Weltdatenbank Abfluß
Bundesanstalt für Gewässerkunde
Koblenz, Deutschland

Global Runoff Data Centre
Federal Institute of Hydrology
Koblenz, Germany

Report No. 16

The GRDC Database
- Concept & Implementation -

Johannes Pauler, Thomas Decouet

July 1997

56068 Koblenz, Kaiserin-Augusta-Anlagen 15-17
Phone +49-261-1306-224, Fax +49-261-1306-280
e-mail (RFC 822): grdc@koblenz.bfg.bund400.de
e-mail (X.400): c=de;a=bund400;p=bfg;o=koblenz;s=grdc

FOREWORD

This report has been prepared in response to numerous queries of users and visitors to the GRDC who wished to have more information about the technical aspects of the database system and its functions. Hence, this GRDC-report provides an insight into the structure and main functionalities of the GRDC database system and add-on tools used for routine operation of the database system and sevice for users.

The report serves two main purposes:

- To inform users of the GRDC about the technical working of the database and

- To present material which can be used to develop relational hydrological database systems by providing a basic understanding of the structural principals and functions of the GRDC database system.

In order to maintain an informative character, the report does not attempt to be a hands-on training manual or operation handbook of the database.

It is quite unusal for a data centre to describe in detail its database system which determines to a large degree its operational utility. After four years of operation and improvement of its functions, the GRDC database system has proven its capability to respond to complex user queries, maintain a high level of data security and adaptability to new developments. This is important to note, because many databases operate on a project base and were not necessarily developed for user-friendliness. The GRDC database has been developed using a relational database concept under INFORMIX and off-the-shelf development tools. Its system's mainstay is a UNIX-system with WINDOWS 3.11 and WINDOWS NT user interfaces. There exist probably as many ideas of how to structure and implement a database system as there are programmers and database managers. The report offers the exchange of information from the viewpoint of the GRDC and thus sets a reference mark for discussion and setting up of new databases. While the description of the GRDC database forms the core of this report, the main supporting tools are briefly described to give the reader an impression of the processing and information generation capacity of the GRDC. Other, off-line software tools, such as statistical programmes are not discussed in this context.

By publishing this report, the GRDC solicits comments on possible improvements and also invites institutions interested in setting up a hydrological database to enter discussions with the GRDC - crew. Our staff is happy to assist where possible!

Wolfgang Grabs
Head, GRDC

Koblenz, July 1997

INTRODUCTION

This document introduces the reader to the core facilities of the GRDC database. The description is written in an informative way and is not meant to be a user's guide. The document describes basic aspects about the planning and designing of databases with a relational database model. It also shows the implementation of the GRDC database and the facilities of the Database Retrieval and Administration Tool (GRDC Base System). It is hoped that the reader gets a better understanding of the realization of a scientific database. This report reflects the current status. As the GRDC is an evolving system to respond to user requests, changes in technologies and facilities are made whenever necessary.

The GRDC database was developed using up to date technology and facilities to give the users comfort, to maintain data security and to add-on additional tools in future. Therefore the GRDC database is implemented as a relational databank using an SQL-Server-Architecture on a UNIX server and the GRDC clients on MS-WINDOWS. Besides the databank itself there exist other tools for visualising, checking and manipulating the stored data. Figure 1 below shows the different access modes of these tools to the databank. It also generalizes the dataflow in the GRDC.

Readers should note that trade names mentioned in this document are mentioned as example only and are not indicative of a recommodation on behalf of the GRDC.

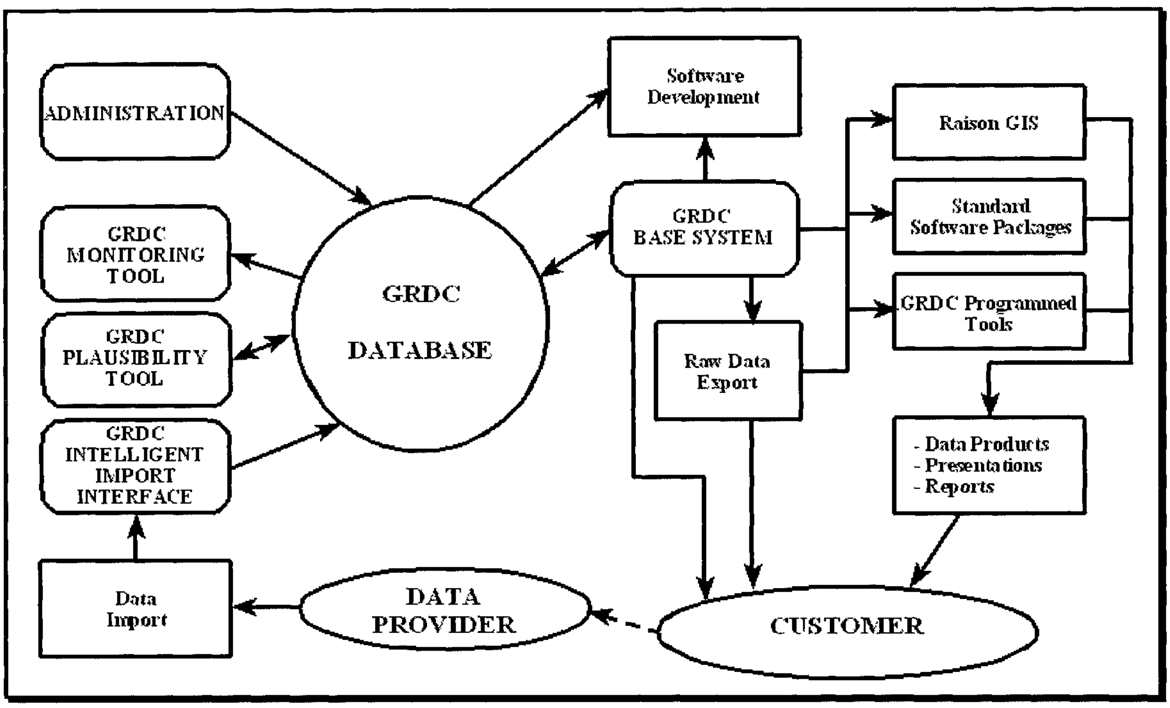


Figure 1: Dataflow and different access modes to the GRDC Database

THE RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS)

GENERAL ASPECTS

Today two different database models are in common use: the hierarchical and the relational model. The hierarchical model is fit for a very large data basis where only few retrieval combinations are meaningful. This is because it is impossible to combine objects defined in different tree paths. So if a new query has to be generated, the complete database structure has to be changed. For example the UNIX file system or large library catalogues are based on this model. The great advantage besides the possibility of storing large quantities of data records is the response time of the queries. This is because the database structure and the query scripts are optimised for speed by the database staff.
Since the late 80's the relational database model displaces more and more the hierarchical model. The systems using this model are capable of managing large data quantities and ensure a high efficiency of the query languages. The greatest advantage is the flexibility of the data handling. These database systems normally work on a UNIX platform.

DATABASE SYSTEMS

The overall purpose of a database system is to maintain central control over a large quantity of data required by many users or programmes. This leads to higher efficiency in work and to better query results for stored data, because

- ☞ the information is stored only once
- ☞ the data is stored in only one location
- ☞ updates are available to all users at once
- ☞ the updating and the administration of the system and data is done by trained staff.

Each database system consists of two main parts:

- ☞ the data basis (all collected and stored data)
- ☞ the data management (software for configuring, controlling and manipulating the data)

The GRDC stores daily discharge data of nearly 2,500 stations out of about 3,300 stations in 1996. This leads to about 400,000 records (an ASCII file of more than 100 Mbytes). Three years before there were only about a fifth of the data available. The use of a state-of-the-art database system is therefore a precondition for speedy updates, queries and user services.
Although it is possible to create a database system with any programming language like C/C++, FORTRAN, COBOL or PASCAL it is highly recommended to use a standard database system such as INFORMIX, ORACLE or IBM DB/2 and for smaller projects e.g. DBASE, PARADOX or MS-ACCESS. This is because these systems have already implemented mechanisms for data security, data access, index organisation and handling. These mechanisms are strongly influenced by the used hardware and operating system.
Nearly all UNIX or VAX/VMS based database systems use SQL (Structured Query Language) as query language. This is a 4th generation language and rather easy to learn and handle. ESQL is a library containing the SQL commands to link them to programs written in a programming language like C/C++ or FORTRAN. This helps the customer to write programs in his preferred programming language and realize the database queries with embedded SQL scripts. The latest standardization of SQL was in 1992 (SQL-92).

DATABASE TABLES

Typically for the relational database system is the grouping of information in a logical way and store these groups in different tables. This prevents data inconsistency and redundancy because data is changed at only one point and is then available all through the system. These tables can be seen as logical units.

GRDC-NO	RIVERNAME	STAT-NAME	LATITUDE	LONGITUDE	AREA
6229500	Vaenern	Vaenersborg	58.38 N	12.32 E	46830
6337100	Weser	Vlotho	52.17 N	8.85 E	17618
2907500	Pur	Samburg	67.08 N	78.15 E	95100

Table 1: Example of a database table

Each table is divided in rows (records) and columns (fields). Each column in a table is of constant length. For example the column GRDC-NO is 7 characters of length, the Column RIVERNAME 40 characters and so. This again leads to a constant length of every row in that table. Now the following problems can arise:

☞ The information for one field is smaller than the column width: In this case disk space will be wasted. The information will be stored correctly, the missing characters will be filled with blanks.

☞ The information for one field is longer than the column width: Here the information will be cut.

Therefore it is necessary to fix a proper column length when designing a table so that neither information will not be cut nor too much disk space will be wasted. The design of the most important tables used in the GRDC database are shown in APPENDIX A

INDICES

Each row in a table is represented by it's primary key. This is a special field or a group of fields. It is not allowed to use one primary key more than once in a table. The primary key field(s) must normally be the first column(s) in the table. The database system software uses these keys to build relations between different tables and for fast querying a table. The primary key field(s) always needs a valid content (NULL values are not allowed). In the GRDC database the GRDC-No. is the content of nearly all primary keys. In addition, all other fields can be declared as secondary keys. These keys don't need to be unique and may be of NULL value.

The system stores the primary and secondary keys and the physical storage addresses of the relating rows in a special table, the so called index. These tables are very useful (and sometimes the only way) for data manipulation. For example, if the above table is queried for a specific river and the column RIVERNAME is not indexed the system will query the table sequentially (this means column by column) until the end of the table. If there exists an index the system will query the index table by using the optimal search algorithm and then directly accesses the fitting rows by using their addresses. The system itself decides which algorithms it will use (self-optimising query language). Care must be taken that not too many secondary key fields are declared for one table and therefore the rows grow too long.

From time to time (after updates or value changes) the database must be re-indexed. By doing this the database tables will be written to the harddisk in a new order and the index tables will be newly defined and generated. The system allocates a certain diskspace for each table. When this diskspace is filled the additional records will be written all around the harddisk. If now a new record is added to the table the index table will be newly sorted and the relative position of the new record in the diskspace is computed from the index and the storage adress. The indices of the records outside the allocated diskspace are only appended at the index table. When a query is started the system first looks in the diskspace by using special algorithms. If it doesn't find the record the remaining indices are looked up sequentielly. By re-indexing the table a new diskspace is computed and allocated and the records outside the former diskspace will be integrated in the system. Another point is to prevent or mend address conflicts in the index tables.

STEPS TO BUILD UP A RELATIONAL DATABASE

The following list shows the different steps for planning a relational database. It is recommended to work out these steps in the order below:

1. defining the attributes (fields) needed (all information that shall be stored)
2. gathering the attributes belonging together in tables (objects)
3. examine the attributes for further subdivision (e.g. name -> prename and name)
4. examine the tables for further subdivision
5. defining field types and length (=> record length)
6. defining primary keys (in special circumstances these keys need to be constructed)
7. defining secondary keys
8. ordering the columns of a table in a logical sense

OUTLOOK: OBJECT ORIENTED DATABASE MANAGEMENT SYSTEMS (ODBMS)

In future the object oriented database models will gain importance. But this change will not be drawn as sharp as from the hierarchical to the relational model in the early 90's, because modern relational systems adopt items from the object oriented concept (e.g. Data Warehouse, object oriented query languages, distributed databases...). At the present there exists no standardization of ODBMS and compatibility to other standards is not established. The table below shows the principle characteristic differences between relational and object-oriented database systems.

Relational DBMS	Object-Oriented DBMS
Table oriented	Object oriented
Information splitting (entity principle); sharp seperation between data and functions	Unity of information and functions (class principle)
	Class heritage
Manual generated primary keys	System generated object-identifiers
Descriptive query language; embedded SQL	Object-oriented languages (Smalltalk, C++)
No user-defined data types; Complex data types (maps, pictures, sounds...) only as Binary Large Objects (BLOBs)	User-defined data types and functions (user needs to define the query in Smalltalk or C++ by using class heritage)
Access modes depend on the used operating system	Access modes depend on type declaration as public, published or private
Query result as a temporary table in RAM or on harddisk	Query result as an object in RAM
Short response time	Long response time; resource consuming

THE GRDC DATABASE SYSTEM AND NETWORK IMPLEMENTATION

DATABASE SYSTEM IMPLEMENTATION AND DATA PROTECTION

The GRDC database was developed by the GRDC and TRITON GmbH[1] in close collaboration, where the GRDC designed the database and TRITON undertook the programming and fine tuning part. The access tools to the database have been worked out the same way.
The GRDC uses INFORMIX OnLine 5.03 as system software. This database engine is designed to manage mid-size databases up to a limit of about 5 GBytes. It allows to build up a very comfortable SQL-Server architecture. The engine is implemented on a Pentium PC using SCO-UNIX as operating system. If the databases grows out of limit it is easy to update the system to a higher level.

The administration of the database system is only allowed on the UNIX platform. This is another aspect of data security, because the user needs a special password to enter the UNIX system and he/she needs to know the implementation and the design of the database system. Besides the common administrative tasks the database administrator is also responsible for the data updating.

Even though INFORMIX OnLine allows to run the system in mirroring mode, the GRDC database is at present only configured in log mode. This means that all calls to the database are written to special files and if a database crash happens it is possible to recover the database with the aid of the log files and the backup tapes to the point where the crash occurred. Backup tapes are drawn each week or after a large update. To prevent table fragmentation, twice a year the total database is unloaded to ASCII, then the tables are deleted and afterwards a re-load from the ASCII file is performed. In the mirroring mode the system runs two identical databases situated on different harddisks. Each transaction is made in both databases. So, when an error occurs in one database (or harddisk) the users are able to continue their work with the still operating database, while the administrator is fixing the problem.

The INFORMIX OnLine engine allows the database administrator to grant user-accounts down to field level. The following accounts can be granted to the users:

- ☞ Select fields (for a query or a view)
- ☞ Update fields (change values in an existing record)
- ☞ Insert rows (import new records)
- ☞ Delete rows (delete records)
- ☞ Index rows (manipulate index)
- ☞ Alter table (change the tables design)

To grant or refuse access permission is very useful for data security and the efficiency of work.

The INFORMIX software is installed in the '/usr/informix' home directory. Depending on highest data security level the GRDC database itself resides on a special file system in the UNIX system. This file system has a volume of 730 MBytes and the diskspace allocated by INFORMIX OnLine is of 550 MBytes. This diskspace is handled only by the INFORMIX software and cannot be manipulated by using UNIX commands. One INFORMIX diskspace can contain more than one database and INFORMIX is able to handle more than one diskspace. The GRDC uses only one diskspace where the following databases reside:

[1] TRITON GmbH; Lohbachstrasse 12; 58239 Schwerte; phone: +49-2304-467442; fax: +49-2304-467447

- ☞ the GRDC database
- ☞ project databases (subsets of the GRDC database such as the ACSYS database)
- ☞ the system database (contains system information)
- ☞ the test database (subset of the GRDC for software testing)

table	row size (Bytes)	number of columns	number of rows	total (Bytes)	comment
grdc	198	36	3,635	719,730	station information
mome	153	44	72,030	11,020,590	monthly discharge
tame	347	102	397,900	138,071,300	daily discharge
datv	45	13	5,326	239,670	available stations
datl	42	11	23,414	983,388	missing values

Table 2: *Statistic information about the most important tables (31.05.1997)*

Table 2 shows the statistics of the five most important database tables in the GRDC and how much space INFORMIX needs to store them. For example if the table tame is downloaded to ASCII the resulting file will be about 150 MBytes. The GRDC database is generated of 25 tables (without the index tables).

NETWORK IMPLEMENTATION

As visualised in Figure 2 there are 4 PC's connected to the database server. These clients use MS-WINDOWS 3.11 as operating system. Physically the FIH Local Area Network (LAN) is used, but the GRDC operates its own logical unit. The FIH LAN uses Novel Netware 3.11 as network protocol, while the SQL server needs TCP/IP. This leads to the problem that two network protocols need to be run at the same time when the LAN services are needed in addition to the database access. The GRDC uses Sun's PCNFS 5.01 as TCP/IP net software.

Meanwhile three of the clients are changed to WINDOWS NT 4.0 and therefore all tools having direct access to the database are recompiled for this operating system. The WINDOWS NT conception makes it more easy for the GRDC system administrator to realise the networking because there is a better support of TCP/IP implemented. It is also planned to change the server to WINDOWS NT to prevent complications when running two different net protocols. Using this kind of network architecture helps the GRDC in setting up its own logical network and grant or withdraw access permission to users.

Because the GRDC database server is not visible outside the GRDC and is running on high security level there is no need for installing a special firewall.

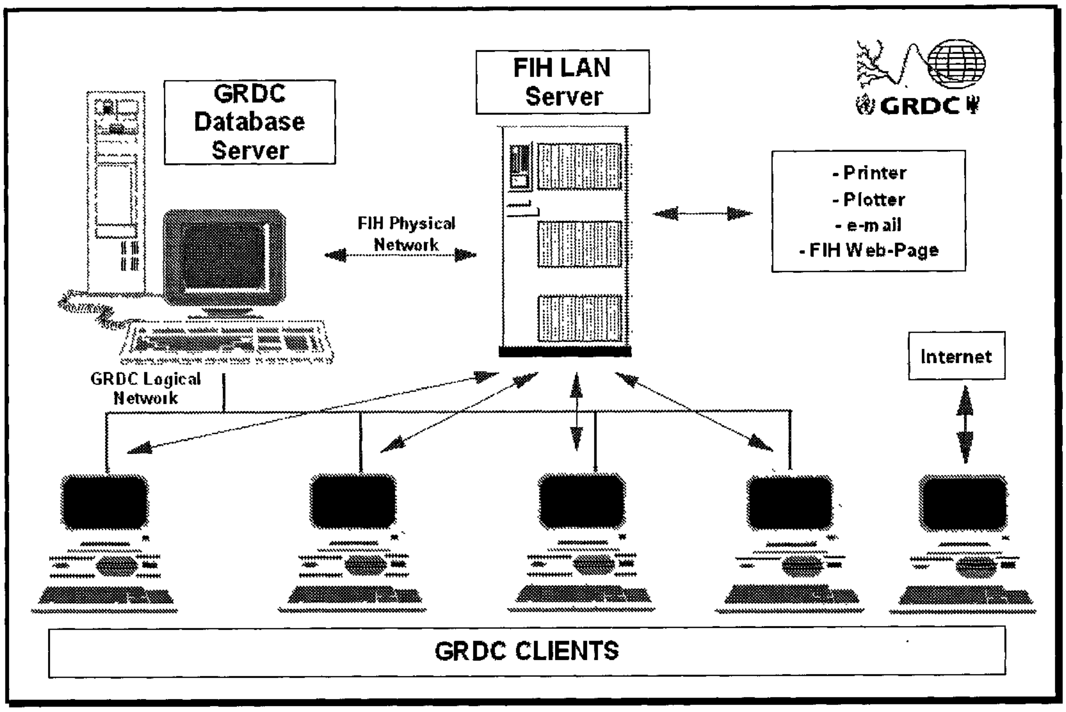


Figure 2: Network configuration of the GRDC

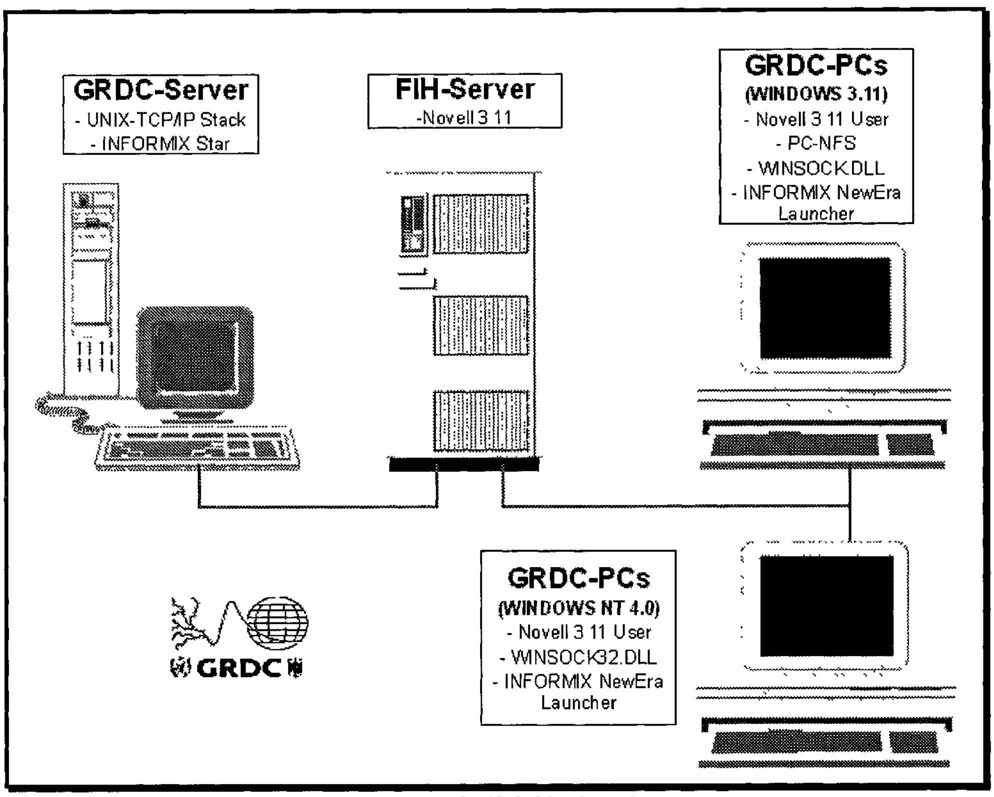


Figure 3: Network protocols and drivers (using WINDOWS 3.11 clients)

DATAFLOW IN THE GRDC

Figure 1 shows the different access modes of the GRDC's tools and programs, and also illustrates a rough overview of the dataflow in the GRDC. Because it can be confusing to visualise the total dataflow of a relational database, in this chapter only the most important parts and how the GRDC solves the problems by using additional tools are described..

DATA IMPORT

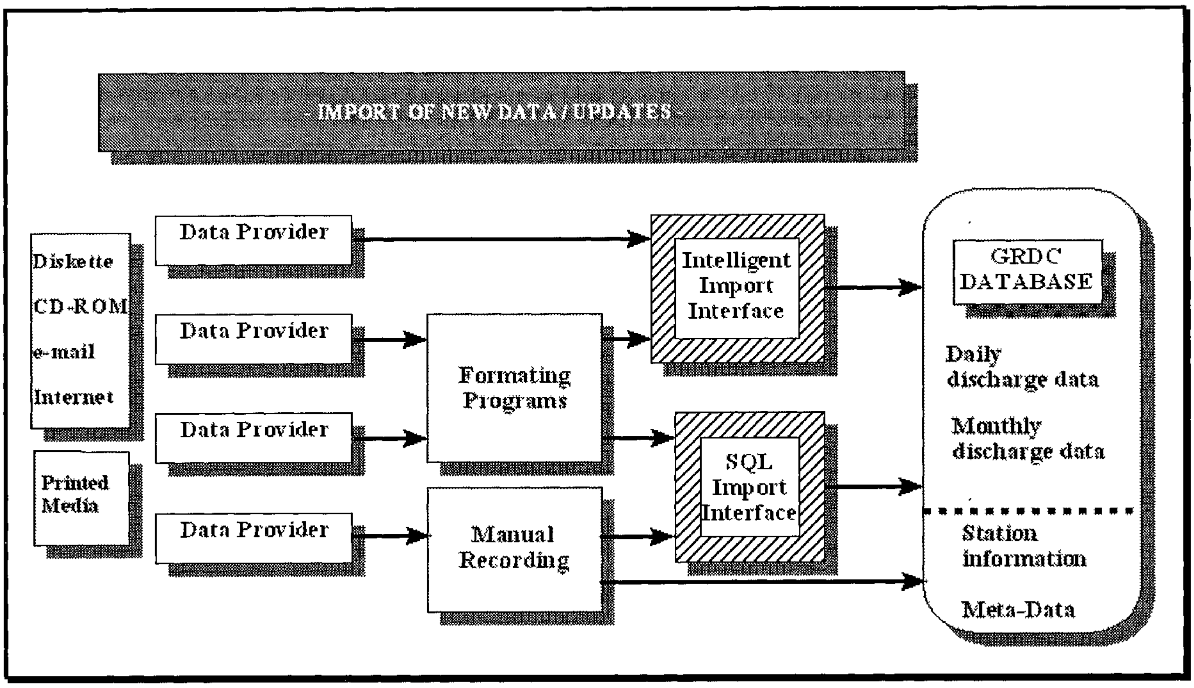


Figure 4: Dataflow Import

To import new data or update existing data sets is one of the most important tasks and with increasing dataflow will become more time consuming. Database updates are now made continuously throughout the year.

The data providers send their data to the GRDC on different medias:

- on diskettes in DOS-ASCII, DBASE, Excel or Lotus format (50%)
- on CD-ROM in DOS-ASCII, DBASE, Excel or Lotus format (10%)
- by e-mail , from FTP-Severs or the World Wide Web (10%)
- on printed media (yearbooks, tables a.o.) (30%)

The data on electronic media are delivered in different formats and cannot be imported to the INFORMIX database directly. To make the import of these files more comfortable the GRDC has developed an *Intelligent Import Interface (GIII)* which stores the structures of the different file formats in a special table in the database. The concept and handling of this tool will be explained later in this chapter. Special formatting programs are also used to re-format incoming data to the standard GRDC-format.

The data delivered on paper media are brought to the standard GRDC import format by digitizing the data to a file which is done twice a year. In few cases the data is typed directly to the database using the functions of the *GRDC Base System* tool.

Figure 4 shows a general overview of the dataflow and the tools for importing new data. The new values will be added to the database table "tame" for daily discharges, or "mome" for mean monthly discharges.
Before the data can be stored in the database tables a unique GRDC-Number must exist for the station. So when importing data for a new station first a record of the main station information has to be generated and, most important, the GRDC-Number has to be allocated. To do this the GRDC uses a GIS application called RAISON. Therefore it is very important that the GRDC gets the geographical description of the stations, such as station name, river name, geographical co-ordinates, size of drainage area, station altitude and name of the first order drainage area. Of course additional (hydrological) information is welcome, too, and can be stored to the database. These data are always typed manually using the function 'Add' in the "Register of Station" dialogue in the *GRDC Base System* and the database table "grdc".
After doing this or if only an update of existing stations has to be done, the datafiles need to be brought to the proper format or are directly imported to the database. If the format is not DOS-ASCII, the GRDC uses either MS-Excel or Quattro Pro to bring it to a text format with fixed field length. In this way it is often possible to generate the GRDC import format without great effort. In other cases either the GIII tool is used when the file format is consistent, not too complicated, and the values are of m^3/s. In all other cases a formatting routine has to be written in BASIC, PASCAL or C. Here normally the standard GRDC import format is used.

GRDC Station Numbers

The most important roll of the GRDC Station Number (GRDC-No) is its function as primary key or part of a primary key in most of the GRDC database tables. The number is of 7 digits and is a combination of 4 different information blocks. The following example of the station "Rhine at Rees (6335020)" will visualise this:

GRDC-No.: 6||3||35||020

6 WMO-Region (1-6)
3 Code of Country (1-9)
35 Code of Subregion (1-99)
020 Code of Station (000-999)

This combination allows the storage of 1000 stations per country and subregion. Another advantage is a rather easy specification of a station on hand of its GRDC-No. While the Region-No, the Country-No and the Subregion-No is fixed, the GRDC is free to generate the Station-No. Here the GRDC orders the stations in direction downstream (000-099) and main tributaries (100-999) to store some more information. This system is gradually applied to those stations with numbers generated in the early days of GRDC development.

GRDC Standard Import Format

There is a distinction between daily and monthly datasets. For both formats it is important that only the data for one station are stored per file.

First there is the description of the monthly data file format because it is easier to understand:

```
PINEGA                                  KULOGORY
1978123.1281.45278.290134.961197.7667.80325.93210.29209.51445.22304.41142.83
197981.06556.38764.64564.8711656.2679.80337.45232.74297.45414.06181.64178.16
1980114.2572.41974.323110.541638.0576.45227.71107.29154.87158.71108.09100.00
198184.16178.38777.51677.7101930.51142.3231.96120.45184.22257.38483.61154.77
198295.03270.22668.903141.901938.6325.32227.54100.87177.19236.38367.87295.67
1983162.4193.29091.0321371.01205.6395.83350.38263.45167.87452.74284.22183.29
1984141.12105.3894.226143.901826.9281.67425.71362.74558.64371.22290.58178.96
1985105.3273.22677.25878.6451364.6564.51373.16166.45226.00530.38200.58133.06
198693.03270.45267.452106.582349.6688.80255.19104.16184.77317.38248.58204.35
198790.51664.03263.58170.5811614.2542.80503.41952.29401.00255.96105.9686.516
```

In the first line the river- and stationnames are placed. The rivername starts at column 1, the stationname at column 41. This leads to a maximum length of 40 characters for both entries. If the rivername is shorter then blanks must be added at its end.
The following lines contain the data. Each line starts with the year (4 digits) followed by the values for each month (6 digits). Because of the global nature of the GRDC it is not possible to use hydrological years. Calendar years need to be generated. Missing values appear in the form of ' 9999.'.
This format is equal to the standard GRDC Export Format that will be explained later in chapter 'Data Export' below.

The standard GRDC import format for daily data looks like this:

```
DNESTR                                  MOGILEV-PODOLSKY

   275   787   275   204   316   234   275   262   422   406 9999.  XXXX 1 11986
   338   200   322   394   400   369   455   335   214   369 9999.  XXXX 2 11986
   448   444   465   444   435   210   378   478   478   444   489  XXXX 3 11986
   366   206   397   492   438   415   325   196   189   296 9999.  XXXX 1 21986
   390   319   350   308   168   175   194   160   147   151 9999.  XXXX 2 21986
   260   208   147   218   305   296   332   338 9999. 9999. 9999.  XXXX 3 21986
   238   181   296   458   403   338   285   234   196   275 9999.  XXXX 1 31986
   316   341   285   267   322   160   196   244   280   253 9999.  XXXX 2 31986
   238   251   181   179   177   183   192   164   162   156   154  XXXX 3 31986
```

Here again the river- and stationname is entered to the first line in the same format as described at the monthly data sets. The second line leaves blank.
The values are entered in the following lines. Here there exist three lines for each month. In the first two lines of each month there are the entries of ten values (6 digits) followed by a missing value (' 9999.'), a filler (' XXXX '), the line counter (1 or 2), the month (2 digits) and the year (4 digits). The remaining values are entered in the third line and the line is filled up to eleven entries with missing values if necessary. Then again follows the filler, the line counter, the month, and the year.
This format is nearly the same as the standard GRDC Export Format and will be explained in chapter 'Data Export' below.

The GRDC Base System: Import Function

For both daily or monthly datasets the import function works the same way. First the data need to be copied to the directory "./grdc/import" in the UNIX system. Now the file to be imported must be formatted to UNIX-ASCII and renamed to "monmess" for monthly or "tagmess" for daily data. To do this a UNIX shell script is used. Then the dialogue for monthly or daily data is selected. The dialogue entry fields will be cleared and the river and the stationname are entered. The INFORMIX system now checks if there exists a valid station number for the entry and if not an error message appears and the import is denied. In the other case the Code of Measurement must be entered, too. This Code of Measurement specifies the type of value (discharge, runoff, water level) that will be imported. After doing so the data import starts. Existing time series will be overwritten by the new data even though there are missing values imported. Therefore it is very important to check this before starting the import. After the import has successfully ended, the table "grdc" gets a checkmark in the field 'New daily/monthly data' in the fitting record. Additionally the date and the name of the importer are stored to each record.

GRDC Intelligent Import Interface (GIII)

This additional tool works on the UNIX platform only and is written in INFORMIX 4GL. To improve the speed of data import, the tool will be re-programmed in C++ once funding is secured. The concept of this program avoids the following disadvantages of the normal data import as described below:

- the checking of the data format and consistency is time-consuming,
- to write and check the conversion routines may take a long time,
- there is no register of formats
- a check if the data is totally stored to the database is not possible at runtime of the routine,
- the check of still existing time series is complicated and time-consuming,
- the user must be trained well in working on a UNIX system.

To prevent these disadvantages the GRDC has developed together with the company TRITON GmbH the *GRDC Intelligent Import Interface*.

This tool copies the import files from a diskette to the proper directory on the harddisk and converts them to UNIX-ASCII automatically. Afterwards it scans the files to find out the format. Now it queries its database table to find a correct or a similar format description. If there exists one the first datasets are shown in a dialogue. The user now commits the format or makes changes interactively. If no fitting format description is found the user has to describe the new format interactively. This description will be stored to the database automatically after finishing. Then again the program scans the first datasets, shows them in the dialogue and the user can commit or make additional changes if necessary.

After committing the routine starts to import the values directly to the proper database tables. If there are still existing data the user is asked whether he wants to overwrite them or not. Missing values are not imported so that a possibly existing value will not be overwritten.

Here again the table "grdc" gets the checkmark for 'New daily/monthly data'.

Even though there is of course a limitation in the number of items in the format description and the files to be imported must be consistent in their format, this tool is very useful in the daily work.

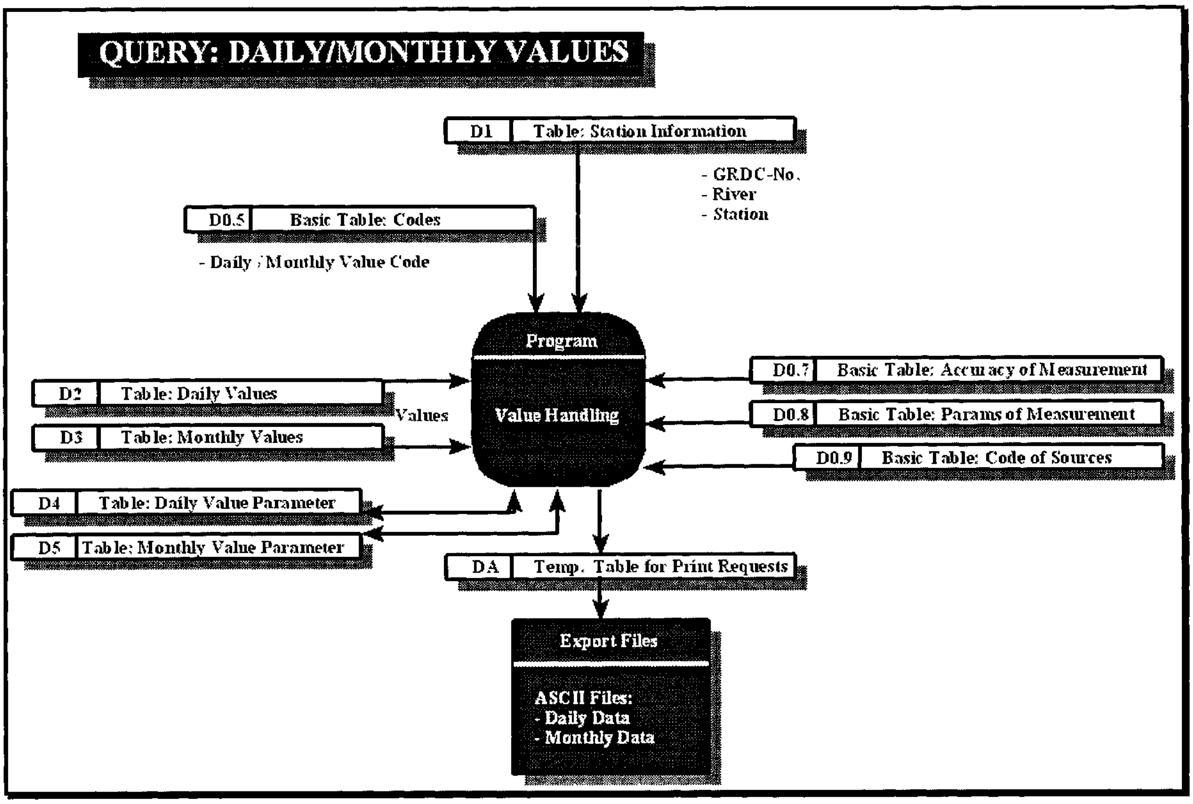


Figure 5: Dataflow Export

The data export is another important aspect in the GRDC. Even though it is possible to export the contents of the tables in all combinations by using SQL scripts, there exist several functions in the *GRDC Base System* for exporting daily and monthly datasets and the station catalogues (catalogue of station and catalogue of missing data). The results of these queries are written to UNIX-ASCII files of special formats. If these standard queries do not fit the requests then of course SQL scripts must be written. Figure 5 shows the dataflow and the affected database tables for exporting daily or monthly discharge data using the standard *GRDC Base System* dialogue.

The GRDC Base System: Export Daily / Monthly Data

There are two ways to export daily / monthly data using the GRDC Base System:

☞ Export of data for each station one by one
☞ Export of data for all stations at once

Which possibility is taken depends on the later demands. For both, daily or monthly data, the export routines work the same way.

When exporting the data station by station the user first needs to know the GRDC-Number for each the stations. After entering this number in the proper field in the dialogue he can start the query by using the function "Search". The data are now visualised in the dialog and real export can be started by using the function "Export". For writing the ASCII file two formats exist, the standard GRDC export format or a screen copy of the dialogue. The GRDC always uses the standard format. When exporting daily data there is an option to compute and export monthly values (arithmetical means). Now there is the option to store the data to a new file or

to append it to an existing file. Monthly data is always exported to an ASCII file named 'monmess', while 'tagmess' is the name for the daily data export file. When exporting the data 'station by station' it is necessary to rename these files. This is done by UNIX shell scripts which also transform the files to DOS-ASCII and copy them to a special directory or directly to diskette. If there are lots of stations to be exported it is better to copy them to a special directory where they can be packed with a standard DOS packer like "lha.exe".
When there is the request for all stations of a subregion the following way is used: First a search for the subregion in the dialogue "Station information"is performed. Using the function 'Mark' in this dialogue will store a flag to all retrieved stations. Now the user has to change to the dialogue "Daily Data" or "Monthly Data" and use the function "Adopt" to mark all records in the table "tame" or "mome" that have the same GRDC-Number as the marked stations in the table ″grdc" automatically. After this function is completed the query can be started for the mark by using the function "Search". Now all the marked records will be shown and can be exported to an ASCII file as explained above. Here the data of all stations are written to one file. Afterwards it is necessary to unmark the records in the affected dialogs.

GRDC Export File Formats

When planning the standard export formats the GRDC tried to find a format that can easily be worked with in either standard software packages or self-developed programs. Even though we know well that these formats will not fit all purposes.

The standard *monthly data export file format* looks like this:

```
PINEGA                                  KULOGORY
1978123.1281.45278.290134.961197.7667.80325.93210.29209.51445.22304.41142.83
197981.06556.38764.64564.8711656.2679.80337.45232.74297.45414.06181.64178.16
1980114.2572.41974.323110.541638.0576.45227.71107.29154.87158.71108.09100.00
198184.16178.38777.51677.7101930.51142.3231.96120.45184.22257.38483.61154.77
198295.03270.22668.903141.901938.6325.32227.54100.87177.19236.38367.87295.67
1983162.4193.29091.0321371.01205.6395.83350.38263.45167.87452.74284.22183.29
1984141.12105.3894.226143.901826.9281.67425.71362.74558.64371.22290.58178.96
1985105.3273.22677.25878.6451364.6564.51373.16166.45226.00530.38200.58133.06
198693.03270.45267.452106.582349.6688.80255.19104.16184.77317.38248.58204.35
198790.51664.03263.58170.5811614.2542.80503.41952.29401.00255.96105.9686.516
```

The format is the same as described in the data import chapter. The river- and stationname is written to the first line. They both are of 40 characters of length. The following lines contain the year (4 digits) and the twelve values (each 6 digits). The missing values are shown as ' 9999.'. If the data of more than one station is stored to one file the stations are seperated by the river- and stationname line.

The *standard daily data file format* looks like the standard import file format, too. The only difference is that there are no additional missing values and fillers written behind the data.

Besides these formats there exist an other format for either daily or monthly data. In this format each line contains the year (4 digits) [and month (2 digits)], the data for one month / year followed by the minimum, maximum and mean. There each value is of 10 digits (######.###). Optionally a delimiter is inserted. This format is fitting well for importing the data to a standard spreadsheet like Excel, Lotus or Quattro Pro. So if the customer preferes these formats this should be noted in the data request. Thus GRDC data can be released to users in a variety of formats.

GENERATING CATALOGUES

Generating the catalogues of available and missing data is done after each major data import. These interim catalogues are for the internal use of the GRDC staff. Public new catalogues are published twice a year. These catalogues are available by e-mail, diskette or can be downloaded from the GRDC website on the WMO or FIH webserver. In addition, a *Catalogue Query Tool for WINDOWS* completes this package. This tool allows the user to query the catalogues in an easy way. It will be explained on page 36 later in this report. Because the catalogues are too voluminous they are not sent as a printout.

Figure 6 shows the dataflow and the affected tables when a catalogue is generated.

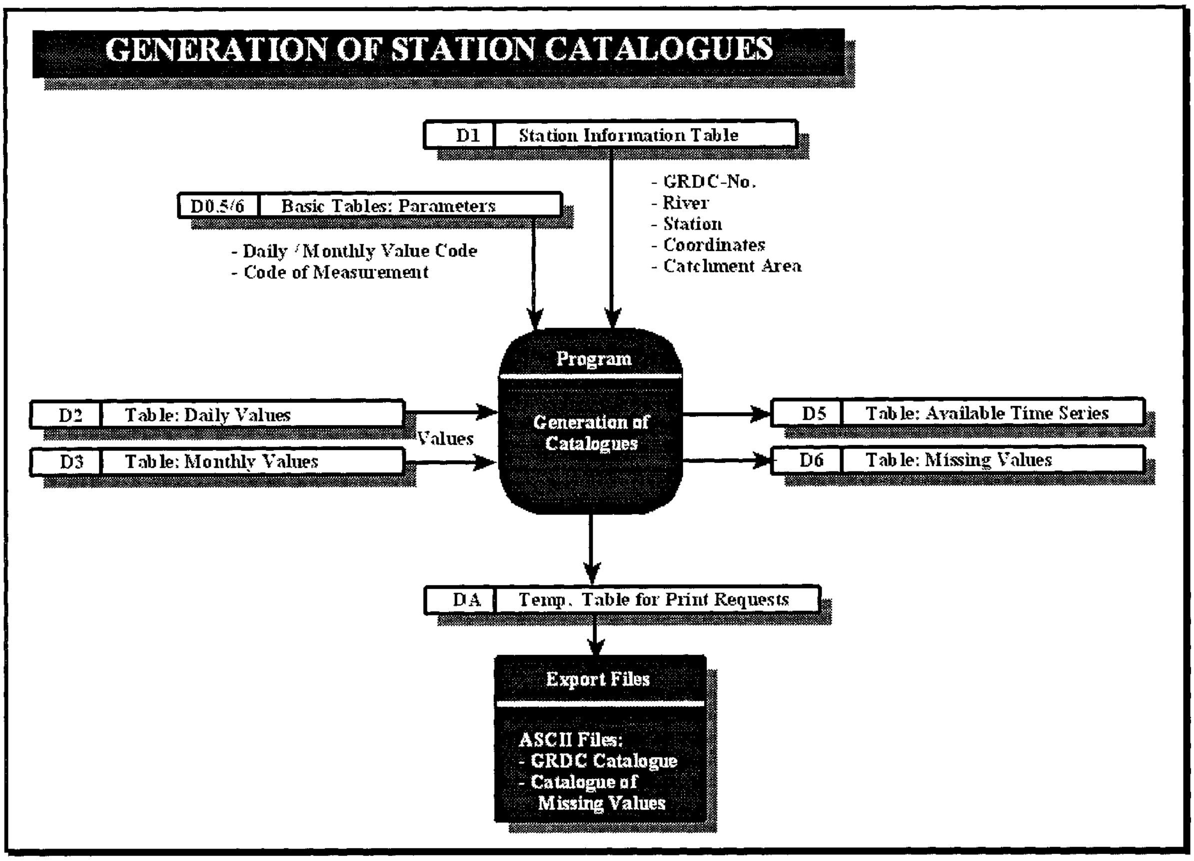


Figure 6: Dataflow Generating Catalogues

The GRDC Base System: Generating Catalogues of Available or Missing Time Series

The generation of the catalogue of available data works the same way as with the catalogue of missing values. The only difference is that the catalogue of available data shows the complete time series (date of the first available record to the date of the last available record) including the missing values and is more thought as a station catalogue, while the catalogue of missing values lists all missing time intervals of each station.

To generate a catalogue the function "Generate Catalogue" in the dialogue 'Catalogue of Stations' or 'Catalogue of Missing Values' is used. Here the following options are available:

- Generate Catalogue for new Daily Data: When new daily values are imported, the corresponding record in the table 'grdc' will be marked. Now only the information for these marked stations will be updated (or added) in the catalogue tables.
- Generate Catalogue for all Daily Data: Here the catalogue tables are cleared and newly filled by scanning all stations in the table 'grdc'.

- Generate Catalogue for new Monthly Data
- Generate Catalogue for all Monthly Data
- Generate Print File: Here an ASCII file is produced containing the information of the tables 'grdc' and 'datv' (total time series) or 'datl' (missing value).

Example of a GRDC Station Catalogue File

```
1201100 Medjerda    Ghardimaou     TS 3627N   843E    1480  1 1976 12 1979 M 1
1201150 Mellegue    K13            TS 3612N   850E    9000  1 1976 12 1979 M 1
1201500 Medjerda    Sloughia       TS 3658N   952E   20895  1 1976 12 1979 M 1
1104150 Cheliff     Sidi Belatar   AL 3602N   027E   43750  1 1976  8 1978 M 1
1104200 Mina        Oued El-Abtal  AL 3550N   068E    6635  1 1976  8 1978 M 1
1104500 Isser       Lakhdaria      AL 3662N   358E    4149  1 1976  8 1979 M 1
1104530 Sebaou      Baghlia        AL 3680N   387E    2501  1 1976 12 1979 M 1
1104600 Bouselam    Sidi Yahia     AL 3642N   460E    4309  1 1976  8 1978 M 1
1104700 Rhumel      Oued Athmania  AL 3623N   630E    1220  1 1976  8 1978 M 1
1104800 Melah       Bouchegouf     AL 3645N   772E     552  1 1976  8 1978 M 1
1204900 Joumine     Djebel Antra   TS 3695N   947E     235  1 1976 12 1979 M 1
1304100 Emsa        Emsa           MC 3552N   530W     110 4  1971 2   1988 D 1
1304800 Kert        Dar Driouch    MC 3490N   329W    1353 6  1969 9   1987 D 1
_______________________________________________________________________________
A       B           C              D  E       F      G      H       I       J K
```

```
COMMENT:

    A = GRDC-Code
    B = Name of river
    C = Name of station
    D = Code of country
    E = Latitude in decimals (please divide by 100)
    F = Longitude in decimals (please divide by 100)
    G = Catchment area in km²
    H = first available record in database
    I = last available record in database
    J = Daily/Monthly data
    K = Missing values in % to total time series

GRDC-Code (for example 1304800):
    1      = WMO-Region (1 = Africa)
     3     = GRDC-Country code (3 = Morocco)
      04   = GRDC-Subregion; main river basin (04 = Mediterranean Sea Coast)
        800= GRDC-Station code
```

The files of the Catalogue of Stations and the Catalogue of Missing Values are nearly of the same format. In H and I the Catalogue of Missing Values shows the time interval of a missing time series in day, month and year. In K the Code of Measurement is shown (1 = m^3/s). In the published catalogues the number of missing values are written as percentage of the total time series instead of the Code of Measurement.

THE GRDC BASESYSTEM V 2.0

The GRDC Basesystem is divided into two separate Tools:

- Database Administration Tool
- Database User Tool

This separation is used to ensure data security.

As mentioned in the chapters above, the GRDC Basesystem allows the user to query the GRDC database easily. Therefore all often used functions are hard-coded. The Database User Tool even allows special queries where the user uses the hard coded relations but can change the conditions of the SQL 'where' clause.

The following SQL script will show the river- and stationname and the monthly values from January of all stations in Asia which are on a altitude higher than 1000 m (key words are bold):

```
Select s_gewa_bez, s_stat_bez, mo_monat_1
From grdc,mome
Where s_pegel_null > 1000 and s_reg_grdc = 2 and s_grdc_nr =
        mo_grdc_nr
```

THE DATABASE ADMINISTRATION TOOL

This tool is UNIX based and is available on the GRDC database server. It is written in ROSI SQL. Even though it contains the functionality of the Database User Tool its main purpose is the administration of the GRDC database. Here the following tasks need to be mentioned:

- user administration
- database updates
- generation of new station catalogues
- export of data to ASCII files

All other administration tasks, as backups, restores, re-indexing of the database, consistency check etc. are done with INFORMIX tools. Also special queries which need to be formulated in SQL scripts must be done with the INFORMIX tools by the system administrator.

The dialogues for manipulating the tables needed for checking the entries and the data consistency can be found here, too. There exists for example a special table for the rivernames to keep them in a consistent appearance. Another example is the table where the code of measurement is defined.

User Administration

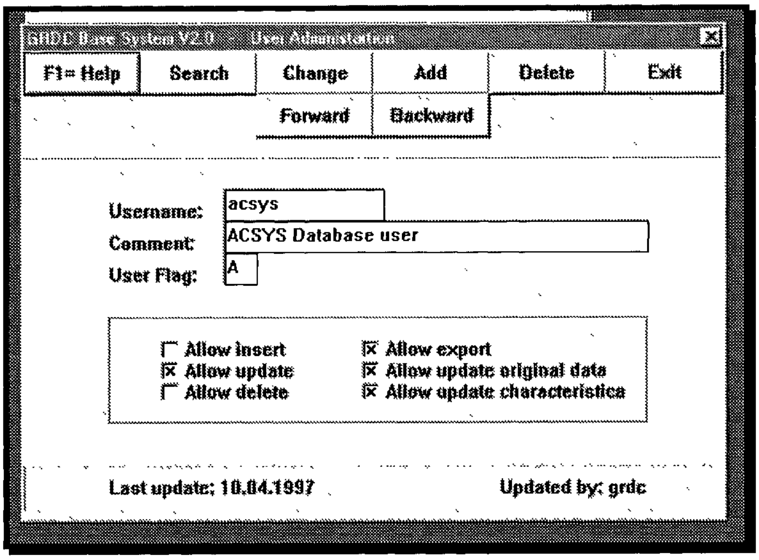


Figure 7: Dialogue User Administration

This task increases with the growth of GRDC's cooperation in international projects. The administrator is able to define a special flag for each project. So the GRDC does not need to generate special databases for each of these projects (and consequently, the need to administer these databases!).
The administrator defines a database user 'acsys' and identifies him/her with the flag 'A'. Now he/she marks all stations in the Register of Stations that belong to the ACSYS project with this 'A'. When the user 'acsys' connects to the database, the system recognizes this and will show only the information and data of the ACSYS project.
The second aspect of the user administration is the management of user accounts like changing, adding or deleting values in the database. So for example the administrator may give the right to change values for the user ACSYS but not the right for adding or deleting, this user can now use the Plausibility Tool and overwrite the original data with corrections but he/she is not able to delete suspect values or add manually new data sets. By doing this the original data are still available in special fields of the table. The GRDC uses this possibility to prevent accidental manipulation of its data by users who are not confident in the use of the GRDC tools e.g. guest scientists.

Generation of new Station Catalogues

The generation of a new set of catalogues (Catalogue of Available Time Series and the Catalogue of Missing Values) is time consuming and only needs to be done after updating a large amount of stations. Therefore it is useful to do this on a special machine with an operating system that allows real multitasking. For example the generation of a catalogue of missing daily values will take more than 9 hours (end 1996). The resulting ASCII files are transferred to the GRDC's staff computers by FTP or telnet.

Export of Data to ASCII Files

This function is used whenever a user needs the data of special stations. Here the user marks the data by using the Database User Tool and if he is not allowed to export the data by himself addresses the administrator to extract them for him. The purposes may be a data request from customers or the generation of special data products with tools which cannot directly access to the database. This task splitting makes sense with reguard to monitor data requests. Here again the generated ASCII files are sent to the requesting user by FTP or telnet.

Handling of Meta Information

Besides the storage of discharge data a variety of meta information can be stored to the databank. Information like the national station numbers or station altitude are stored directly to the table 'grdc' in the dialogue 'Register of Stations', others like the 'Code of Source' or the 'Accurasy of Measurement' are administrated in special dialogues and shown in the main dialogues as flags (numbers). Of course it is possible to query for these informations in the main dialogues. A great problem for the GRDC is to get these meta information from the data providers.

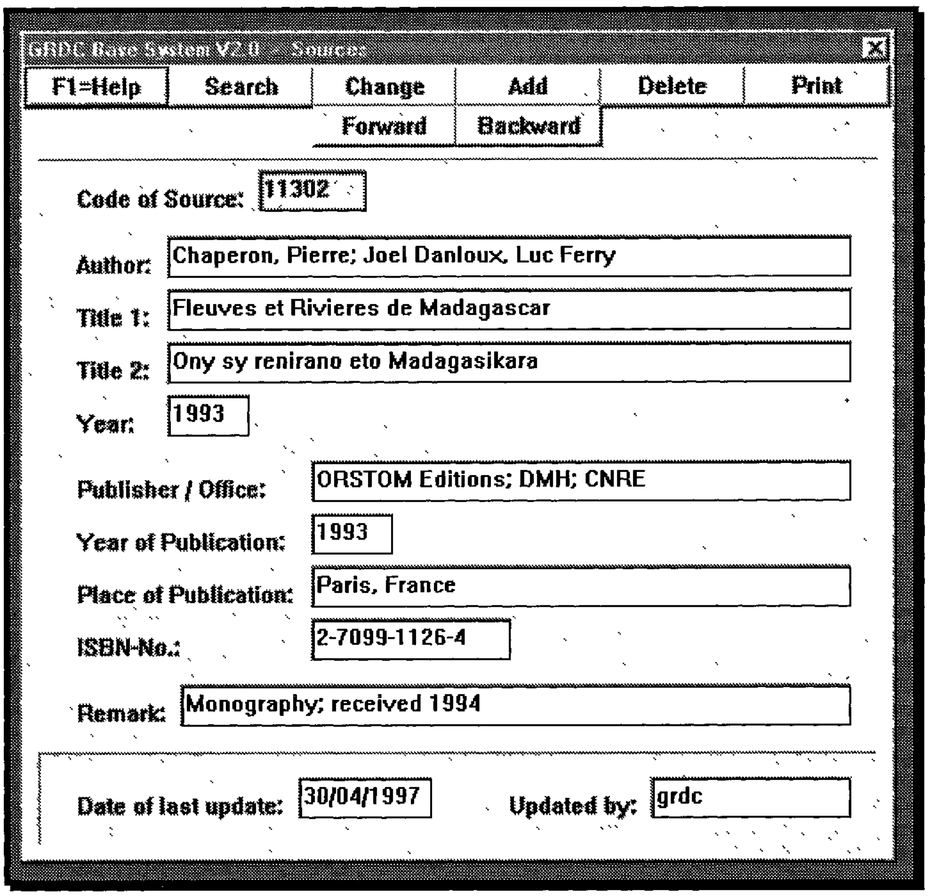


Figure 8: Dialogue 'Data Sources'

THE DATABASE USER TOOL

This tool works on a WINDOWS NT / WINDOWS 3.11 platform and is written in INFORMIX NewEra. It is installed on each GRDC computer.

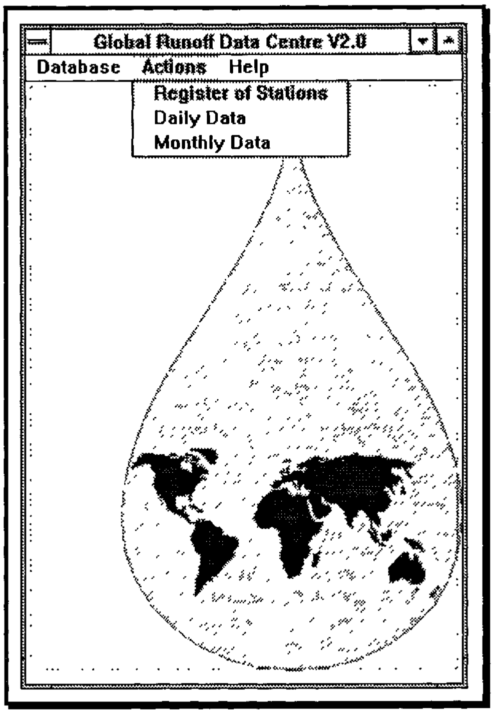


Figure 9: Base System V 2.0 for MS-WINDOWS

After starting the program the user needs to identify himself. The database system connects the user to the database and manages his/her accounts as mentioned above. After the connect was successful the menu item 'Action' in the main menu is enabled. The user may select one of the three options:

- Register of Stations
- Daily Discharge Data
- Monthly Discharge Data

Register of Stations

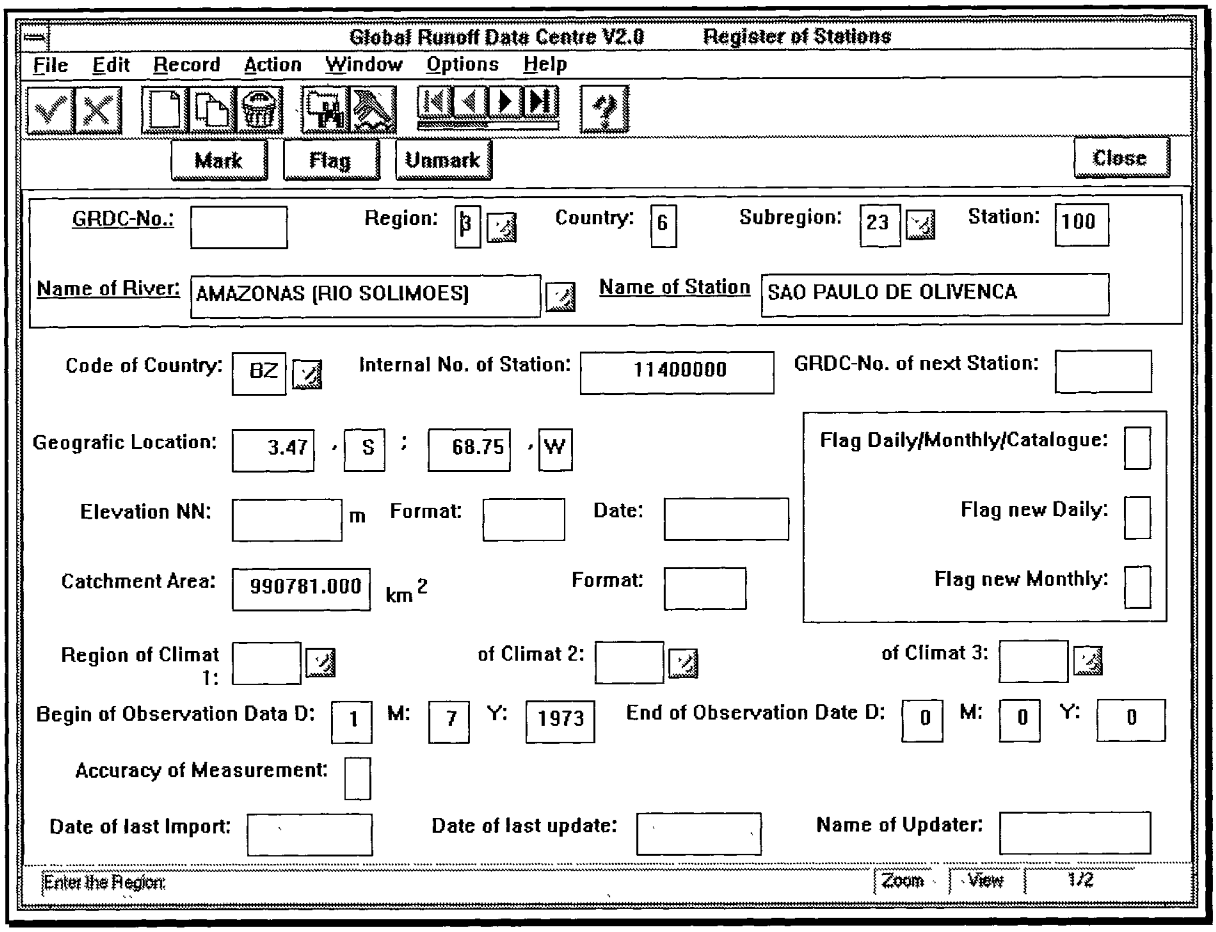


Figure 10: Dialogue "Register of Stations"

Figure 10 shows the dialogue of the function 'Register of Stations'. [It shows all information stored in the database about the stations the GRDC has data from. It also shows the information the GRDC should receive from the data suppliers!]

In this dialogue all fields are enabled for querying. It is possible to query for more than one criteria. Wildcards and specifiers (<>...) are allowed. There exists a comfortable option for more complicated queries in the menu item '*Record*'. Here a SQL script can be generated using self-defined relations and where clauses.

After the query is started the number of found records are shown in the very right field of the status bar. Using the multimedia buttons one can navigate through the view.

With the '*Flag*' Button the present record will be marked, with the '*Mark*' Button all records of the view will be marked. This is very useful if the discharges of a couple of stations shall be shown or extracted because this mark can be adopted by the daily/monthly data tables. With the '*Unmark*' button all marks will be deleted in this table.

Other fields like 'Flag new Daily Data' or 'Flag new Monthly Data' are only of interest for the database administrator. These flags are set when new discharge data is imported and will be deleted when a new catalogue is generated. The field 'Date of last Import' shows when the last discharge data update was done.

The 'Z' buttons open browser windows, where the information is generated in a table-like form.

Daily Data Dialog

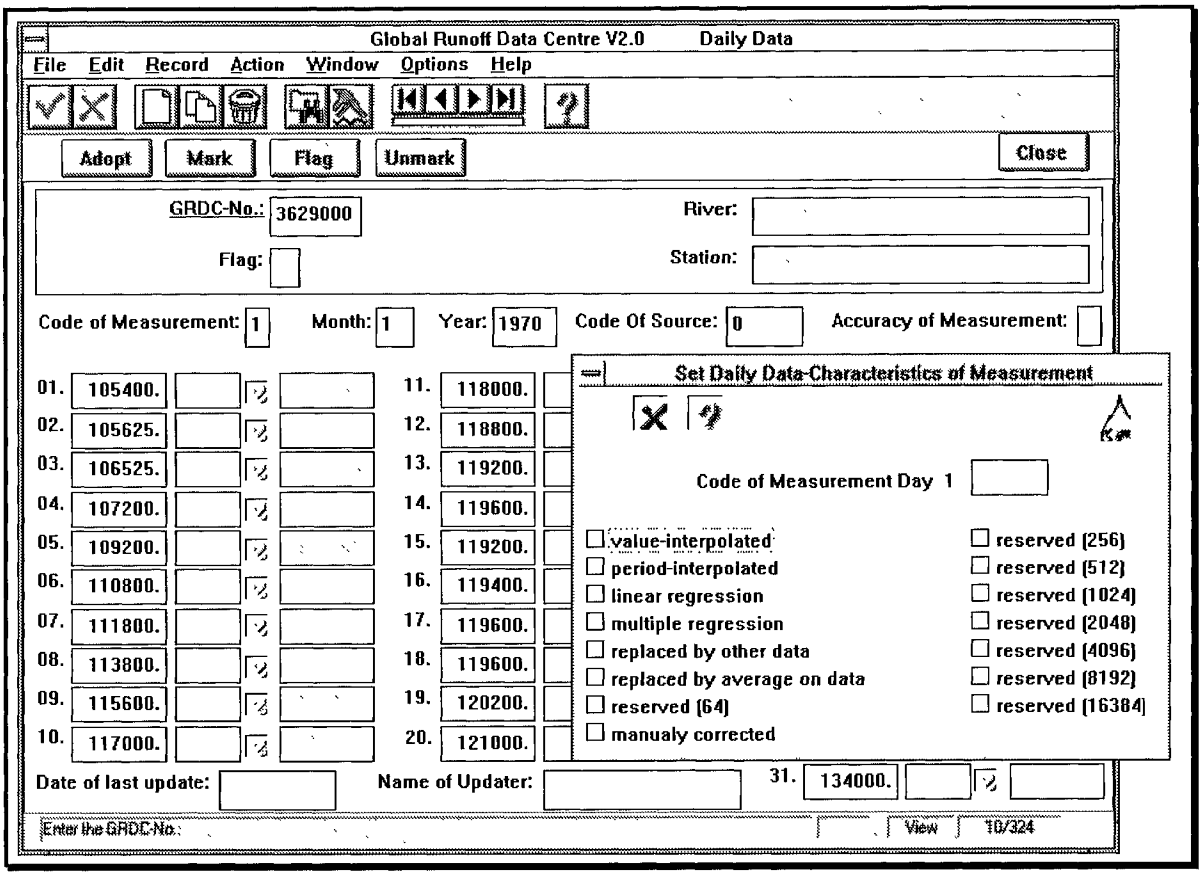


Figure 11: Dialogue "Daily Data"

The Daily Data Dialogue shows the user an overview about the daily discharges stored for a specified station. The normal query is to search for the GRDC-Number. River- and/or station-names cannot be entered directly in the dialogue. Of course, when using the menu-item '*Record*' a user-specific SQL script can be generated as described on page 19 above.
If there are stations marked in the Register of Stations dialogue, one can make a link by clicking the '*Adopt*' button and afterwards make a query about the flag.
After the query has ended the data is shown month by month. For each day there exists one field with the corrected value, one field that shows the way how the value was corrected and one field showing the original value as it was imported. For correction the GRDC Plausibility Tool is taken which automatically sets the proper flag to the table. If a value is corrected by using the basesystem's '*Change*' option the user opens the '*Set Characteristics*' window and click on proper characteristic.
About the '*Code of Measurement*': This field is of interest if there are not only discharges in m^3/s stored, but for example ft^3/s. The GRDC is always storing m^3/s and this is "1". The '*Code of Source*' shows if the data is shipped by diskette, e-mail, yearbook or other media.
Here again the '*Date of last Update*' and the '*Name of Updater*' are integrated.

Dialogue Monthly Data

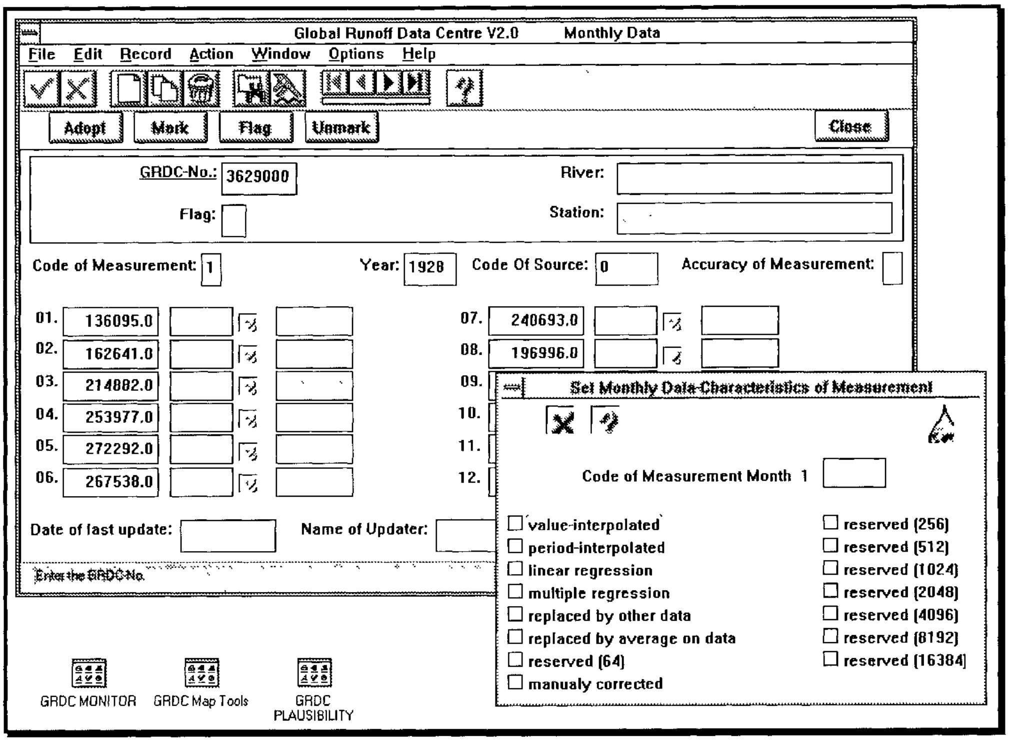


Figure 12: Dialogue Monthly Data

The dialogue of monthly data shows the mean monthly discharges for one year. It has the same appearance and functionality as the dialogue for daily data.

ADDITIONAL GRDC TOOLS

The tools briefly described in this chapter improve the utilities of the GRDC Database System with regard to data queries by users, quality check and the visualization of the database for monitoring purposes. The description of the tools is presented in an 'overlook' way and is not meant to describe the tools and their functionality in detail.

GRDC PLAUSIBILITY TOOL

This chapter describes in an informative way the main functions of this tool as an extension of the GRDC database system.
The Plausibility Tool allows the graphical check of mean daily and mean monthly discharge hydrographs of one or several stations. The tool is meant to be operated by an experienced hydrologist who can make sound plausibility judgements on the basis of graphical inspections and filling of data gaps using the methods described below. Due to the large variability of hydrological regimes an automated process is not desireable.
The tool was developed by the GRDC in close collaboration with TRITON GmbH.

Purpose

The main idea for developing the GRDC Plausibility Tool was:

- ☞ Checking the time series of the GRDC Database for inconsistency and suspicious entries
- ☞ Filling the missing values or correcting suspicious data by using different algorithms or manually

Suspicious data normally can be recognized by a visual inspection of the time series curve as:

- ☞ a divergence between chronological 'neighbourhooded' data of the same station
- ☞ a divergence between the long year means of the same station
- ☞ a divergence between neighbour stations in the same time series.

Implementation

The tool consists of two seperate programs:

1. the database access and the computing routines (written in INFORMIX NewEra)
2. the graphical visualization (written in C++)

This is because INFORMIX NewEra works as an interpreter and is for this reason slower than a compiled program. The tool itself directly accesses the database and is implemented on Windows 3.11 and Windows NT 4.0.

Software handling

As mentioned above this application is directly accessing the database. This means that the data is directly read from the database and the changes are directly written back to the proper table if the user owes the rights to do so. To prevent loosing the original values there exist two columns for each discharge value in the tables 'mome' and 'tame' (see Appendix A). When importing discharge data each value is written in both columns. So when changes are made by using the Plausibility Tool the corrected value is only transferred to one column while the original value still remains in the second column. Additionally a flag is set for the changed value to show the method of correction. If for example a typing error is detected the original value can be corrected by the database administrator, too.

Presently customers receive the original datasets if they do not ask for the corrected data, while the GRDC uses the corrected data when generating data products. One handicap at present is that the customer is not able to get to know about the correcting method, but it is planned to implement new export file formats where the flags for each value are exported, too.

The checking of daily and monthly data work the same way. Before starting a session the username and password need to be entered and will be checked by the database system. Then the user is connected to the database. Now the 'Check Data' window (figure below) appears and the user needs to enter the information for the session. Now the requested data are drawn from the databank and the data of the year to check is shown in the data selection table (figure below). In the 'Suspicious' column a flag is set by the program if a value is detected that is out of the threshold limitation ('Threshold Value' in the figure below). This threshold value can be pre-determined to adapt the tool to different hydrological regimes.

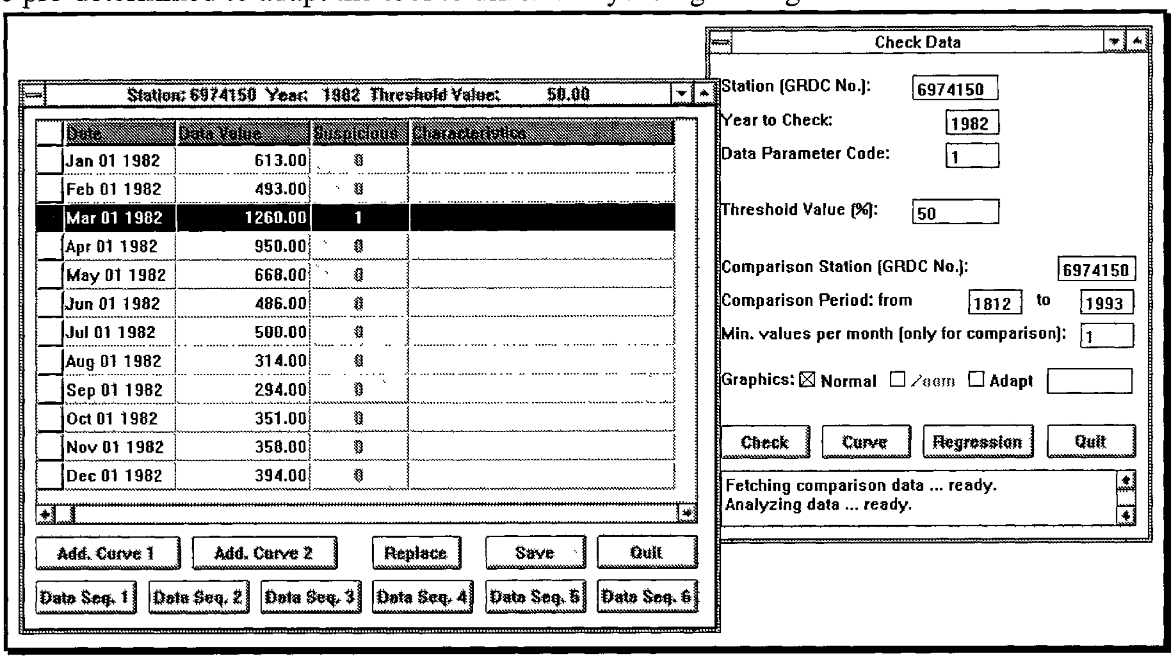


Figure 13: Session selection and data selection table

By using the buttons 'Add. Curve' two additional time series can be added for graphical inspection, while additional time series for computation can be added with the buttons 'Data Seq. 1' to „Data Seq. 6".

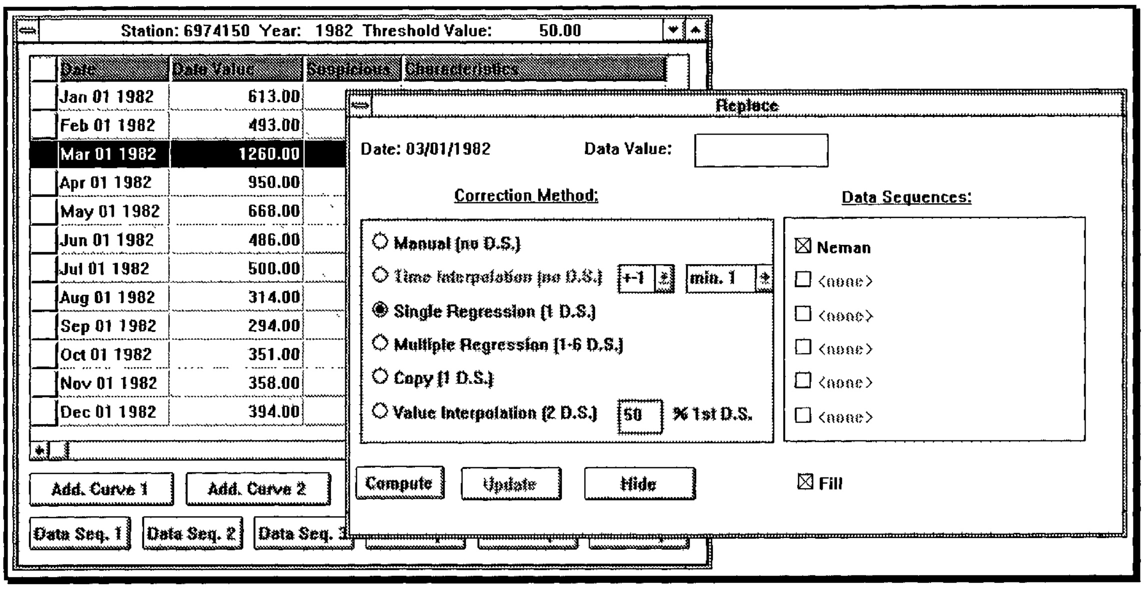


Figure 14: Data selection table and computation window

By clicking the 'Replace' Button, the computation window will appear (figure above). Here the user can select the correction method and the additional data sequences which shall be used as basis for correction. If the box 'Fill' is marked then the computed value is written to the data selection table.
A zoom function allows the close-up screening of daily discharge values and long time series can be scrolled over the screen.
The corrected value is written temporarily to the database by clicking the 'Update' button. To ensure highest data security level, all data exchanges of a session must be committed by leaving the program, then they are conclusively stored in the database. If the user clicks the 'Rollback' item, then the changes are discarded.
The discharge curves can be examined at all time by starting the graphics program of the Plausibility Tool. The figure below shows an example of the graphic window.

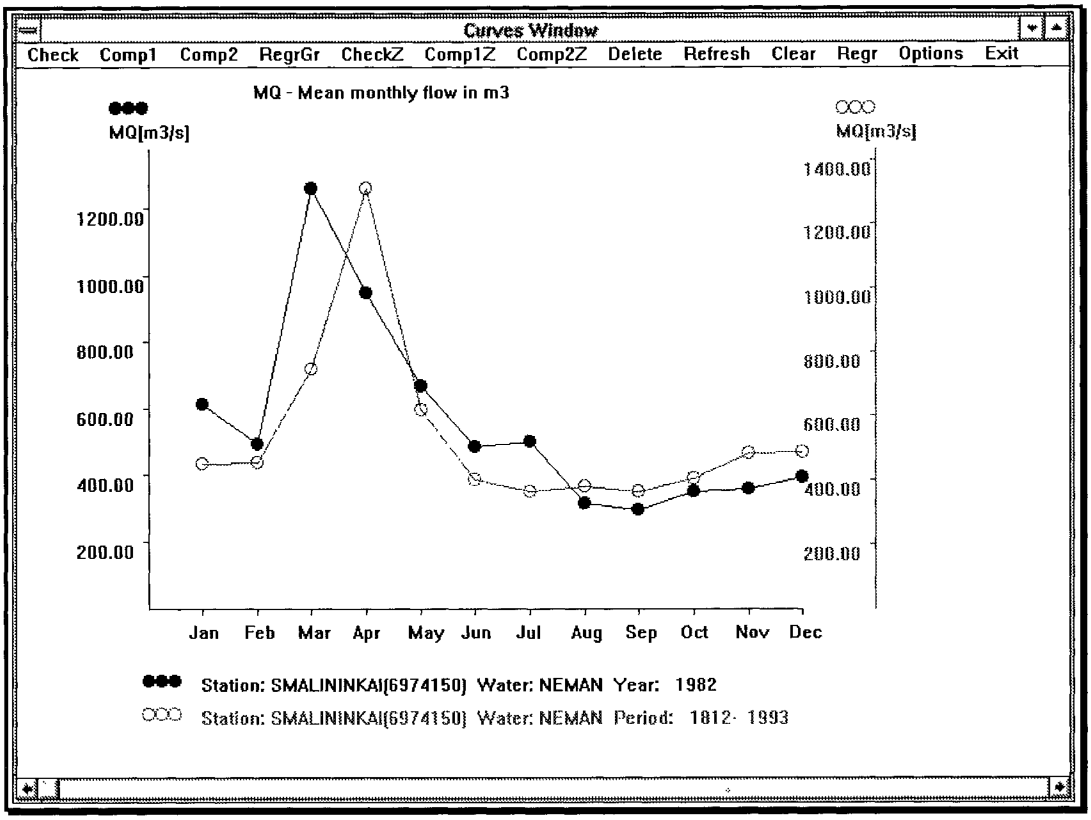


Figure 15: The graphic window

THE GRDC MONITORING TOOL

This chapter will not describe all options of this tool, but only the technical aspects in the view of the GRDC database. An example of a map theme is shown in Appendix B.
The tool was developed by the GRDC in close collaboration with TRITON GmbH.

Concept of the Global Runoff Monitor

The GRDC has developed a software tool to monitor continental and global runoff on a grid-based, comparative basis. The software system is used mainly on request of users and for global runoff analysis initiated by GRDC.

The software system "Global Runoff Monitor" allows the visualization of partial or the entire content of the GRDC database. The main objective for the development of the tool is to assess the temporal and spatial development of the runoff situation of continents or the entire globe on a comparative basis.

The principal utility of the runoff monitor is to identify regions of relative water surplus or deficit in a given year or period of years in comparison with different time series. Depending on the analysis mode, the monitor allows the visualization of long-term development of runoff on a monthly basis, indicate the deviation of runoff values from the long-term monthly average, create slide-shows of the runoff development over regions, continents or the entire globe for any given year where data are available and prepare maps of the classified differences of runoff as a result of the comparison of two different time-series. In addition, the spatial distribution and density of discharge stations can be shown as well as the number of stations per grid. Several additional functionalities allow the user to select or discard stations for analysis or allocate stations to neighboring grid cells if this is necessary.

For this reason, a grid overlay with 2.5° for the entire globe and 1.5° for the continents is created. All discharge stations which are located within the grid-cells are counted and a mean monthly runoff value is calculated as arithmetic mean from the time-series available or used for all of these stations. This runoff value is calculated on the base of the size of the grid cell. In this way, runoff of a grid cell is defined as grid-referenced discharge which occurs within a grid at a given time. Lateral transports are not considered. It is clear from the description that the monitor produces gridded runoff as a first estimate where only discharge data is used as input into the runoff monitor.

The calculated mean monthly (grid-based) runoff values are classified in accordance to pre-set threshold values and then plotted onto global or continental maps. For special purposes, the regions can be zoomed in to obtain a more detailed image of the station and runoff situation. Only those grid cells are used for the computation, where stations and data are available. It is evident from the figure that determined efforts are necessary to obtain the input discharge data for a truly global coverage.

Implementation

The GRDC Monitoring Tool is implemented on WINDOWS 3.11 and WINDOWS NT 4.0. It is directly accessing the GRDC database. The graphical and computation part is written in C++, while the data accessing part is written in INFORMIX NewEra.

Map Generation

The GRDC Monitoring Tool allows to work with the following map themes:

- global in a 2.5° grid
- each of the WMO regions (continents) in 1.5° and 2.5° grid

The maps include the coast lines, the large rivers and the mid-range rivers. Of course other features like boundaries can be added to the maps.

To generate maps an additional program was developed which reads the input vector files and computes the affected grids (coordinates, area) and stores the results back to a special table in the database. Additionally the ASCII-formatted vector files are merged to a resulting mapfile which is binary formatted. This has to be done only once for each map theme.

Session Features

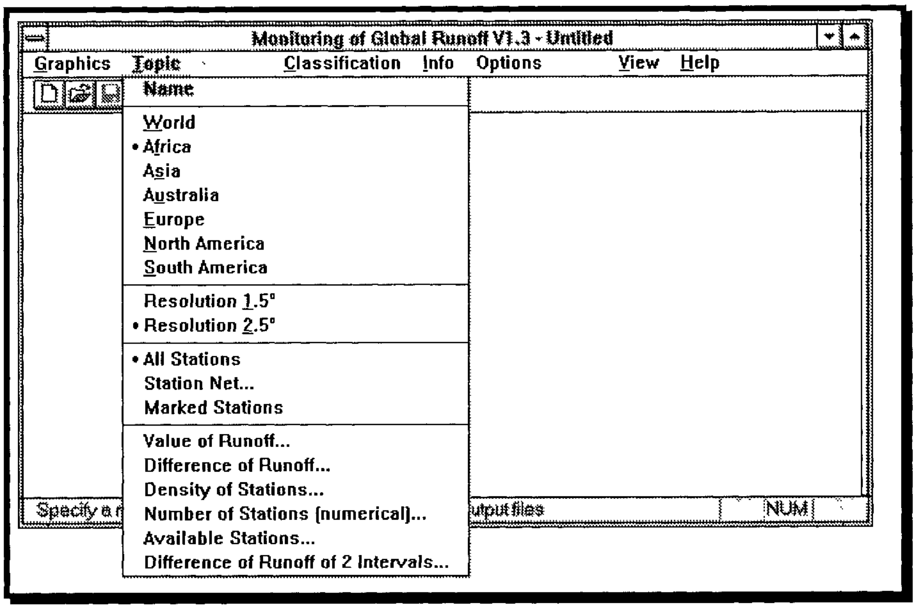


Figure 16: Session features

For each session the features must be selected in the following way:

1. the map theme (world or one of the WMO regions)
2. the grid resolution (1.5° or 2.5° gridding)
3. the kind of stations that shall be affected
4. the kind of information that shall be visualized

The user now starts the query and the necessary information is drawn directly from the GRDC Database. After the query has ended a map or a series of maps will be drawn. The selected information will be visualized in coloured grids depending on different categories. The classification is computed automatically, but the user owes the possibility to change the category limits and colours manually (figure 17).

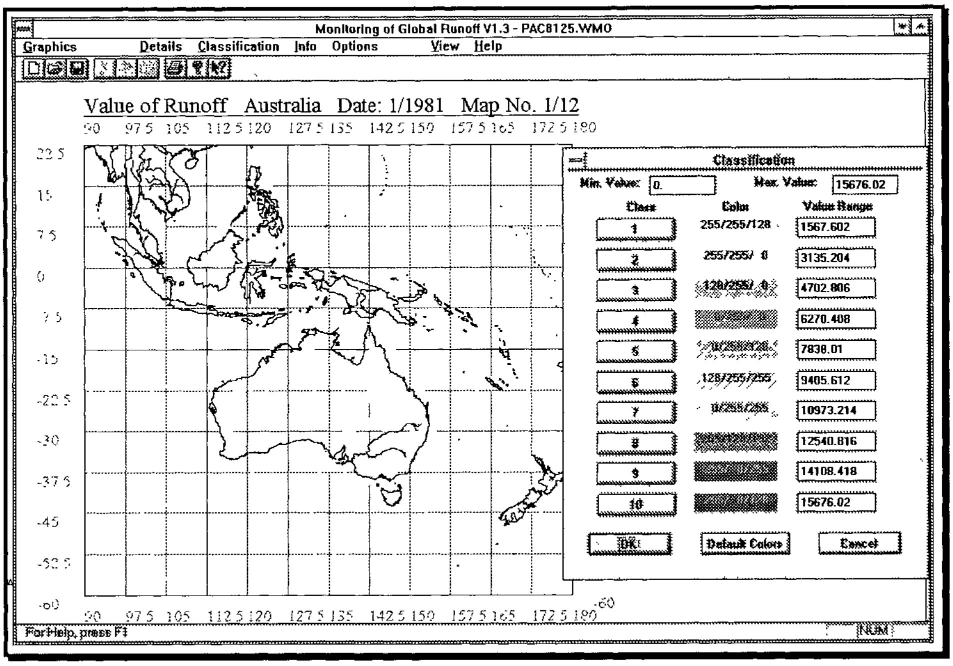


Figure 17: Classification of a map theme

The menu item „Info" allows the user to gain additional information about the stations in a selected grid. Therefore this information is visualized in list form.

The sessions can be stored and reloaded. In this case it is not necessary to start the query on the database system again.

THE GEOGRAPHICAL ANALYSIS TOOL „ RAISON FOR WINDOWS (Vers.1.0)"

RAISON (Regional Analysis by Intelligent Systems on Microcomputers) has been developed at Enviroment Canada's National Water Research Institute.
RAISON offers basic functionalities of a Geographical Information System (GIS) as well as of a database. This together with its extensive use in the water quality monitoring programme of the United Nations Enviroment Programme (UNEP) make it an ideal tool for the GRDC for analysis and visualization of georeferenced hydrological data and information.

The following paragraphs give a selective overview of functions which are currently used in conjunction with the GRDC database system.

RAISON integrates a database, worksheet, map module, contouring module and an expert system with the possibility to make some statistical analysis with the integrated analysis tool. It is particulary suitable for applications which involve point data. Point data are usually samples of an attribute, gathered as discrete observations over a period of time.

GRDC uses RAISON to visualize gauging station locations on maps in context with metadata information from the GRDC catalog and to perform statistical analysis with discharge data from the GRDC main database system. Its extensive display facilities and framework to integrate data, text, maps and images are vital to enhance GRDC's product generation capability. Notably, the display of stations and metadata allows the assessment of gauging station status and monitoring of changes at stations in its geographical context.

RAISON allows to create several projects. A RAISON project consists of data stored in the RAISON database, a set of one or more maps which correspond to the data, and site definitions linking the data to the map. Sites are point data connected to the maps through the use of site locations. In the GRDC, site locations are the geographical longitude and latitude of gauging stations. RAISON allows to store the site locations as X -Y coordinates or Universal Transfer Mercatur (UTM) units.
Within the project, the different modules of RAISON can be used to present the data and analysis results in context to the local and the regional geography. RAISON operates using projects consisting of the project database, a set of maps, site definitions linking the data to the maps and other site or map specific metadata.
All spatial visualization, gauging station status and monitoring, statistical analysis and other advanced data and image processing is performed within a defined RAISON project.

Data Import and Storage

RAISON is linked to the GRDC database system through a transfer programme which has been developed by GRDC. This transfer programme converts the UNIX-based data into ASCII. Required data from the GRDC database are exported and - using the transfer programme are formatted in a way which allows the import module of RAISON to load the data in the internal worksheet of RAISON.
From thereon , all functions provided in RAISON can be used for data analysis and display.
RAISON has the option to import ACCESS, DBASE, PARADOX and BTRIEVE formatted data directly into the database.

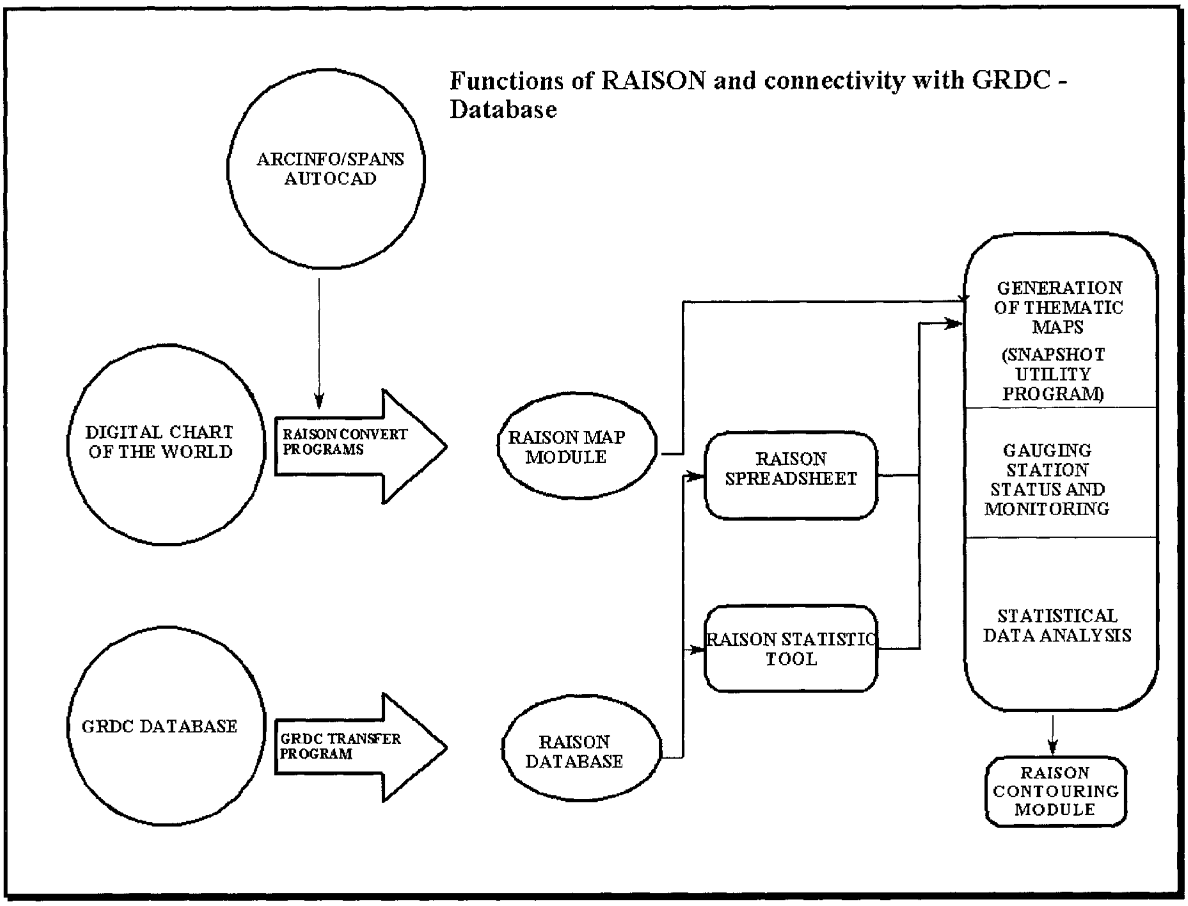


Figure 18: Functions of RAISON and connectivity with GRDC Database

The Map Module

A characteristaical feature of any GIS software is to provide a georeferencing interface to allow the user to view their information in context to the local geography. RAISON integrates a Make Snapshot utility program, which works with map files of different GIS systems such ARC INFO, AUTO CAD, and SPANS.

The GRDC works mainly with map files exported from the DIGITAL CHART OF THE WORLD, containing river lines, coast lines, political boundarys and polygons of catchment areas from the whole world map. Other vector maps are incorporated where necessary (e.g. digital basin boundarys).
With the Make Snapshot utility program users can zoom in a selected area from the world map and create a new snapshot file in the bitmap format (*.bmp). (Figure 19).

Using other applications of WINDOWS like PAINTBRUSH or PRESENTATIONS the created Snapshot may be further processed and restored in the RAISON project.

Figure 19 shows an example of the Amazone watershed snapshot, where a site file with all available GRDC stations is loaded in. Clicking on the station of Obidos, on the right side of figure 19 appears the site information, which was imported from the GRDC databank before.

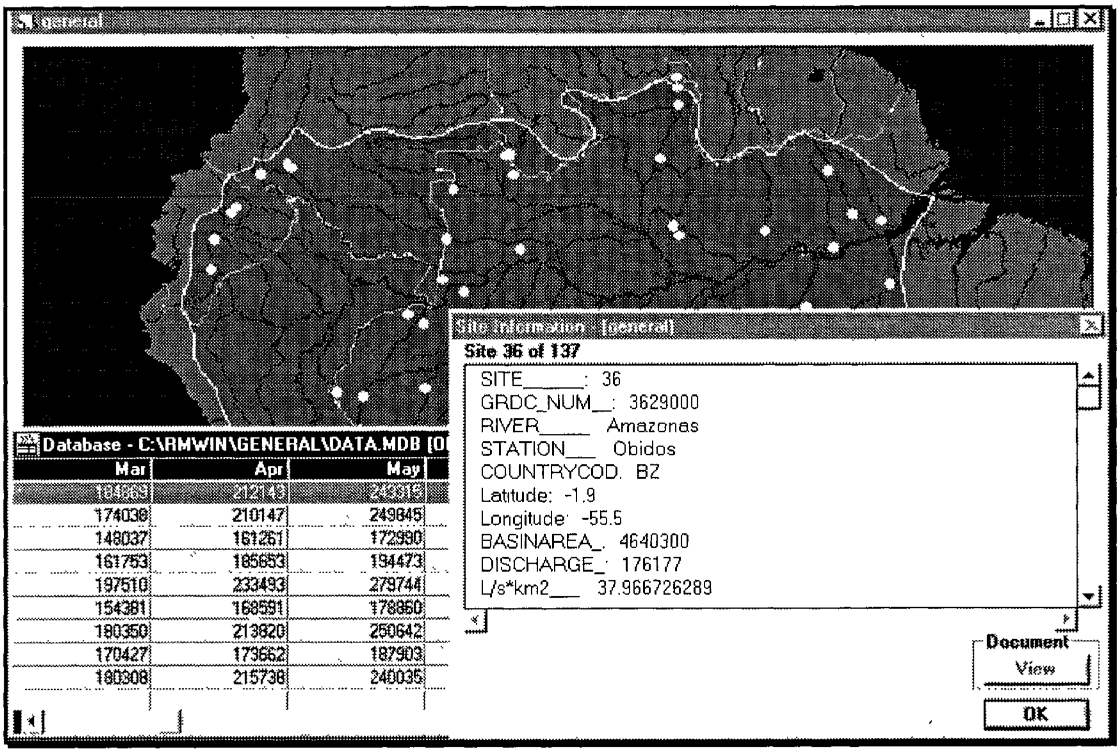


Figure 19: The RAISON Snapshot

The Spreadsheet and Analysis Module

Not only site files with coordinates can be stored as RAISON databank files, but also the discharge data corresponding to the stations. (Window in Figure 19 left side).

Loading the discharge data into the worksheet the user is able to compute averages, maximum or minimum values and some other functions like in a spreadsheet.
From this spreadsheet it is possible to open a graph window, where curves, regression lines, boxplots and moving averages can be created.

After creating a graphic, the user can paste it into a loaded snapshot or it can be stored in a bitmap file. Examples of both facilities are shown in figure 20 and 21.

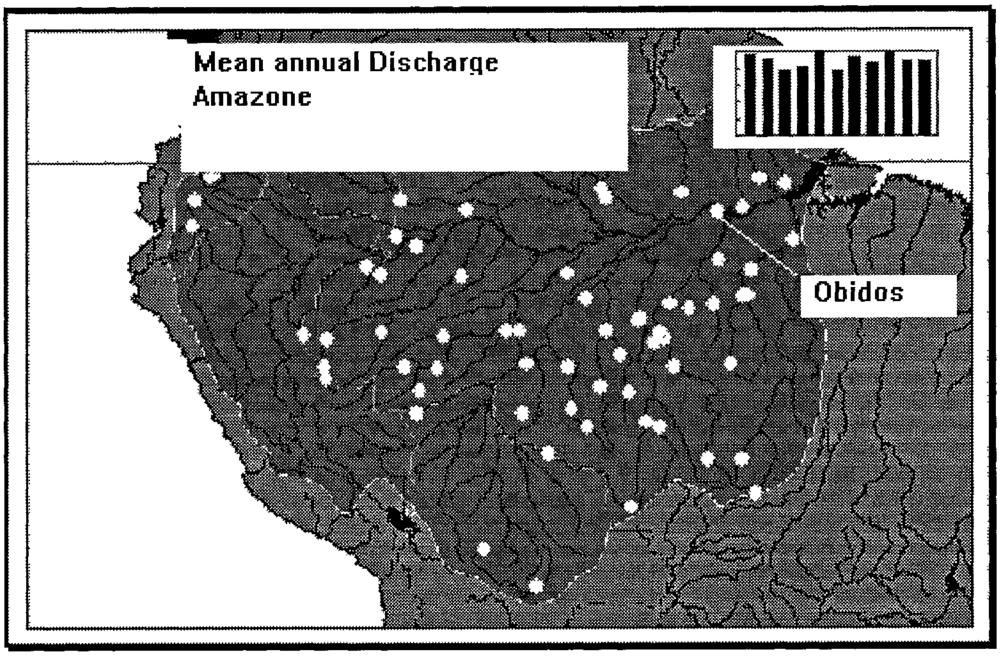


Figure 20: Snapshot with statistical graphs

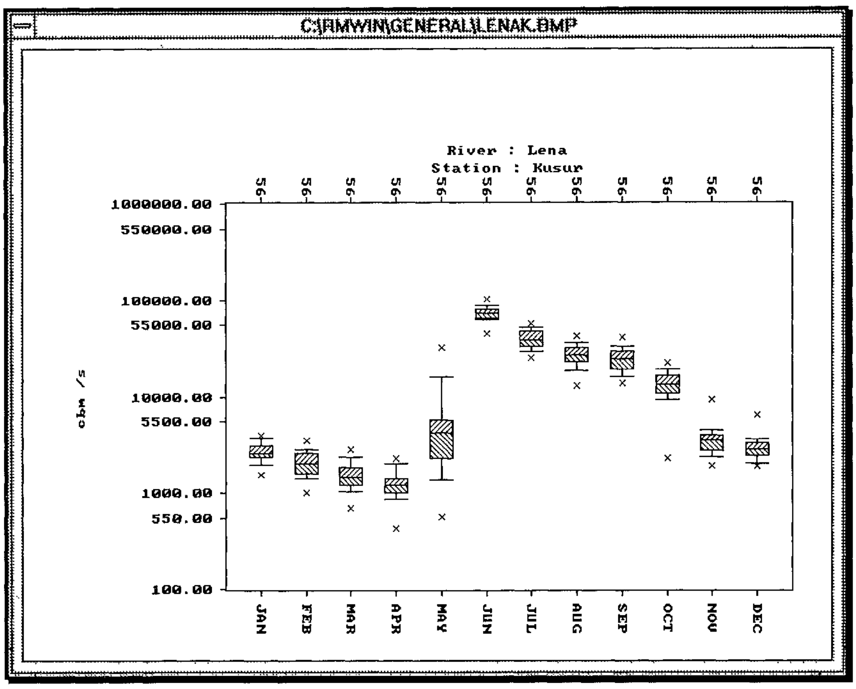


Figure 21: Statistical Graphic as a bitmap

The statistic tool provides the tasks Basic Statistics, Hypothesis tests, Multiple linear regression and Frequency distribution plots:

- Basic Statistics (mean, mode, median, percentiles)
- Hypothesis tests (parametric / non parametric tests, ANOVA and time series)
- Frequency distribution (Normal, lognormal and cumulative distribution)

All statistical functions require at least one set of data, some require more data sets. Therefore the first task is to add one or more data sets into the statistic window (Figure 22).

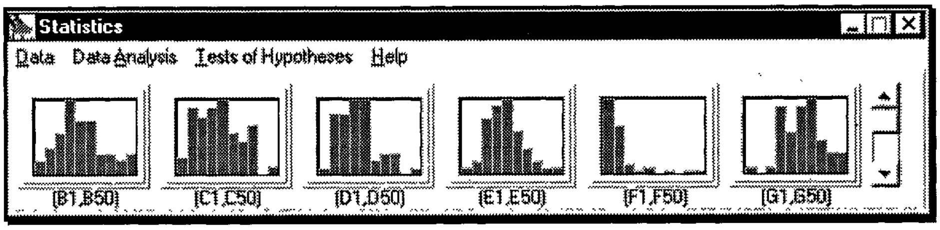


Figure 22: Statistic window containing 6 datasets

The user may add up to 12 sets of data. The data are graphically displayed within the boxes with parameter names underneath. The histogram represents the frequency distribution of the data.
The GRDC uses the statistic tool to characterise the attributes of a hydrograph for example by computing the percentiles. (Figure 23)

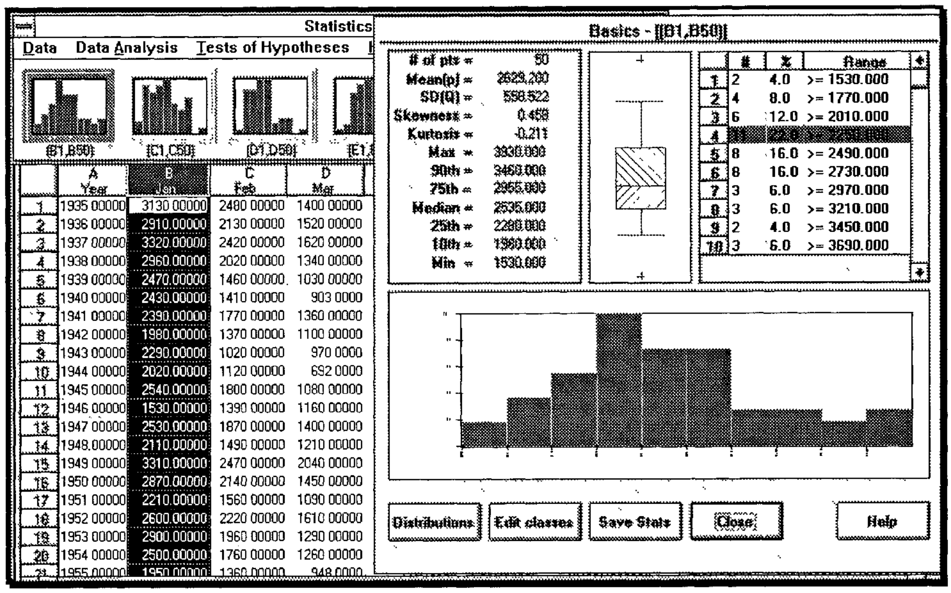


Figure 23: Spreadsheet and statistical results

THE GRDC CATALOGUE TOOL

The GRDC Catalogue Tool is a MS-WINDOWS program that allows easy querying the GRDC catalogue for various options. The Catalogue Tool assists users to make appropriate data selections. It is written by GRDC staff to help them in answering customer requests and to assist users in defining their data needs. The GRDC Catalogue and the Catalogue Tool are distributed on diskette, e-mail or the World Wide Web (from WMO webserver or FIH webserver). Even though it is compiled for WINDOWS 3.11 it also runs on WINDOWS 95 and WINDOWS NT. As programming language BORLAND DELPHI is used. Figure 24 shows the main window of the program.

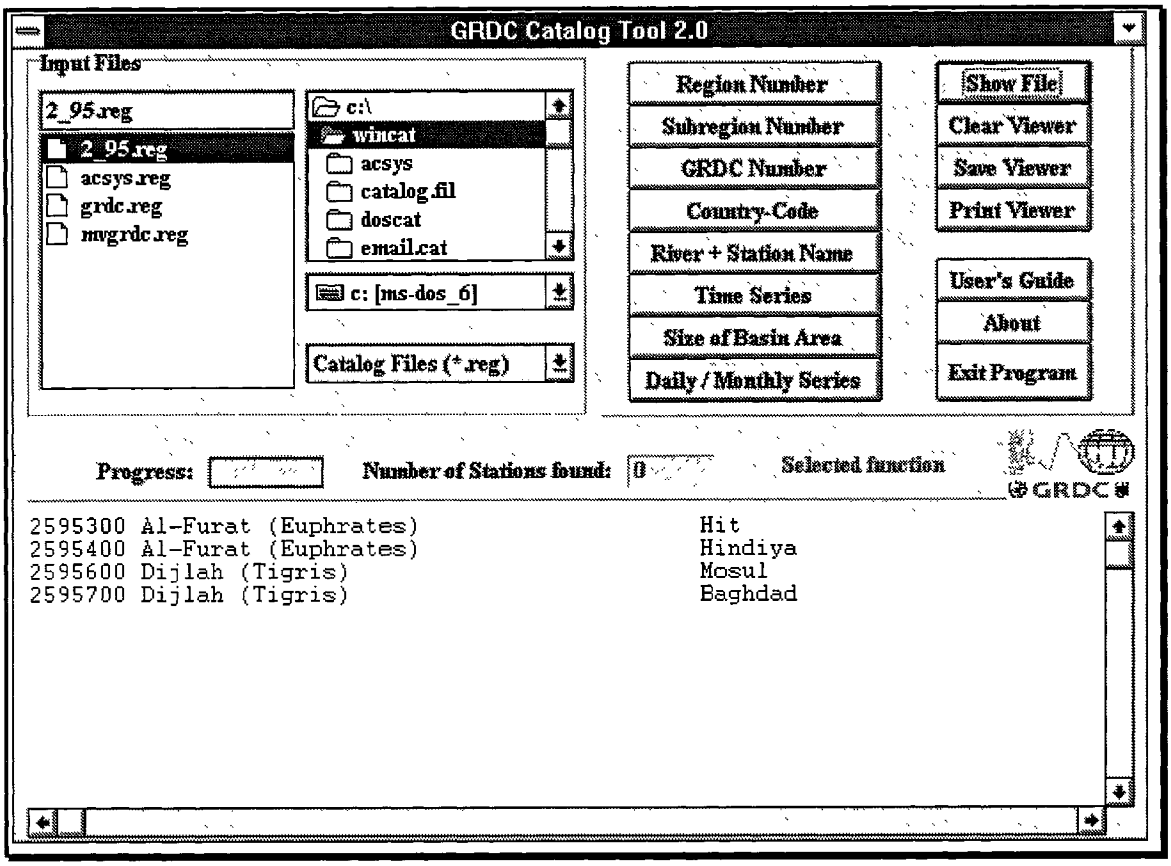


Figure 24: The GRDC Catalogue Tool

The following functions are available for querying the catalogue:

☞ Region Number (WMO Regions)
☞ Subregion Number (WMO Regions and Subregions)
☞ GRDC Number (the 7 digit GRDC Number)
☞ Country Code (the 2 character country code)
☞ River + Station Name (River and/or Station name)
☞ Time Series (stations within a certain time interval)
☞ Size of Basin Area (stations within an area size interval)
☞ Daily / Monthly Series (stations with daily / monthly data)

All functions work as well with the Catalogue of Stations as with the Catalogue of Missing Values, except the function "Time series". This is because of the different date formats in these files.
The results of all queries are added to the viewer and can be stored as a new catalogue file by using the option "Save viewer". After saving, the file-list on the left will be updated immediately and therefore an input file must be clicked before starting a new query. Besides this the viewer's content can be printed out on paper.
The file-list is masked by default for GRDC catalogues (*.stn). Using the file type selection window, any other ASCII file can be shown in the file list window and loaded to the viewer.

APPENDIX A: IMPORTANT DATABASE TABLES IN THE GRDC

Table „*grdc*" (Station Information)

Col.-Name	Type (Width)	Key	Comment
s_grdc_nr	char (7)	A	GRDC station number
s_reg_grdc	char (1)	B	WMO region number
s_land_grdc	char (1)	B	Country number
s_ureg_grdc	char (2)		Subregion number
s_stat_grdc	char (3)		Station number
s_gewa_bez	char (40)		River name
s_stat_bez	char (40)		Station name
s_land_code	char (2)		Country code
s_stat_int	char (16)		National station number
s_klima_reg_1	char (4)		Climate region 1
s_klima_reg_2	char (4)		Climate region 2
s_klima_reg_3	char (4)		Climate region 3
s_geo_breite	decimal (4,2)		Latitude
s_geo_breite_r	char (1)		North/South
s_geo_laenge	decimal (5,2)		Longitude
s_geo_laenge_r	char (1)		East/West
s_pegel_null	decimal (7,3)		Station height NN
s_pegel_null_g	smallint		Flag: Code of measurement (m...)
s_pegel_null_d	date		Date of Station of calibration
s_area	decimal(10,3)		Drainage area
s_area_g	smallint		Flag: Code of measurement (km^2...)
s_beob_datu_ta	smallint		Day of station establishment
s_beob_datu_ma	smallint		Month of station establishment
s_beob_datu_ja	smallint		Year of station establishment
s_beob_datu_te	smallint		Day of station closed
s_beob_datu_me	smallint		Month of station closed
s_beob_datu_je	smallint		Year of station closed
s_n_grdc_nr	char (7)		GRDC-No. of next downstream station
s_me_ge	char (1)		Flag: Marked station
s_datv_m	char (1)		Flag: New monthly data import
s_datv_t	char (1)		Flag: New daily data import
s_flag	char (1)		Flag: Characteristics of measurement
s_updater	char (12)		user who generated or changed this record
s_merkmal	smallint		Flag: Accuracy of measurement
s_last_import	date		Date of last data import (daily or monthly values)
s_last_update	date		Date of generation or changing of this record

Table „*mome*" (Monthly Values)

Col.-Name	Type (Width)	Key	Comment
mo_grdc_nr	char (7)	A, B	GRDC station number
mo_me_code	smallint	A, B	Flag: Code of measurement
mo_jahr	integer	A	year
mo_lite_nr	smallint		Flag: data from yearbook, diskette ....
mo_me_ge	char (1)		Flag: Accuracy of measurement
mo_monat_1	smallfloat		Corrected value, January
:			
mo_monat_12	smallfloat		Corrected value, December
mo_update	char (12)		user who imported or corrected this record
mo_flag	char (1)		Flag: Marked record
mo_last_update	date		Date of last update
mo_mmonat_1	smallint		Flag: Correction methode January
:			
mo_mmonat_12	smallint		Flag: Correction methode December
mo_umonat_1	smallfloat		Original (imported) value January
:			
mo_umonat_12	smallfloat		Original (imported) value December

Table „*tame*" (Daily Values)

Col.-Name	Type (Width)	Key	Comment
ta_grdc_nr	char (7)	A, B	GRDC station number
ta_me_code	smallint	A, B	Flag: Code of measurement
ta_jahr	integer	A	Year
ta_monat	integer	A	Month
ta_lite_nr	smallint		Flag: data from yearbook, diskette...
ta_me_ge	char (1)		Flag: Accuracy of measurement
ta_tag_1	smallfloat		Corrected value 1st day of Month
:			
ta_tag_31	smallfloat		Corrected value 31st day of Month
ta_updater	char (12)		user who imported or corrected this record
ta_flag	char (1)		Flag: Marked record
ta_last_update	date		Date of last update
ta_mtag_1	smallint		Flag: Correction methode 1st day
:			
ta_mtag_31	smallint		Flag: Correction methode 31st day
ta_utag_1	smallfloat		Original (imported) value 1st day
:			
ta_utag_31	smallfloat		Original (imported) value 31st day

Table „*datv*“ (Available Time Series)

Col.-Name	Type (Width)	Key	Comment
da_grdc_nr	char (7)	A	GRDC station number
da_ta_mo	char (1)	A	Flag: Daily / Monthly data
da_me_code	smallint	A	Flag: Code of measuremnt
da_jahr_a	char (4)		Starting year
da_monat_a	char (2)		Starting month
da_tag_a	char (2)		Starting day
da_jahr_e	char (4)		Ending year
da_monat_e	char (2)		Ending month
da_tag_e	char (2)		Ending day
da_updater	char (12)		user who generated this record
da_flag	char (1)		Flag: Marked record
da_last_update	date		Date of record generation

Table „*datl*“ (Missing Time Series)

Col-Name	Type (Width)	Key	Comment
dl_grdc_nr	char (7)	A	GRDC station number
dl_ta_mo	char (1)	A	Flag: Daily / Monthly data
dl_me_code	smallint	A	Flag: Code of measurement
dl_jahr_a	char (4)		Starting year
dl_monat_a	char (2)		Starting month
dl_tag_a	char (2)		Starting day
dl_jahr_e	char (4)		Ending year
dl_monat_e	char (2)		Ending month
dl_tag_e	char (2)		Ending day
dl_updater	char (12)		user who generated this record
dl_last_update	date		Date of record generation

Table „*dbuser*“ (Database User Administration)

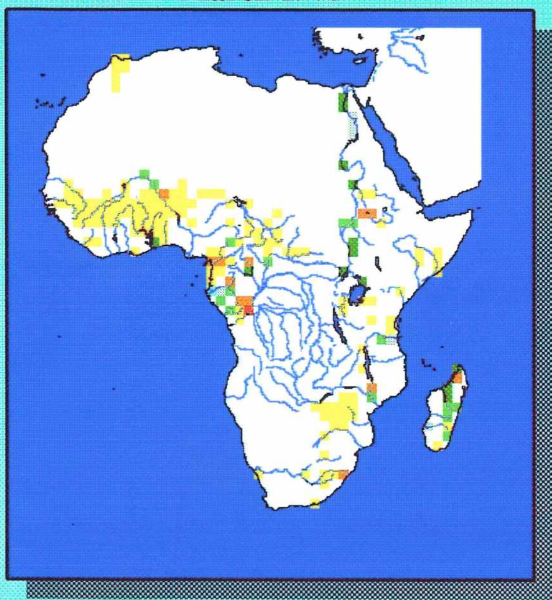
Col-Name	Type (Width)	Key	Comment
us_name	char (12)	A	User name
us_del	char (1)		Flag: allow delete
us_ins	char (1)		Flag: allow insert
us_upd	char (1)		Flag: allow corrected data update
us_exp	char (1)		Flag: allow data export
us_imp	char (1)		Flag: allow data import
us_upd_org	char (1)		Flag: allow original data update
us_upd_mwmerk	char (1)		Flag: allow update methode flag
us_bez	char (40)		comment
us_markz	char (1)		user flag (A=ACSYS...)
us_updater	char (12)		user who generated or changed this record
us_last_update	date		Date of record generation or change

APPENDIX B: EXAMPLE OF THE GRDC MONITORING TOOL

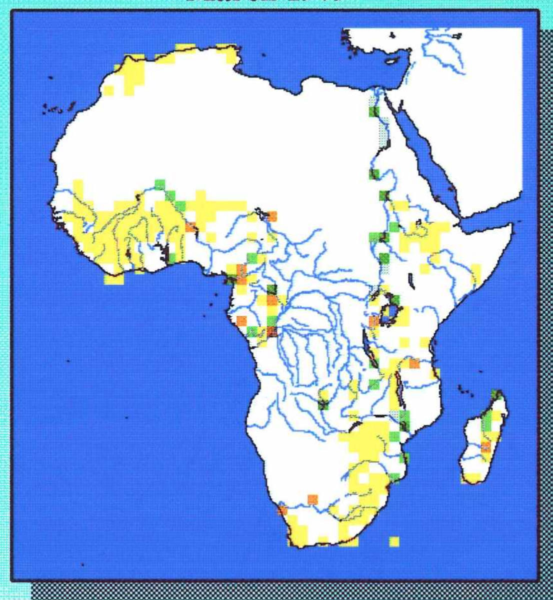
COMPARISON OF RUNOFF IN A 1.5° GRID

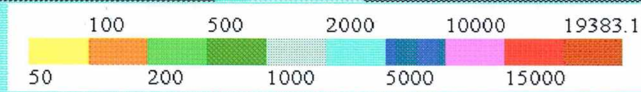
GRDC

Value of Runoff in mm
March 1973



Value of Runoff in mm
March 1979





Difference of Runoff in mm
Mar. '73 to Mar. '79