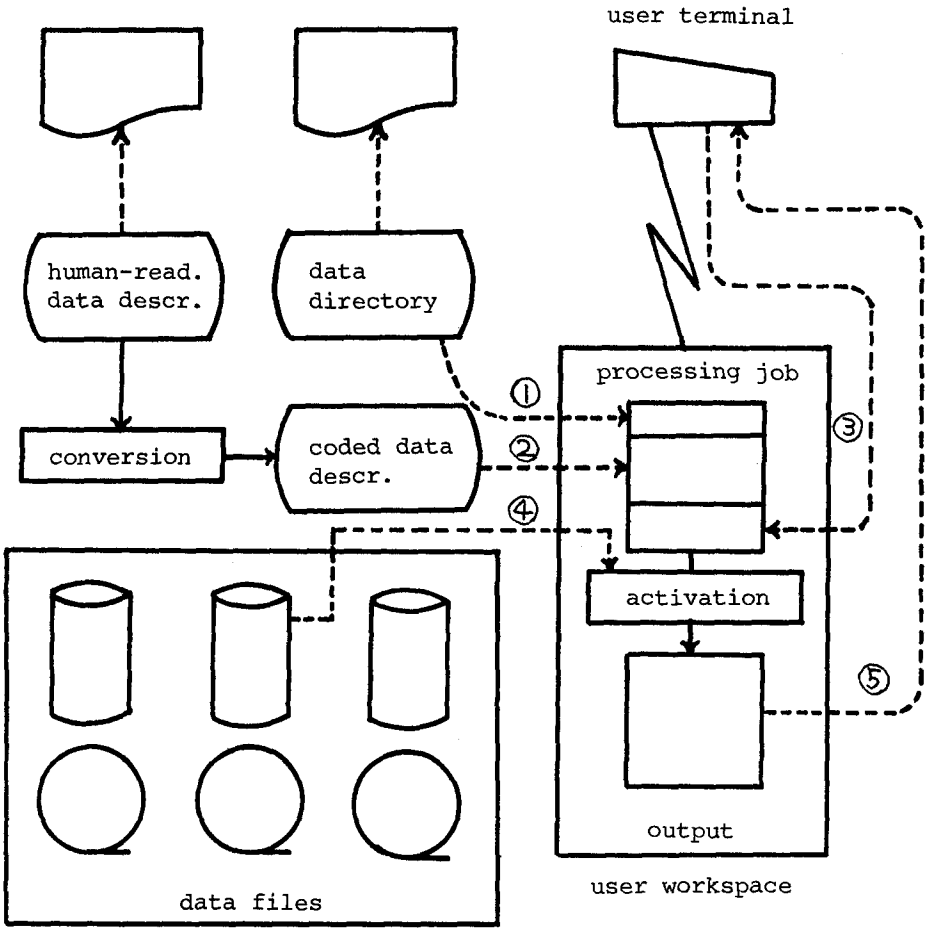

Figure 1. Structure of the Astronomical Data Management System.

on-line, only about half of them are currently resident on disks, for the reason of economy. Nevertheless, files on off-line tapes can still be accessed in the same fashion, though, in this case, assistance from the computer operator is required.

3. FILE PROCESSING LANGUAGES

While almost any computer language can be used to process high-volume data files such as star catalogs, the amount of programming effort for a given task varies depending on what language one should choose. Thus, it can safely be said that COBOL and PL/I are more useful than FORTRAN and ALGOL. However, the former languages are not necessarily the best. There are now available a certain number of languages specifically designed for file-processing applications. Unlike

https://doi.org/10.1017/S0252921100082610 Published online by Cambridge University Press

```
.gastro c50021 example
*** DISK OR TAPE ? ('D' OR 'T')
.d
*** CATALOG C50021 READY FOR PROCESSING ***
INPUT:
.9999
.if del20des gt 40
.sort mv
.list lj limit 6 seq alp20 del20 mv name
.
*** SUBMIT ?
.yes

.elist d

 7/01/81
  SEQ     ALP20     DEL20       MV          NAME
  1708     51641     46  0        9       1301   AUR
  7924    204126     45 16      126       5001   CYG
  4905    1254 2     55 57      176       7705   UMA
  1017     32420     49 51      179       3301   PER
  4301    11 344     61 45      179       5001   UMA
  5191    134732     49 19      186       8507   UMA
          6   RECORDS PRINTED
```

Figure 2. Sample session using ADMS.

general-purpose programming languages such as COBOL and PL/I, these languages offer simplified sets of instructions in such a way that one no longer needs to specify details of operations (the non-procedural feature). This feature is especially noticiable for the looping structure, sort operation, and output editing: that is, the operations always needed in file processing jobs.

EASYTRIEVE, which is currently used in ADMS, is one of such languages. Before this one, we used GIS, another (and one of the earliest) of them (see Ref.1). FOCUS is one of the more recently developed, which we are interested in but has not yet been tested. Statistical languages, such as SPSS and SAS, can also be used for file processing, but their direction is somewhat different and they usually lack in efficiency (we plan to connect SAS to ADMS anyway, since statistical processing is, as a matter of fact, highly demanded in astronomical works).

These file processing languages are not without their problems. Their main purpose has been for business applications. Thus, FORTRAN-like numeric types are not popular among them (standard data types are packed- and zoned-decimals). In business applications it is usually taken for granted that data files are well organized and stored in accordance with established standards, while this is not the case for astronomical (or scientific in general) data files. In the latter

https://doi.org/10.1017/S0252921100082610 Published online by Cambridge University Press

case, it is often necessary to perform operations involving two or more files concurrently, while such operations are not always easy with the languages mentioned here.

Although it is certainly advantageous to use such languages, our conviction has been that there should be file-processing languages that are more scientifically oriented than those now available. As an attemt, one such language is being developed at the KIT Computer Center.

4. DATA COMMUNICATIONS FACILITIES

The means of data dissemination at Kanazawa has been predominantly off-line, that is, magnetic tapes and printed outputs (produced under ADMS). On-line access to Kanazawa data files by way of public telephone lines has been experimented, but the results are not very encouraging. One reason is the quality of the line, which exhibits considerable variations depending on the geographical location of the caller. Another is the cost for such a connection (charged as a usual long-distance call).

The only solution to these problems seems to be to use a dedicated digital communications network. In Japan, the DDX system, a packet-switching digital network developed by the Nippon Telephone and Telegraph Corp. (NTT), is now available at certain areas of the country, and is expected to cover the Kanazawa area in the near future. An interface to the DDX-based N-1 Network (a nation-wide inter-university computer network) has been developed at the KIT Computer Center, and is being tested.

We are indebted to colleagues at CDS for their continuing assistance and encouragements.

REFERENCE

1) Terashita, Y. 1981, in 'Scientific Information Systems in Japan' (Ed. H. Inose), pp.164-169, North-Holland Publ. Co., Amsterdam.