Jal Vigyan Sameeksha—a publication of
High Level Technical Committee on Hydrology
Vol. IV, No. 2, December 1989

Data Storage & Retrieval System

Deepa Chalisgaonkar
Scientist
National Institute of Hydrology,
Roorkee - 247 667

P.K. Mittal
Foreign SRA (Prog.)
National Institute of Hydrology
Roorkee - 247 667

Abstract : *Most of the hydrological studies being data based necessiates collection and storage of relevant data for further analysis and use. The consistency and timely availability of the required data are influenced by the data handling procedures. Manual data handling methods become less efficient and troublesome when the database grows larger and larger. Computer Oriented data management techniques assist in rapid, easy and economic handling of these data. The paper describes a Data Storage and Retrievel system for hydrological data, developed at NIH, Roorkee, using which data can be stored and retrieved through a common and controlled approach. The system has been developed using dBASE III Plus package.*

1.0 Introduction

The term data refers to symbols used to represent a fact, event, or entity. Information is the meaning given to a set of data. The data can be treated as coded information which has a definite and predecided interpretation. For example, the group of digits '17041962' may not give any explict meaning but in fact it may be the date of birth of a person, a bank account number, discharge of a particular stream, or enrolment number of a student and so on. Therefore the group of digits '17041962' can be considered as a data item having different possible interpretations depending upon applications.

The past few decades have witnessed tremendous growth in utilisation and management of water resources. The successful and efficient execution of these studies / decisions requires a vast amount of data of interrelated phenomenon. The accuracy of results obtained depends upon the quality and quantity of data used. Although techniques are available for synthetic generation of data, the generated data cannot be perfect substitute of the observed data.

Once data are collected, the next important and necessary step is their proper storage. In the conventional ways, the data are mostly kept in manuscript form in registers or files. This form of data storage is easiest, does not require any technical skill and is independent of such factors like capital investment, availability of sophisticated machines and electricity. Usually no preprocessing is done before the data are stored in the registers/files.

In spite of its simplicity, this practice of storing the data has a number of limitations and disadvantages. It is very difficult to keep an up-to-date inventory of data. Many times, data are lost due to physical damage to the storage media. Further, whenever, data are required for any application they are manually copied which has possibily of introduction of human errors besides being time consuming. Any change of format of data requires complete rewriting of the data in the desired format.

2.0 Need of Data Storage and Retrieval System

Hydrology is an applied natural science and therefore hydrologists have to deal with vast emount of data. Moreover, with the increase in population and industrial activities, water is gradualig becoming a scarce commodity rather than a freely available natural resource. Due to this, the analysis of water resources systems is becoming more and more complicated and detailed and this demands, inter alia, more frequent observations of larger number of variables. It has been estimated that the annual volume of primary, information received from a single gauging station is in the neighbourhood of 150,000 characters. Other observations such as water quality might produce between 500 to 600,00 characters. The summarization of these data in year books is a time consuming and voluminous job. Moreover, several types of statistical information is required for the analysis of hydralogical system, i.e., flood frequency determination, design flood estimation, low flows analysis and reservoir operation. The proper storage of these data in conventional way is next to impossible because of shear volumes of labour involved.

3.0 Data Storage

3.1 Conventional ways

In the conventional way, data are mostly kept in registers or files and whenver demanded they are manually copied and supplied. Usually, no data inventory is available and many times, data are lost with the passage of time due to several reasons like physical damage to the storage media. With the introduction of digital computers, data handling is done more rapidly, easily and economically.

3.2 Use of Computer Media

The punched cards may be used for storing the data but this type of data storage has a number of limitations like the cards may be damaged, they are inconvenient is carry, visual checking is difftcult and time consuming and and proper sequence is to be maintained.

Another possible way of data storage is by use of computer files. In the simplest way, data are organised in sequential files which are stored on magnetic media. Whenever needed the data are picked up from this media and are used. To access a specific data record from a sequential file, it is required to check each record in sequence until the desired one is encountered. Clearly, sequential access holds a distinct disadvantage for a user interested only in specific records. Further, the contents of a record on tape cannot be changed, nor can the records be added or delated in between.

To overcome the drawbacks of magnetic tapes, magnetic disks may be used as the storage devices. The magnetic disks are the direct access medium on which data are recorded in concentric circles and a read/write head is used to directly locate, the place where the read/write operation is to be performed. While handling large sequential files on disks, the access to data in several files is awkward and time consuming. Further, there is little asurance that the same data items on different files are defined similarly. This results in lack of data integrity. Program and data are dependent which means that a change in one causes corresponding changes to be made in the other.

In the sequential files, there is no data independence. If a change is made to the data organization, the user has to modify his programme accordingly, recompile it and then test thc modifications. The files thus maintained are useful for one application only. Other applications requiring the same data in a different form cannot use the same data file. There is, therefore, a high level of redundancy with many different files containing the same data element in different form.

Since the same data are stored in different files, any change in data requires that all files should be simultaneously updated. This does not happen always and this creates confusion since different files will have different values of the same data. Further the data on files are mostly updated by use of editing programmes and as the size of the files increases, larger disk space is required to edit them.

In view of these problems of data handling, it is essential to use a systematic approach for storage and retrieval of the data. The software which is employed to manage this task is known as Data Storage and Retrieval System.

3,3 Data Storage and Retrieval System

A Data Storage and Retrieval (DSR) system may be defined as a collection of interrelated data stored together without harmful or unnecessary redundancy to serve multiple applications. Specific data item(s) may be searched/retrieved from this system very quickly and easily,

The concept of a DSR system is that an organization keeps all its possible items of data in a large reservoir from which a number of users with different data requirements can retrieve the data. Each user will have his own view of the data which is derived from a common overall data structure. However, their methods of accessing or searching the data may differ. Such a DSR system would be highly complex because of a wide variety of requirements. In reality, today most of the DSR systems serve a limited set of applications.

The data in a DSR system are stored in such a way that they are independent of the programmes which use them. A common and controlled approach is used in adding new data and modifying / retrieving the existing data within the designed databases. Infact a DSR system is a software interface between a data base and a user. The data is structured so as to provide future application development simple.

4.0 Characteristics of Data Storage and Retrieval System

The immediate objective of DSR System is to make application development easier, cheaper, faster and more flexible. Some of the features of a DSR system are as follows :

1. Performance :—A DSR system designed for interactive use must give quick response.
2. Minimum Cost :—To keep the cost down. the data should be represented in such a way so that the total storage requirement is minimum.
3. Minimal Redundancy :—The DSR system should eliminate redundant data where it is economical to do and should control the inconsistencies that are caused by redundant data values.
4. Search Capability :—The DSR system should achieve fast and flexible search capability.
5. Integrity :—The storage, updating and retrieval procedures should be such as to avoid harm to the data. Range checks and other controls should detect data inaccuracies where possible.
5. Privacy and Security :—Data must be kept secure and private. The data must be protected from a person who may falsely update them. Unauthorized access to the data must be prevented.

In practice, no software package gives all the characteristics that an ideal Data Storage and Retrieval System should have and the designer has to compromise among different qualities.

5.0 Present System

In the present DSR system, the relational data base model has been used. This model

has been preferred over other models because it, generally includes powerful tools for selecting, indexing, sorting and reporting the data. Besides this, the relational structure is the simplest, easily extendable, less redundant and therefore economical. Further, addition of new data to the existing database is also easiest in this structure.

The hydrological data has been divided in to 7 categories :

(i) Station Description
(ii) Meteorological Data
(iii) Land Use & Vegjtal Cover Data
(iv) Surface Water Data
(v) Ground Water Observations
(vi) Geological Parameters
(vii) Water Quality Data

These data have been again subdivided in to different categories and so on. Each type of data have been stored in its own format.

The basic operations which can be performed on these data are :

(i) Edit Data
(ii) Delete Data
(iii) Add More Data
(iv) Get Report (Retrieve Data)

The data have been stored Station wise. A five digit code has been developed for storing the station name. The advantages of using the code are :

(i) A code requires less space for storage.

(ii) Two stations may have the same name. Such type of stations may be distinguished using the codes.

(iii) The station names may be misspelled. The use of code avoids it.

The entire country has been divided in to 10 major river basins and a single digit numeric code has been designed to each river basin. The river basins have been divided in to sub basins and because the number of subbasins in a partlcular river basin may be more then 10, a single digit alphanumeric code has been assigned to them :

The format of the code is as follows :

BASIN CODE	SUBBASIN CODE	STATION TYPE	STATION NO
(1 digit)	(1 digit)	(1 digit)	(2 digits)

The range of different digits of the code is as follows :

(i) BASIN-CODE may vary from 0 to 9.
(ii) SUBBASIN-CODE may vary from 0 to 9. and A to Z
(iii) STATION-TYPE may vary from 0 to 9
(iv) STATION-NO may vary from 0 to 99.

Some validation checks are also available in the system, which checks the format and range of the inputs supplied by the user. If the format is incorrect or the input is out of the range, the system displays the error message.

The developed system is an interactive type of menu driven, user friendly package. Pass word control has been provided in the system to prevent the unauthorized access to the data. The system can display in multiple colours if the colour monitor is used. If a user forgets the code for a partlcular station, with the help of the package he can know it. The user need not to remember the commands because it is a menu driven package,

6 0 Hardware and software requirement

The package has been developed on IBM compatible personal computer, having minimum of 640 of memory, a floppy drive and a winchester disk drive, using dBase-III Plus soft-

ware. The disk operating system DOS must be version 2.0 or later for the program to operate. Any printer compatible with the computer can be used with the program.

dBase III Plus is a file management system with relational features added. It contains it own programming language.

Some of the features of d Base III Plus are :

(i) Program and data are largely independent of each other. The user can change the structure of the database without making many program changes.

(ii) The user can easily, add, edit, delete, or report data using minimum of programming.

(iii) The user can quickly create many reports from the data, using as necessary, mathematical operators such as multiplication and division. Average, totals, and subtotals can be generated easily.

7.0 Conclusions

A computerised Data Storage and Retrieval System has been developed using the physical and logical data descriptions to extract the data items required by the users from the data base.

The characteristics of the developed System are :

* Physical and logical data independence
* Protection from loss or damage
* Easy data updating procedure
* Accuracy and consistency
* Privacy of use
* Timely data availability
* Controlled redundancy
* Fast searching and retrieval.

References

1. Martin, J., Database Organization, Prantice Hall of India Private Limited, New Delhi, 1984.

2. Townsend C., Mastering dBase III Plus, a Structured Approach.

3. Unesco, Hydrologic Information Systems UNESCO - WMO, Paris, Geneva 1972.

4. Wisar, E.H., HILRAS, "Hydrologic Information Storage and Retrieval System Reference Manual.

5. WMO, Case Studies of National Hydrological Data Banks, Operational Hydrology Report No. 17. World Meteorological Organisation, Geneva, 1981.