Development of a national ground-water data base

Jürgen Kirchner*, Jeff W Morris and VE Cogho
Institute for Ground-water Studies, University of the Orange Free State, P.O. Box 339, Bloemfontein 9300, South Africa.

Abstract

A data base has been developed in which more than 200 different parameters per site can be stored in 33 data sets. Communication is interactively through terminals via the Transparent Government Network or dial-up. Software is developed for data processing on the mainframe. Special geohydrological processing is done on IBM-compatible microcomputers for which programs have also been written. The system is explained and examples of the processed data are given.

Introduction

For the identification of ground-water resources, the development of new water supply schemes and the siting of boreholes, it is necessary to utilise all relevant ground-water data if manpower and funds are to be used in the most efficient way. The availability of a wide range of data also facilitates research and reduces its cost. The generation and collection of these data are very expensive as it requires the drilling of boreholes and field observations. At present a great deal of these data are not formally recorded. Therefore the creation of a ground-water data base where all these data can be stored, was stated as one of the priority tasks in the Master Plan for Research and Development Requirements in Geohydrology (Water Research Commission, 1981).

After preliminary studies of other existing ground-water data bases, mainly the United States Geological Survey (USGS, 1980), the Australian one and others (Unesco, 1983; Bundesanstalt, undated), and in consultation with the Department of Water Affairs (DWA), the Institute for Ground-water Studies (IGS) at the University of the Orange Free State, under contract to the Water Research Commission (WRC), started the development of the data base on the mainframe computer of the Department of Agriculture in Pretoria in 1984.

As the National Ground-water Data Base is now in operation and the IGS and the Directorate of Geohydrology, DWA, have started entering data into the data base, it is appropriate to introduce it to the profession. On the following pages the structuring of the data base, the communication with it and the further processing of the data are discussed.

The data base

Structure

At the beginning of the project it was decided that the data base should be point-orientated, that it should contain the original data and that it must be machine-compatible with other relevant data bases. It was desirable to make the scope of parameters to be accommodated in the data base as wide as possible to comply with the needs of a maximum number of users. It was necessary, on the other hand, that valuable historical data which are generally limited in accuracy, compared to what can be put in the data base, could be accommodated too and, finally, disk-space requirements and processing times also had to be considered.

With this in mind the parameters to be included in the data base were determined. In order to reduce the disk space necessary for the storage of data they were grouped in 32 data sets. These data sets comprise the items numbered between 10 and 119 in Figure 1.

The basic information about a site is contained in the data set ENTRY. Of the 28 fields provided there (see Table 1), at least the allocated identification code, the code of the map on which the site is situated, latitude, longitude, the accuracy of the position, the drainage region, the site type and the reporting institution must be given.

The remaining data sets consist of headers like LOGS, CONSTRUCTION etc. and repeating groups, e.g. 'Penetration rate', 'Aquifer' and 'Geology', which resort under their respective headers and which can be added when required.

Only one ENTRY data set is allowed for each site. Headers and repeating groups, however, may be repeated. Headers contain general information which applies to all repeating groups which are to be stored under them. The validity of other new data in the various sets, e.g. dates and water-levels, is checked against information already stored in the data base. Certain constructional information in the data set ENTRY is updated as more recent information is added in the repeating groups under the header CONSTRUCTION. Because most of the data are stored in repeating groups, it is possible to build up historical records for the different sites.

As stated above, the Ground-water Data Base can store a wide variety of data which have been grouped into different data sets. All the parameters that can be accommodated in the data base are shown in Table 1. Apart from the basic information in ENTRY, *geohydrological data gained during drilling* are stored under the header LOGS: i.e. the penetration rate recorded, the thickness and type of the geological strata encountered and the aquifers struck.

Constructional information is grouped under the header CONSTRUCTION. Data regarding the contractor are placed under the header, while details about depths and diameters are stored under the repeating group 'Hole'. Similarly, the information regarding the casing can be entered under 'Casing', information regarding size and distribution of perforations under 'Openings' and information on filter pack or seals placed around the casing under 'Fill'.

Data regarding installation are stored under the header INSTALLATION.

For various *operational data* of a site, provision is made under five headers and repeating groups: DISCHARGE RATE and WATER LEVEL for the recording of time of measurement and measured discharge rate or water level respectively; FIELD MEASUREMENT and METER READING for storage of times and a

*To whom all correspondence should be addressed.
Received 9 October 1986.

TABLE 1
THE PARAMETERS OF THE DATA BASE

1.0 ENTRY
1.0. 1 Site identification code
1.0. 2 Map code
1.0. 3 Number on 1 : 50 000 map
1.0. 4 Province, district and farm number
1.0. 5 Site name
1.0. 6 Topographic setting
1.0. 7 Latitude
1.0. 8 Longitude
1.0. 9 Coordinate accuracy
1.0.10 Drainage region
1.0.11 Altitude
1.0.12 Method of measurement
1.0.13 Site type
1.0.14 Site selector
1.0.15 Nominal diameter
1.0.16 Depth completed
1.0.17 Completion date
1.0.18 Depth information source
1.0.19 Status of site
1.0.20 Purpose of site
1.0.21 Water use: consumer
1.0.22 Water use: application
1.0.23 Reporting institution
1.0.24 Equipment
1.0.25 Potability
1.0.26 Date record entered
1.0.27 Date record updated
1.0.28 Comment

2.0 LOGS
2.0. 1 Site identification code
2.0. 2 Header sequence number
2.0. 3 Reporting institution
2.0. 4 Date record entered

2.1 REPEATING GROUP: PENETRATION RATE
2.1. 1 Site identification code
2.1. 2 Header sequence number
2.1. 3 Repeating group number
2.1. 4 Depth to top of interval
2.1. 5 Depth to bottom of interval
2.1. 6 Diameter of interval
2.1. 7 Penetration rate
2.1. 8 Comment

2.2 REPEATING GROUP: AQUIFER
2.2. 1 Site identification code
2.2. 2 Header sequence number
2.2. 3 Repeating group number
2.2. 4 Depth to top of aquifer
2.2. 5 Depth to bottom of aquifer
2.2. 6 Aquifer code
2.2. 7 Comment

2.3 REPEATING GROUP: UNCONSOLIDATED
2.3. 1 Site identification code
2.3. 2 Header sequence number
2.3. 3 Repeating group number
2.3. 4 Depth to top of interval
2.3. 5 Lithology code
2.3. 6 Primary colour
2.3. 7 Secondary colour
2.3. 8 Texture
2.3. 9 Primary feature
2.3.10 Secondary feature
2.3.11 Feature attribute
2.3.12 Sorting
2.3.13 Roundness
2.3.14 - 20 Sieve analysis
2.3.21 Comment

2.4 REPEATING GROUP: CONSOLIDATED
2.4. 1 Site identification code
2.4. 2 Header sequence number
2.4. 3 Repeating group number
2.4. 4 Depth to top of interval
2.4. 5 Lithology code
2.4. 6 Primary colour
2.4. 7 Secondary colour
2.4. 8 Texture
2.4. 9 Feature
2.3.10 Comment

3.0 CONSTRUCTION
3.0. 1 Site identification code
3.0. 2 Header sequence number
3.0. 3 Piezometer number
3.0. 4 Reporting institution
3.0. 5 Date record entered
3.0. 6 Date of construction
3.0. 7 Name of contractor
3.0. 8 Construction data source
3.0. 9 Method of construction
3.0.10 Type of finish
3.0.11 Method of development
3.0.12 Duration of development
3.0.13 Special development treatment
3.0.14 Cost of construction
3.0.15 Comment

3.1 REPEATING GROUP: HOLE
3.1. 1 Site identification code
3.1. 2 Header sequence number
3.1. 3 Repeating group number
3.1. 4 Depth to top of interval
3.1. 5 Depth to bottom of interval
3.1. 6 Diameter of hole

3.2 REPEATING GROUP: CASING
3.2. 1 Site identification code
3.2. 2 Header sequence number
3.2. 3 Repeating group number
3.2. 4 Depth to top of casing interval
3.2. 5 Depth to bottom of casing interval
3.2. 6 Diameter of casing interval
3.2. 7 Casing material
3.2. 8 Wall thickness of casing
3.2. 9 Comment

3.3 REPEATING GROUP: OPENINGS
3.3. 1 Site identification code
3.3. 2 Header sequence number
3.3. 3 Repeating group number
3.3. 4 Depth to top of open interval
3.3. 5 Diameter of open interval
3.3. 7 Material in this interval
3.3. 8 Type of openings in interval
3.3. 9 Length of openings
3.3.10 Width of openings
3.3.11 Horizontal distance between openings
3.3.12 Vertical distance between openings
3.3.13 Method of openings made

3.4 REPEATING GROUP: FILL
3.4. 1 Site identification code
3.4. 2 Header sequence number
3.4. 3 Repeating group number
3.4. 4 Depth to top of interval
3.4. 5 Depth to bottom of interval
3.4. 6 Type of fill
3.4. 7 Comment

4.0 INSTALLATION
4.0. 1 Site identification code
4.0. 2 Header sequence number
4.0. 3 Reporting institution
4.0. 4 Date record entered
4.0. 5 Date of installation
4.0. 6 Type of installation
4.0. 7 Depth to pump intake
4.0. 8 Type of power
4.0. 9 Power rating
4.0.10 Manufactor
4.0.11 Serial number
4.0.12 Power meter number
4.0.13 Monitoring facility
4.0.14 Installation data source
4.0.15 Comment

5.0 DISCHARGE RATE
5.0. 1 Site identification code
5.0. 2 Header sequence number
5.0. 3 Reporting institution
5.0. 4 Date record entered
5.0. 5 Type of discharge
5.0. 6 Method discharge measured
5.0. 7 Discharge information source

5.1 REPEATING GROUP: DISCHARGE RATE
5.1. 1 Site identification code
5.1. 2 Header sequence number
5.1. 3 Repeating group number
5.1. 4 Date of measurement
5.1. 5 Time of measurement
5.1. 6 Discharge rate
5.1. 7 Comment

6.0 WATER-LEVEL
6.0. 1 Site identification code
6.0. 2 Header sequence number
6.0. 3 Piezometer number
6.0. 4 Reporting institution
6.0. 5 Date record entered
6.0. 6 Method of measurement
6.0. 7 Water-level status
6.0. 8 Collar elevation
6.0. 9 Water-level information source

6.1 REPEATING GROUP: WATER-LEVEL
6.1. 1 Site identification code
6.1. 2 Header sequence number
6.1. 3 Repeating group number
6.1. 4 Date of measurement
6.1. 5 Time of measurement
6.1. 6 Water-level
6.1. 7 Comment

7.0 WATER SAMPLE
7.0. 1 Site identification code
7.0. 2 Header sequence number
7.0. 3 Reporting institution
7.0. 4 Date record entered

7.1 REPEATING GROUP: WATER SAMPLE
7.1. 1 Site identification code
7.1. 2 Header sequence number
7.1. 3 Repeating group number
7.1. 4 Date of sampling
7.1. 5 Time of sampling
7.1. 6 Sample number
7.1. 7 Comment

8.0 FIELD MEASUREMENT
8.0. 1 Site identification code
8.0. 2 Header sequence number
8.0. 3 Repeating group number

8.1 REPEATING GR. FIELD MEASUREMENT
8.1. 1 Site identification code
8.1. 2 Header sequence number
8.1. 3 Repeating group number
8.1. 4 Date of measurement
8.1. 5 Time of measurement
8.1. 6 Parameter measured
8.1. 7 Reading
8.1. 8 Comment

9.0 METER READING
9.0. 1 Site identification code
9.0. 2 Header sequence number
9.0. 3 Reporting institution
9.0. 4 Date record entered
9.0. 5 Type of meter read
9.0. 6 Comment

9.1 REPEATING GROUP: METER READING
9.1. 1 Site identification code
9.1. 2 Header sequence number
9.1. 3 Repeating group number
9.1. 4 Date of reading
9.1. 5 Time of reading
9.1. 6 Reading
9.1. 7 Comment

10.0 PUMPING TEST
10.0. 1 Site identification code
10.0. 2 Header sequence number
10.0. 3 Reporting institution
10.0. 4 Date record entered
10.0. 5 Date pumping test started
10.0. 6 Method tested
10.0. 7 Depth to pump intake
10.0. 8 Recommended abstraction rate
10.0. 9 Transmissivity
10.0.10 Storativity
10.0.11 Comment

11.0 MISCELLANEOUS
11.0. 1 Site identification code
11.0. 2 Header sequence number
11.0. 3 Reporting institution
11.0. 4 Date record entered

11.1 REPEATING GR.: SITE SELECTION
11.1. 1 Site identification code
11.1. 2 Header sequence number
11.1. 3 Repeating group number
11.1. 4 Description and recommendation

11.2 REPEATING GROUP: OWNER
11.2. 1 Site identification code
11.2. 2 Header sequence number
11.2. 3 Repeating group number
11.2. 4 Ownership date
11.2. 5 Name of owner

11.3 REPEATING GR. OTHER IDENTIFIER
11.3. 1 Site identification code
11.3. 2 Header sequence number
11.3. 3 Repeating group number
11.3. 4 Other identifier
11.3. 5 Assignor

11.4 REPEATING GROUP: OTHER DATA
11.4. 1 Site identification code
11.4. 2 Header sequence number
11.4. 3 Repeating group number
11.4. 4 Data type
11.4. 5 Type of logger
11.4. 6 Location
11.4. 7 Format

11.5 REPEATING GROUP: SITE VISITS
11.5. 1 Site identification code
11.5. 2 Header sequence number
11.5. 3 Repeating group number
11.5. 4 Date of visit
11.5. 5 Visitor

11.6 REPEATING GROUP: SPECIAL CASES
11.6. 1 Site identification code
11.6. 2 Header sequence number
11.6. 3 Repeating group number
11.6. 4 Number of hole etc. in group
11.6. 5 Depth of the deepest hole
11.6. 6 Depth of the shallowest hole
11.6. 7 Method of holes constructed
11.6. 8 Diameter of group
11.6. 9 Length of pond/tunnel/drain
11.6.10 Strike of pond/tunnel/drain
11.6.11 Dip of tunnel or drain
11.6.12 Depth of lateral collector
11.6.13 Length of lateral
11.6.14 Mesh of screen in lateral

11.7 REPEATING GROUP: REFERENCES
11.7. 1 Site identification code
11.7. 2 Header sequence number
11.7. 3 Repeating group number
11.7. 4 Reference (map/section/report)

11.8 REPEATINGGROUP: SEARCH STATUS
11.8. 1 Site identification code
11.8. 2 Header sequence number
11.8. 3 Repeating group number
11.8. 4 Record type & period searched

11.9 REPEATING GROUP: COMMENTS
11.9. 1 Site identification code
11.9. 2 Header sequence number
11.9. 3 Repeating group number
11.9. 4 Comment

variety of different parameters or different meters respectively. To avoid duplication, the details of the chemical analyses are stored in the chemical data base of the Department of Water Affairs. Under WATER SAMPLE only the date and time of sampling as well as the sample number are recorded. Correlation between the two data bases is by means of the sample number, the identification code of the site and the coordinates.

PUMPING TEST. Deviating from the principle of accommodating only unprocessed data, provision is made to store calculated values for *aquifer parameters* and recommended abstraction rates. The original data which were used to determine these parameters are stored under the operational repeating groups DISCHARGE and WATER LEVEL.

Other data which do not fall under any of the previous headings are grouped under MISCELLANEOUS. They comprise information on *site selection*; *ownership*; *numbers other than the site ID code*; they refer to the kind and whereabouts of *other geohydrological data* which are not site specific; *site visits*; *special cases* like information recorded from tunnels, shafts etc.; *references* to more general information about the area of the site; the *search status*, i.e. information on the sources of the data entered into the data base; and lengthy *comments* which cannot be recorded in the comment of the relevant data set.

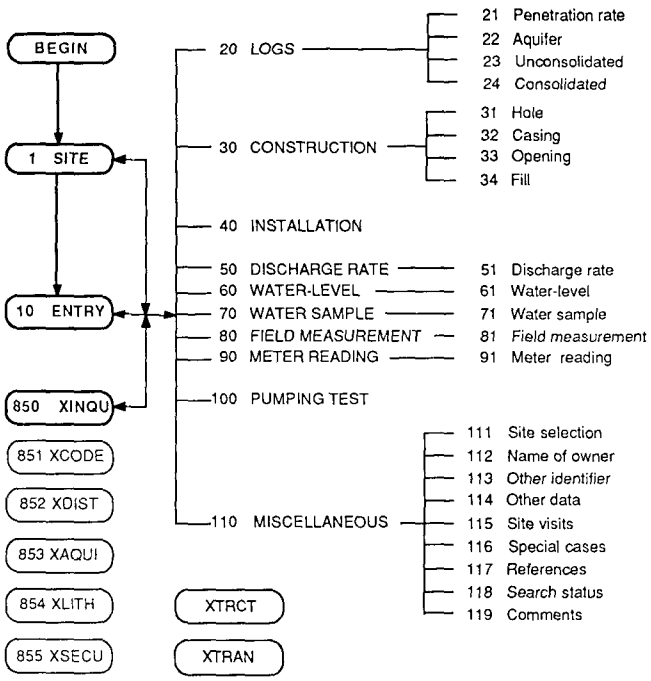


Figure 1
The structure of the Ground-water Data Base.

Communication with the data base

At present data are entered, updated and extracted interactively by terminals connected to the mainframe via the Government Network. For entry, change, deletion and retrieval of data, use is made of screens on the terminal which depict coding forms that have to be filled in. All items shown in Fig. 1 are screens and most of them correspond with the data sets in which the actual values are stored. The BEGIN screen is for logging on; SITE for determination of the site identification code for which information is to be updated. The screen numbered 850 is for inquiries. The higher numbered screens contain various code lists etc. and are, together with the unnumbered screens, for the use of the data base administrator only. Details, describing step by step all the procedures a user must follow to enter, retrieve and process data, are contained in a users manual (Institute for Ground-water Studies, 1986a); the technical specifications are given in a second volume (Institute for Ground-water Studies, 1986b).

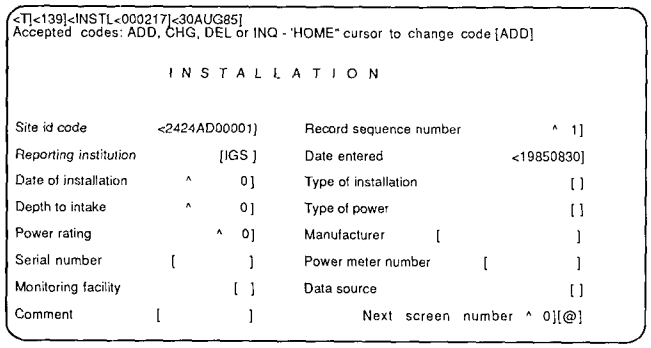


Figure 2
Example of a screen: 40 INSTALLATION.

Data collection, preparation and entry

Data to be entered into the data base must satisfy certain minimum requirements and must of course be presented in the correct format. Under normal circumstances data are entered on the coding forms for the different data sets. Details describing the required format of the data and guidelines regarding the completion of the coding forms are contained in the users manual. A list of the fields and codes of the data base is also available.

Once an authorised user has logged onto the system, he determines the SITE he wishes to access. He then moves from screen to screen by entering transaction codes, data and next screen numbers using the numbers given in Fig. 1. As an example of the layout of the screens, the one for the data set INSTALLATION is shown in Fig. 2.

Retrieval

A variety of options exist for the retrieval of data: when the user wishes to extract information using the data base programs, the inquiry screen XINQU is called up by entering 850 as the next screen number (see Fig. 3).

Here, data are selected according to identification codes or area. The area can be defined by either two coordinate pairs; by province, district or farm or by drainage region. Optionally the extracted data can further be sorted by specifying the site type, e.g. borehole, well, fountain etc. and the completion date.

Data can be printed out in various printed lists or in statistically processed forms, by choosing 'Printer file' from the

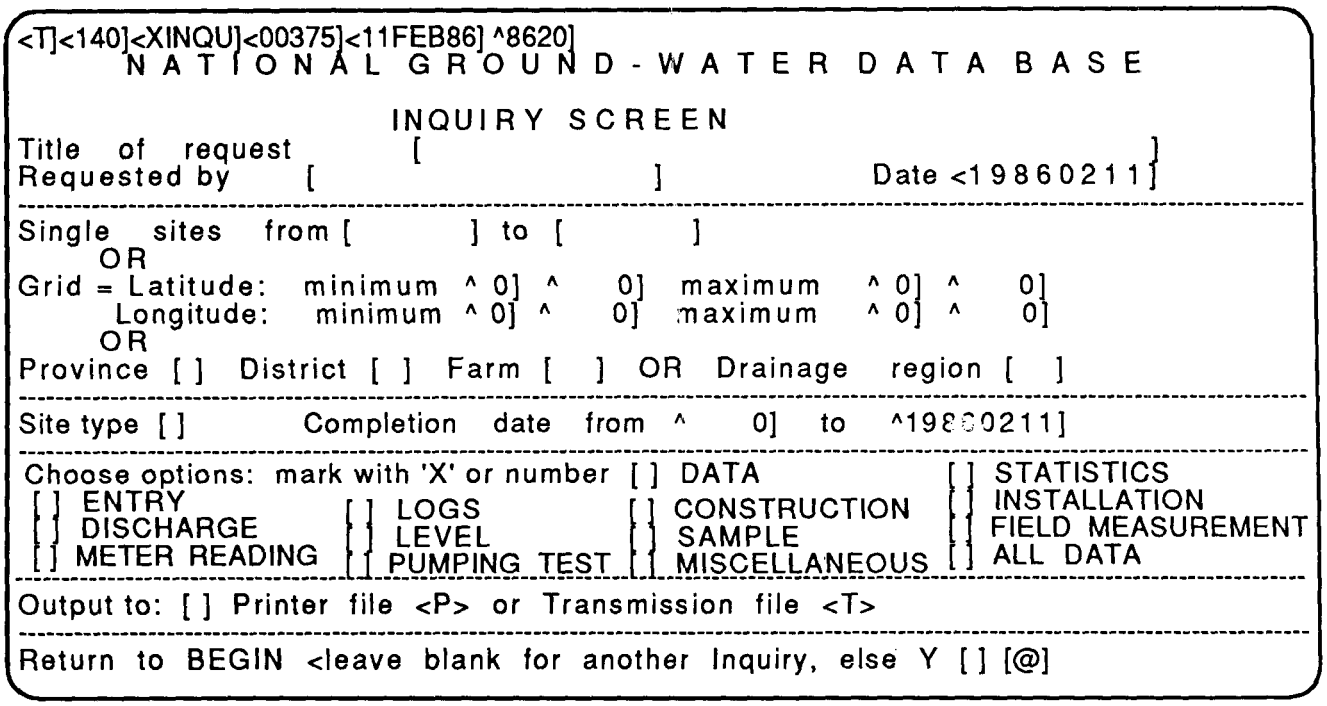


Figure 3
The inquiry screen 850 XINQU.

menu. Printed lists contain tables displaying all data of the data sets selected on the inquiry screen and are sorted according to either coordinates, identification codes, farm numbers or drainage regions.

If the option 'Statistics' instead of 'Data' is chosen one of the following printouts can be obtained:

- the number of entries within the specified area broken down according to 5 accuracy classes of the position;
- the number of ENTRIES; the number of the headers including INSTALLATION and PUMPING TEST and the number of each of the 22 repeating groups within the selected area;
- in addition to the information under point 2 the respective number of repeating group entries per ID code within the selected area; and
- selected information about each site intended amongst others for use of the technician or geologist in the field.

Machine-readable data files can be created by selecting the option 'Transmission files'.

If data are to be selected by other parameters than the options provided by the inquiry screen of the data base, special software like ERGO, INQUIRY and LIRC must be used which is provided on the mainframe.

Software

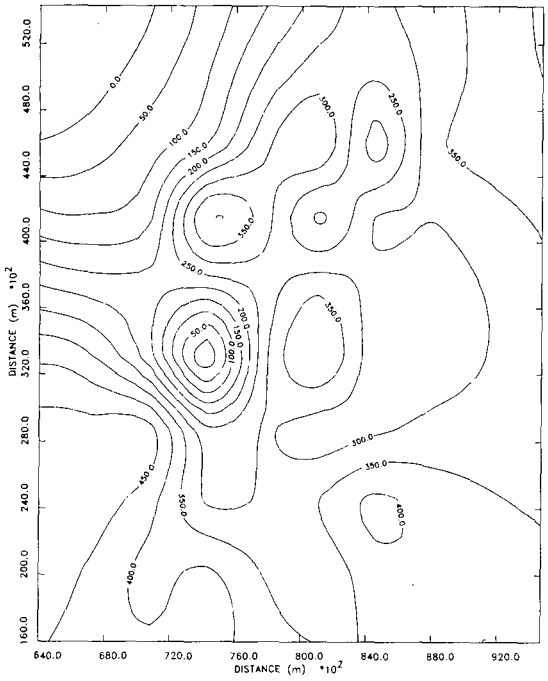
Because of the enormous mass of data to be stored by the system, a large mainframe computer was essential. The National Groundwater Data Base has been created on such a mainframe and is written with the LINC II (Logic and Information Network Compiler) fourth generation system development package. By means of special programs (or inquiry software programs like LIRC, etc.), data can be processed on the mainframe as has indeed been done for the creation of the transmission and printer files.

It has been found, however, that special geohydrological processing of the data is more effectively done on microcomputers. Amongst others, these microcomputers are more user-friendly and a high-quality graphical output can easily be obtained. Waiting times and the cost for data transmission can also be reduced, especially if graphics programs are frequently re-run after output parameters have been changed or data have been further edited. Microcomputers linked to the mainframe are therefore used for the geohydrological processing and presentation of the data.

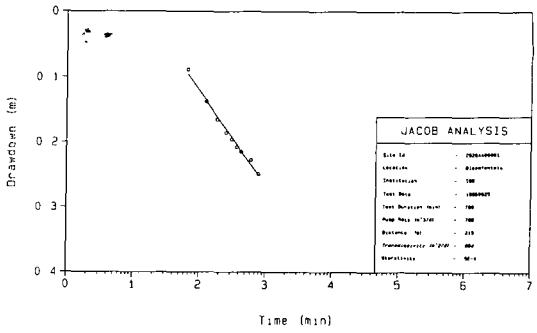
For the geohydrological user the IGS is presently developing software which facilitates the processing of geohydrological and chemical data on the microcomputer. Programs for the presentation of chemical data (Piper, Schoeller, Durov, SAR-diagrams etc.); graphical display of geological and other borehole information in sections, on plans and contour maps; as well as a variety of single and multivariate statistical tests are developed. Some examples of already available programs for graphical presentation of processed data are shown in Fig. 4.

Hardware

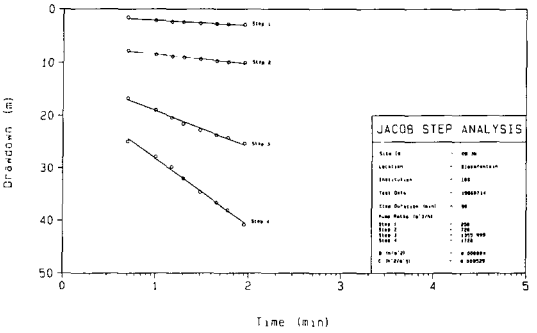
The data base is running on a mainframe computer with the usual peripherals. Terminals, printers and microcomputers are linked either directly or via the Transparent Government Network and concentrators. BURROUGHS' ET 1100 and other compatible products which can emulate the ET 1100 screen as, for instance, the Olivetti M 24, the Qume QVT 109 and ADM 32A terminals are all supported. For those users who are not permanently linked with the data base, dial-up will also be possible, although there may be certain restrictions, e.g. in respect of the speed at which data can be transferred.



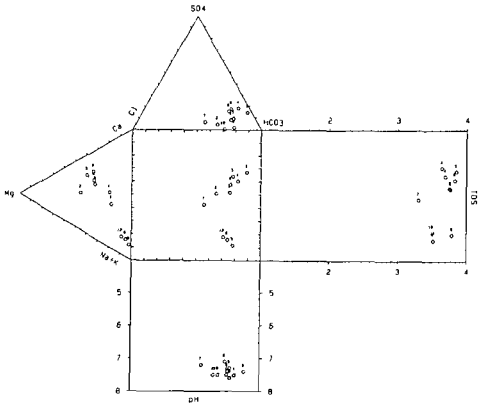
WATER-LEVEL CONTOURS



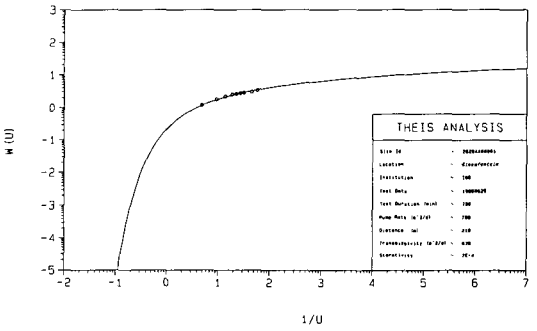
PUMPING TEST : JACOB



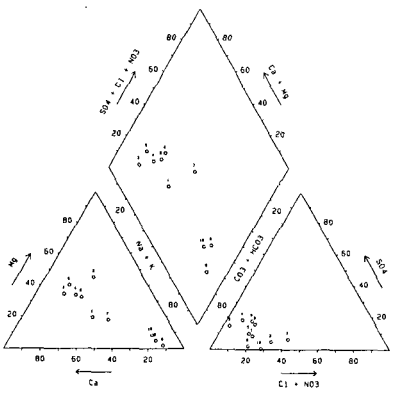
PUMPING TEST : STEP TEST



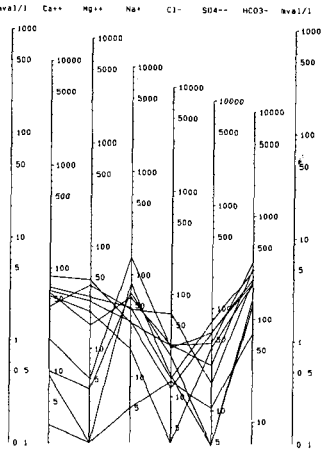
DUROV DIAGRAM



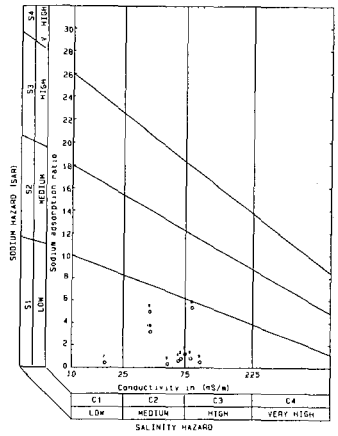
PUMPING TEST : THEIS



PIPER DIAGRAM



SCHOELLER DIAGRAM



SAR DIAGRAM

Figure 4
Examples of graphical displays provided by the Data Base.

For local processing IBM compatible (Olivetti M 24) and HP 17 microcomputers with printer and plotter attachments have been used so far.

Ongoing development

In the near future it will be possible to extract data from the data base also via the newly founded Computer Centre for Water Research in Pietermaritzburg.

While the data base is already fully in operation, the IGS is presently busy rounding off the user programs and updating the manuals. This will be completed at the end of this year. A report which demonstrates the abilities of the data base, using the provided software to process the stored data is in preparation and will be available by the end of the year. Future aims could include batch loading facilities, the refinement and extension of the existing software and possibly the addition of further software modules, e.g. extended CAD (computer aided design) facilities for preparation of maps, etc.

Conclusion

The National Ground-water Data Base is a national facility. It has been completed and is available for the use of geohydrologists, as well as other professionals working in the field of ground-water development, supply and research for the government as well as the private sector. Potential users are encouraged to use the new data base and are invited to contact either the Directorate of Geohydrology in Pretoria or the IGS for further details.

Acknowledgements

The authors wish to thank the Water Research Commission for the funding of this project and are indebted to the numerous professionals from both the Directorate of Geohydrology and the private sector for suggestions and for critical reviews of the work carried out so far. This paper is published with the permission of the WRC.

References

BUNDESANSTALT FÜR BODENFORSCHUNG (undated) Coding forms for Well Data Base, personal communication P.E. Groba, Hannover, West Germany.

INSTITUTE FOR GROUND-WATER STUDIES (1986a) National Ground-Water Data Base, Volume 1: User Guide.

INSTITUTE FOR GROUND-WATER STUDIES (1986b) National Ground-Water Data Base, Volume 2: Technical Specifications.

UNESCO (1983) Computer processing of ground-water data (Chapter 8.5) in *Ground-water studies*.

U.S. GEOLOGICAL SURVEY (1980) National Water Data Storage and Retrieval System.

WATER RESEARCH COMMISSION (1981) Research and development requirements in geohydrology.