USAGE OF MANIPULA AT THE NETHERLANS CENTRAL BUREAU OF STATISTICS

L.P.M.B. Hofman
Netherlands Central Bureau of Statistics, Heerlen

1. Introduction

In October 1989 the CBS started the development of the general file manipulation program Manipula. One of the reasons for developing this program was the shift of the statistical production process from the mainframe computer to the local area network (LAN). On the LAN no general file manipulation program was available. At some departments a new Pascal program was written for each file manipulation. Some departments tried to develop their own manipulation programs. Everybody wrote their own sorting program. This was of course not very efficient and it did not contribute to the CBS standardization. For efficient implementation of the statistical production process on the LAN at the CBS a general program was a necessity. One of the goals of such a program should be to relieve the user of all the extra around the manipulations. The user must be able to concentrate on his manipulations without having to worry for instance about opening and closing files.

In February 1990 a test version was installed on the LAN. Based on the comments and wishes of the users a lot of changes and extensions were made and in November 1990 version 1.0 was released. One of the important features of this version was the possibility of reading data stored in Blaise files directly. In 1991 a version was released for the mainframe computer. This version has roughly the same possibilities as the LAN version.

The usage of the program at the CBS is widespread. Because the program is so versatile, a lot of different (and for the makers of the program) unexpected applications can be found. Manipula is also used in situations for which it was not primarily intended, for instance tabulation and

checking. There is some overlap between Manipula and other tools in the Blaise system and there even is some overlap between Manipula and Blaise. In practice Manipula has proved to be a jack-of-all-trades. By combining different kind of usages of Manipula and other standard tools in the Blaise system complex statistical production systems can be and have been built.

It would have been possible to describe in this paper one of the statistical production systems which makes use of Manipula, but that is really a task for that system's author. In this paper a different approach is chosen. Some examples are given of the kind of usage of Manipula encountered by the developers of Manipula during consulting and support at the CBS. The following kinds of usage will be described briefly: preparation for Abacus, conversion from ASCII to Blaise, conversion from Blaise to ASCII, reports, tabulation, linking data, manipulation of several input files in one run and checking data.

2. Usage of Manipula

2.1. Preparation for Abacus

Sometimes a table has to be produced from data not directly suited for tabulation. Abacus however offers only a limited number of data manipulation functions, for instance only one-dimensional recodings are possible. Manipula has proved to be very handy and powerful if the manipulations are too complex for Abacus. Manipula can prepare the file to be tabulated with Abacus.

2.2. Conversion from ASCII to Blaise

The CBS receives a lot of data in machine-readable form (on tape, diskette, by telephone). Blaise is often used for further editing. But how to get the ASCII data in the right conversion format? Manipula can do this job fairly easily. With the Setup generator a Manipula description of the ASCII import file can be generated. This description is used in the output

paragraph in the Manipula setup. A description of the received file is always available (unless something is very wrong). This description is translated into a Manipula file description and used in the input paragraph. The manipulate paragraph offers possibilities to compute new variables, make a selection and so on. It is also possible to combine several input records into one output record or to create several output records from one input record. This can be important in case subfiles are used. If there is no unique key available in the input record it can be computed, for instance by using the record number.

2.3. Conversion from Blaise to ASCII

Manipula makes it possible to access data stored in a Blaise file directly. It is the only tool in the Blaise system that gives the user full control over the Blaise data. Because Manipula does not use the meta-data of a questionnaire the user has to take care of a number of aspects normally taken care of by the Blaise conversion program Convert. An example of this is the translation of the 'don't know' and 'refusal' codes. If the questionnaire is small (which is often the case) or if the codes are not accepted by the data entry program no special provisions are necessary. In spite of the possibly negative sides of Manipula as a tailored conversion tool it is used a lot for this purpose. An example is the Labour Force Survey (a huge questionnaire!). During the conversion process of this survey all open answers are replaced by a code which indicates whether an answer has been given.

2.4. Reports

It is very easy to display the contents of records or parts of records in a printable ASCII file with Manipula by using a print paragraph. In the Manipula setup headers and footers can be defined which are displayed on a page together with the contents of a number of (selected) records. The headers and footers need not be fixed text but can contain variables. In this way the user has full control over the page contents.

Examples of this usage can be found at various places. For instance, Manipula is used to generate reports with visit instructions for different field employees in the administrative application described in Lammers (1992).

2.5. Tabulation

In spite of the presence of Abacus, Manipula turned out to be used for tabulation purposes also. Sometimes the layout of a table is prescribed by an external client. An example of such a client is EUROSTAT, the statistical agency of the EC. For some surveys EUROSTAT provides a detailed descriptions of the layout of the tables. With Abacus these tables cannot always be made. Using sorting and summing, plus a print paragraph the job usually can be done with Manipula. Although it takes longer to produce a table with Manipula then with Abacus there is no limit to the size of the table that can be produced. Examples of such tables are the Energy Tables of the Foreign Trade Statistic. Descriptions of the different codes used in the tables are usually obtained by linking data files.

2.6. Linking data

Manipula offers an easy way of linking data files of different sources and of different types. File types supported for linking are Blaise, ASCII and Index (created with the external tool of the Blaise system). It is very easy to instruct Manipula to look up data in such a link file. The link data files can be used for different purposes: recoding data, joining files, for making a selection and so on. Linking data files is also used for partial update of a Blaise form. In a Manipula setup an ASCII file with update values of some fields in a Blaise form is linked with the Blaise data file. The output file contains the completely converted Blaise form with the updated values. This form is imported in Blaise by using the conversion tool.

2.7. Manipulation of several input files in one run

By using a wildcard in the input filename Manipula offers the possibility of carrying out the manipulations on all data files which conform to the wildcard in one run. An example is the situation where data is stored in several files, for instance for each week a different file, and a report has to be made concerning all files. The setup which can be used for one file can also be used for all files simultaneously.

2.8. Checking

Manipula offers the possibility of carrying out various checks on a data file. Checks can be made on the level of a data field (is it numeric?) but they can also be relational checks within one data record. It is also possible to carry out checks accross several input data records. We have a number of examples where Manipula is used as a checking tool. In many cases it would also have been possible to do the checking with an integral check in Blaise CADI.

Manipula is used for checking ASCII files. If the checks are simple this is sometimes more efficient then using Blaise because data does not need to be converted to the Blaise format. But Manipula is also used to do extra checks on data stored in Blaise files. In this case a report is generated with the keys of the cases that need to be examined interactively. There is also an example of a complex check program implemented in Blaise on the LAN. The same checks are carried out on data stored on the main frame using Manipula.

References

Lammers, J.P.M., 1992, The use of Blaise for administrative purposes. Proceedings of the First International Blaise Users Meeting, 14-16 October, Netherlands Central Bureau of Statistics, Voorburg.