AN OVERVIEW OF STORAGE AND RETRIEVAL SYSTEMS

by

Kennith E. Foster
Office of Arid Lands Studies
University of Arizona

INTRODUCTION

There is no universally accepted definition for the term data. At times the term refers only to recorded observations of natural phenomena; at other times it refers more broadly to results of calculations and human judgements.

Also, distinctions must be made between "raw" and "processed" data. Raw data are the actual observations made in the field, while processed data are raw data which have been put into a form that will facilitate use and understanding. For example, current meter readings or water quality samples are raw data while tabulations of daily stream discharge are processed data. In this report the term data will refer to processed data.

From a data bank viewpoint, data can exist in two forms: "hard" and "soft". Herein, hard data is the term applied to numerical and narrative data that can be processed by a computer. These are data that are actually filed on magnitic tape, disk, card, drum, or other processing media. Soft data is the term applied to data and information that cannot be processed by electronic data processing equipment.

DEFINITIONS COMMONLY ASSOCIATED WITH STORAGE AND RETRIEVAL SYSTEMS

Information. Information is defined as those data items which have been modified by numerical calculations or human judgement for use in planning, justification, construction, operation, and management of water resource systems. Examples of information are projected streamflow and water uses. Here again, the information can exist in either a hard or soft form.

Basic Data Management. Basic data management refers to the total program for observing, collecting, recording, storing, retrieving, presenting, preparing for use, and publishing basic data to meet, as well as possible, all of the needs of the donor agencies and other participating entities.

Data Bank. Data bank is defined as a collection of data and information stored in specific files that are either numeric or narrative in nature; that may be stored in a centralized location; that are fully described for user access by catalogs and/or narrative index material; that can be retrieved, irrespective of storage form or location by a set of standardized retrieval codes and/or formats; and that have specific security protection features which are responsive to user needs yet protect the supplying entities.

The term data bank implies the need for computer hardware and software required not only to store and retrieve hard data and information but also to provide sufficient supporting computational services to retrieve material in a form most useful to the user entity and within a time period responsive to his needs.

Figures 1 and 2 show the basic structure of the storage and retrieval portion of the data management system. The data bank monitor receives all input requests; determines which processing software will be used; logs in each inquiry into the system; and establishes multiple programming input cues based upon file availability, core requirements, and device requirements.

Next, the system driver program determines which functions are to be performed by the processing software, loads the necessary software into the computer (passing any necessary parameters and receiving control back from the sub-programs), and transfers control to the processing software to perform the required data manipulations are (1) updating a file, (2) extracting selective information, (3) plotting information, and (4) performing statistical analysis on extracted data.

Upon completion of the data manipulations requested, the processing software returns control back to the system driver. The system driver then interrogates the user's request to determine if all activities were performed and appropriately terminates the processing of the user's request.

Because a distinct difference does exist between data handling and information handling systems, a review of the two systems now in operation and user access to these systems is in order.

REVIEW OF INFORMATION SYSTEMS

Smithsonian Science Information Exchange (SSIE). The SSIE annually collects 85,000 to 100,000 single page records of research projects currently in progress. The projects are funded by both public and private sources of support. Each record describes who supports the project, who does it, where and when the research is performed, and includes a technical summary of the project. To obtain more information pertaining to the SSIE, inquires should be addressed 1730 M Street, N. W. Washington D.C. or call 202-381-5511.

Current Research Information System (CRIS), U. S. Department of Agriculture. The CRIS annually collects records of new and continuing research for seven organizations within the USDA. Each record describes who supports the project, who does it, where and when the research is performed, technical summary, and any publications produced during the yearly reporting period. To date about 24,000 project descriptions are on file from the Agricultural Research Service, Forest Service, Economic Research Service, Cooperative State Research Service, Farmer Cooperative Service, the Statistical Reporting Service, and 53 State Agricultural Experiment Stations.

CRIS searches are provided to all agencies of USDA, to State Agricultural Experiment Stations, State Extension Services, and universities.

Generalized Information Processing System (GIPSY). GIPSY is a computerized storage and retrieval system for water resources information, developed by the Water Resources Scientific Information Center, Office of Water Resources Research. Water resource abstracts of current and earlier pertinent monographs, journal articles, reports and other publication formats are on file. The contents of these documents cover the water-related aspects of the life, physical and social sciences as well as related engineering and legal aspects of the characteristics, conservation, control, use, or management of water.

The citations have been housed on computer file at the University of Oklahoma with connecting terminals located at those state water resources research institutes located at the University of Wisconsin, Cornell University, and North Carolina State University.

Interested users in Arizona should contact the Water Resources Information Program, University of Wisconsin, 1324 West Dayton Street, Madison, Wisconsin for retrievals. To date over 60,000 citations are on file.

Oak Ridge National Laboratory. This agency publishes Abstracts of the IBP's Ecosystem Analysis Studies (current issue is V. 3, No. 4, Dec. 1973), which serves as a communication link between the Biome Programs that constitute part of the Environmental Component of the U. S. International Biological Program, and between the various Biomes and the Scientific community. Groups currently participating in this effort include the Coniferous Forest, Desert, Eastern Deciduous Forest, and Grasslands Biomes, and the Island Ecosystems and Upwelling Programs.

It includes abstracts of reports or data sets for each of these biomes, including work in progress as well as publications. It is submitted in machine readable form under a program that forms part of the Oak Ridge Computerized Hierarchical Information System (ORCHIS) operated by the Laboratory

Mathematics Department. From these computer files, material for publication in the Abstracts is selected and manipulated by computer programs to produce the body and indexes in each issue. All material submitted by the various biomes and Programs is stored on data cells to be accessed by a conversational retrieval program.

Not all issues carry abstracts from all biomes. Each issue has indexes for investigator, keyword, taxonomic name, and subject category for each abstract appearing in that number. Some of the subject categories carried in the December 1973 issue were: aquatic, cecompostion, hydrology, invertebrates, modeling, nutrient cycling, productivity, and terrestrial.

It too, like the OALS, has its own Environmental Terminology Index, developed with NSF funding, a much broader tool than the Arid Lands Thesaurus because it covers all environments encompasses by the entire IBP program.

Arid Lands Information System (ALIS). After a development period of NSF funding, the Arid Lands Information System (ALIS) within the OALS is now fully operational, with a modest bank covering worldwide arid lands scientific information relating to the physical and biological aspects of world deserts. One subsystem, relating to specialized information on watershed management, is functioning within the larger framework and will be discussed in the next presentation.

Publications generated by the stored information include Arid Lands Abstracts featuring not only current publications arranged by subject categories such as Environmental Engineering, Geography and Geomorphology, Land Use, Weather and Climate, and Bibliographies, but also carrying in each issue a special retrospective bibliography on a particular arid land topic. Specials appearing to date include those on Burning as a Tool for Arid Range Management, Creosote bush, and Physiological Response to Heat Stress. In addition to this irregularly issued serial is another entitled Arid Lands Information Papers which are specialized bibliographies accompanied by substantive textual interpretation. To date papers have been prepared for Salinity Problems in Arid Lands Irrigation, Geothermal Exploration, Desertification, and Southwestern Groundwater Law. One on Jojoba is in preparation.

In addition to such publications, the ALIS system can retrieve from the machine-readable data base by any mix or match of indexing terminology for customized searches. This operation combines all aspects of a complete index-abstract-information service, from research and selection, to production and promotion of the end product.

The computerized storage-and retrieval system includes verified standard bibliographical information such as author, date, title, and source,

plus abstracts, plus keywork indexing from the Thesaurus of Arid Lands Terminology, developed by OALS under NSF funding. This thesaurus consists of over 3,000 terms, with hierarchical arrangement and scope notes.

The ALIS system is now formalizing alliances with such arid foreign countries as Israel and Australia, to create an international arid lands information network. The Israeli partnership has just been funded by the U. S. - Israel Binational Science Foundation.

Arizona Water Information System (AWIS). The AWIS has been funded by the WRRC, University of Arizona since 1971 on a modest basis as an attempt to draw into a centralized system information regarding water related research and programs within the state of Arizona. The first step necessary in the design of an Arizona Water Information System (AWIS) in 1971 was personal contact with those in the state who would be data or information suppliers and/or users of such a system. Emphasis was placed on benefits in time saved by researchers, planners, consultants and others seeking data to apply to an area of interest or convenience of an automated system to handle water data or information dissemination to the public by State agencies, and the convenience associated with the computer's capability to condense data via programmed computation and presentation of information in a desired format.

Several State of Arizona agencies were contacted to determine which would be interested in such a system. Currently, six Arizona State agencies (Arizona Highway Department, Arizona Game and Fish Department, Arizona Water Commission, Arizona State Health Department, Arizona State Land Department, and the Office of Economic Planning and Development) have specific statutory responsibilities or administrative needs for inventorying, conserving, developing, managing, operating, protecting, or monitoring the water resources and related facilities of the State.

A survey of these agencies and the University of Arizona resulted in an expressed interest for a "water activity file" and various data needs for research. A water activity file is descriptive information pertinent to each water resource activity.

The varied interests of Arizona water information users called for a flexible operating system. A pilot program was developed with the premise that a knowledgeable base of information concerning ongoing and past water resource activities must be initially developed as a referral type system. Once this water activity file was initiated, then data generation cumulating from the water activities was sought. The location of data does not always result in its inclusion into AWIS, however, as many sources are capable of disseminating data themselves. In this case AWIS provides users the location of data through its activity file.

Figure 3 is a Water Resources Questionnaire developed to collect the activity information for entry into the AWIS. The information collected to enter into the questionnaire was secured in several ways. These include: (1) personal contacts; (2) coordination with the Water Resources Research Center; (3) review of files of the Vice President for Research, University of Arizona; (4) utilization of the Current Research Information System (CRIS); (5) information provided by the Office of Water Data Coordination (OWDC), Washington, D. C.; and (6) annual reports.

The water activity file consists of title of project (this may be a data collection program or a research project), person in charge of project, address of person in charge, period of operation, key words to describe activity, brief summary of the work, professional papers or articles resulting from activity, data availability (where data, if any, may be obtained), and province, county, river basin, township, range and section where activity was conducted. Initially many of the projects in the activity file lacked information pertaining to data availability; however, investigators were contacted to determine if data exists and how a potential user can best access it. Water resource information and/or data, if it exists, may be retrieved by key words, or a broad range of geographic locations from Arizona's three provinces down to a quarter, quarter section or latitude-longitude designation.

Sixty-two different agencies in Arizona have been identified as working in some aspect of water. The list is large and quite varied, and includes the consulting firm, state and federal agencies, and the University. A computerized system has been devised whereby information about a research project can be entered and later retrieve any part or all of it. The information we are putting in concerning each project - title, principal investigator and name and address of the institution where the research is conducted. If it is in a university we also include the professional school or the college, a period of operation, the level of funding, who supported it, key words, a summary, and any publications that may have come out of the work, plus a data availability factor are also included. We also entered a geographic locator for the project like province, county, river basin, and range, township, section. All of this information goes into computer entry for every project that on file. What we are trying to accomplish is to keep up in as much totality as possible, what is going on in the way of water resource work in Arizona. These projects may be present and ongoing, or past projects that have since terminated. For example, to date 163 projects have been identified as being conducted in Arizona by the U. S. Geological Survey. This output has been generated from information provided in publication type material. Retrievals can be made on the system by supplying key words, geographical locators, or agency performing the work.

Three graduate students working in fiscal year 1973 at the University of

Arizona, Arizona State University and Northern Arizona University compiled University water resource projects. To date, we have over 1,000 projects on computer file that have been conducted in Arizona dating back to the late 1800's up through today.

DATA STORAGE AND RETRIEVAL SYSTEMS

U. S. Geological Survey Data-Processing System. The U. S. Geological Survey data-processing system was developed between 1963 and 1965 to handle a rapidly increasing volume of hydrologic data more efficiently. This computer-oriented system (Johnson, 1965) uses punch cards for recording data from laboratory tests of geological materials, the quality of ground water, the physical and chemical quality of surface water, spring inventory data, water use, and water level data.

Current access to the system can be made through the USGS, Water Resources Division in Tucson which now has access capability to the data through a medium-speed printing device.

In this system, the forms, punch cards, and procedures for working with well and spring inventories, ground-water quality, water levels and water-use data are of particular interest. The basic elements of the U.S.G.S. ground-water storage and retrievel system are:

"1) A well schedule and other for ms designed to record data in the same order and manner that they will appear on the punch cards, 2) a series of cards for storing the data. Some data have been coded for conciseness and to present the data in a more useful form. Use of the codes will allow direct printouts of certain data from the cards in tabular form suitable for the use in various types of ground-water reports."

The U.S.G.S. system for recoding well-inventory data consists of three cards - designated A, B, and C cards. The A card is a master card describing location, ownership, and data available. Card B is a well-description card, which includes information such as the physical characteristics of the well, water levels, yield, and select quality of water data. The information on card C describes the physiographic and topographic setting, the geology, and the aquifer characteristics.

Water-quality data are punched onto three cards - the Q, R, and S cards. These cards contain the sample location, depth, results of the analysis for most dissolved chemical constituents, and physical properties. Spectrographic analysis of constituents in water is punched on cards T, U, and V. Card Y of the USGS system is for recording ground-water use in an area. Card Z is an aperture card for storing diagrammatic, graphic, and tabulated material; microfilm can be inserted in the aperture, which is a small cutout square.

Water-level data are recorded on dards designated 1, 2, 3-20, 101-999. The 1 card contains data on location, owner, use of well and water, and geologic units. The information on card 2 includes depth, description of measuring point, and records available. Cards 3-20 and cards 101-999 contain a narrative description of the well location, and the latitude-longitude of the well, and the water-level data.

All USGS records are referenced to latitude-longitude, which if properly used will result in a unique number for each location. All records having the same latitude-longitude should be from the same inventory site.

Canadian HYDRODAT System. Recognizing the significance of electronic data processing, the Canadian National Advisory Committee in 1964 requested that the Geological Survey of Canada investigate the feasibility of establishing a geological storage and retrieval system. The objectives of this study were as follows:

1) To study the current and potential uses, requirements, and problems in the recording, storage, and retrieval of geological data using electronic data-processing methods.
2) To recommend the principles to be followed in developing a National system for the recording, storage and retrieval of geologic data.
3) To develop a method of indexing which will serve as a key to the existence and location of geological data and thereby be fundamental to the retrieval, exchange and utilization data.
4) To undertake pilot studies that test the principles on which the National system is based, and that also develop standards for recording data in specific fields.
5) To propose an organization to continue the development of a National system.

Mineral-deposits data, fossil-fuel data, and geophysical and geochemical data were the specific areas of geological field data that were designated for the initial development of computer-processable files.

As a result of the Canadian studies, a computer-oriented system for the storage and retrieval of ground-water data was developed. The Canadian system, termed the HYDRODAT system is similar in format to the USGS system and incorporates many of the same codes and coding techniques.

The basic element of the HYDRODAT system is a well-schedule form for recording and coding data. The well-schedule form is divided into four major sections:

1) Location, ownership, and well-description data
2) Hydrogeologic data
3) Chemical quality-of-water data
4) Lithologic data

Data from the forms are punched onto standard computer cards for

processing.

With the exception of litologic data, the HYDRODAT system is a much shortened version of the USGS system. For example, in the HYDRODAT system, one punch card is used for chemical quality, whereas in the USGS system, three cards are used for water quality and three more cards for spectrographic analysis of trace constituents.

Oil Industry Systems. The oil industry has expended considerable funds and effort in the development of computer oriented systems for storage, retrieval and manipulation of well data. Data from oil and gas wells, of course, differ from water-well data, but there are probably more similarities than differences between the two types of wells. The large investment and the extensive experience of the oil industry with well data cannot be ignored in developing a natural resource information system. Of the data-processing systems developed by oil companies, the Permian Basin Well Data System (PBWDS) is probably the best known because of its scope and wide use. This system represents the efforts of several oil companies to put geological and drilling data from wells in the Permian Basin of Texas and New Mexico onto magnetic tapes. After selection of data formats and development of a manual of instructions in 1963, the job of storing data was begun. To date over from 124, 792 historical wells and 10,858 current wells were put on magnetic tapes.

Alpha and numeric codes were developed for such things as operators, formations, well classifications, producing fields and reservoirs, lithologies, porosity, oil or gas shows, and regulatory bodies. Well data are recorded on coding sheets according to these codes and then punched onto standard 80-space computer cards. About 180 items can be coded for a well by use of 21 different types of cards. The coded data include such information as the unique well number, formation tops, drill-stem tests, location, logs available, drilling problems, type of well and well owner. Once on cards, the data are stored on magnetic tapes. Retrieval of the coded data is facilitated by use of subprograms, which can exclude redundant data, test the records to see whether they contain useful data, and translate the coded data into easy-to-read output complete with headings. Other programs are used to change, add to, or delete from the stored data, to transmit data to x-y plotters according to the latitude-longitude coordinates of the wells, to draw cross sections and contour maps, and for many other user-oriented applications.

STORET System. The Water Quality Data Storage and Retrieval System (STORET) was developed by the U. S. Public Health Service in 1961 and is now being widely used by the EPA. The STORET system enables storage and retrieval of water-quality data in a variety of forms, and for any statistical load and correlation computations that may be desired for data analysis and interpretation.

The basic elements of the STORET system are:

STORET Subsystem II consists of 1) a series of related computer programs designed for the efficient storage and retrieval of data collected in connection with water quality management programs. The system utilizes latitude and longitude to identify the location of data collection points and is intended for use in handling data collected from large open bodies of water and from points on land areas which cannot be associated readily with points on a stream, 2) a stream mileage and indexing method of locating data collection points for storing and retrieving water quality. STORET Subsystem II then, should be considered as a complementary system designed to serve in areas where STORET Subsystems I will not function effectively. Plans are under way to provide a link between the two subsystems so that, as a secondary measure, it will be possible to retrieve data stored in STORET I by using the techniques of STORET II.

The STORET Subsystem I used five cards for the location and description of the sample site. Parameter cards are used to store the water-quality data. On the card, five columns are used for the particular parameter, four columns for the significant figures of the parameter, two columns for the exponent, and one column for remarks. Five parameters can be stored on each card, and any number of cards can be used. Parameter abbreviation cards are used to store the column headings for the data printouts. The parameter control card is used to retrieve particular parmeters, and five station control cards are used to retrieve data from individual stations.

The STORET Subsystem II uses four cards for the location and description of sample sites. Parameter cards are again used to store water-quality data. A parameter control card and two other control cards are used to retrieve the data.

The STORET System is designed primarily to process data from surface-water supplies. The system is flexible, however, can be used for a wide variety of water-quality problems. Data from diverse sources can be stored and several programs have been written to perform analyses of these data.

Montana System. In Montana, with the support from the Montana Water Resources Research Council and the Montana Bureau of Mines and Geology, an investigation of the use of a computer-oriented system for processing ground-water data was begun in 1967. At present, the Montana Bureau of Mines and Geology in cooperation with the U.S. Geological Survey engaged in a program of evaluating the ground-water resources of the state in this ongoing system.

North Carolina System. A computerized system for storage, retrival and

routine processing of hydrologic data has been developed at North Carolina State University. Data on streamflow, rainfall, temperature, snowfall, evaporation and peak flow are stored. System capability permits listing of data and various statistical operations.

Nevada System. A storage and retrieval station was developed by the Center for Water Resources Research in order to centralize data used in research. Data currently being collected are water analyses, well lots, and surface-water discharge.

Texas System. The Texas Water Development Board has been designated by the State legislature to be the centralized agency for collection and dissemination of water data in Texas. The Board is currently in the process of developing means to store and retrieve U. S. Geological Survey stream flow data and NOAA weather data for 52 recording stations in Texas. Their current needs call for retrieving data by date or station number.

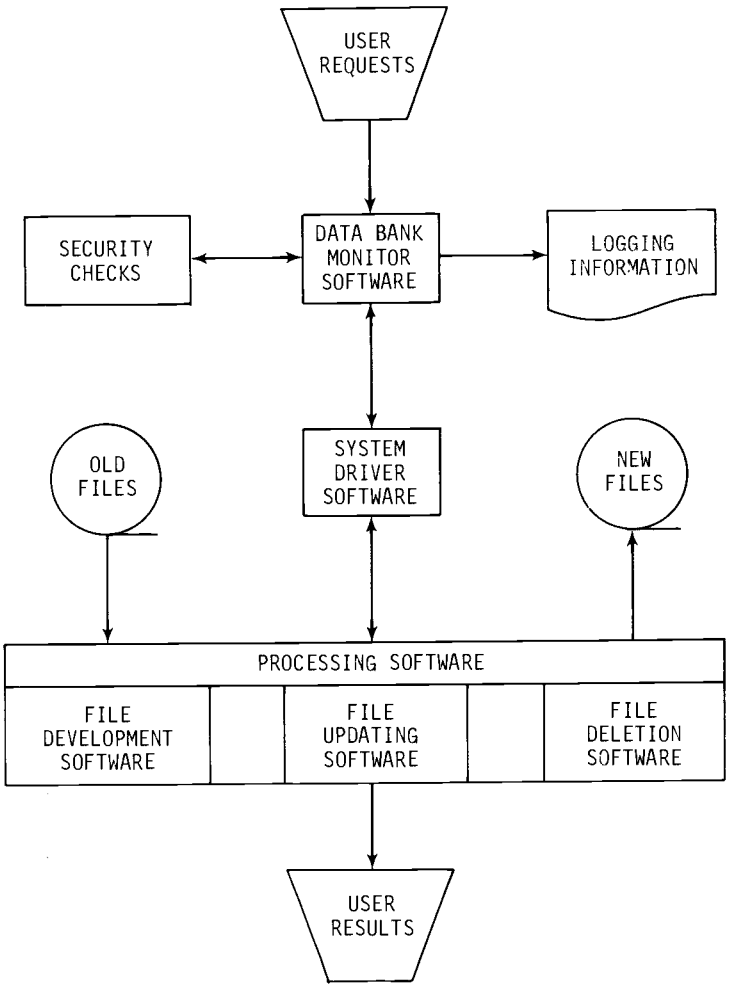

Figure 1. General Design Features of the Hard File Development and Update Sub-System.

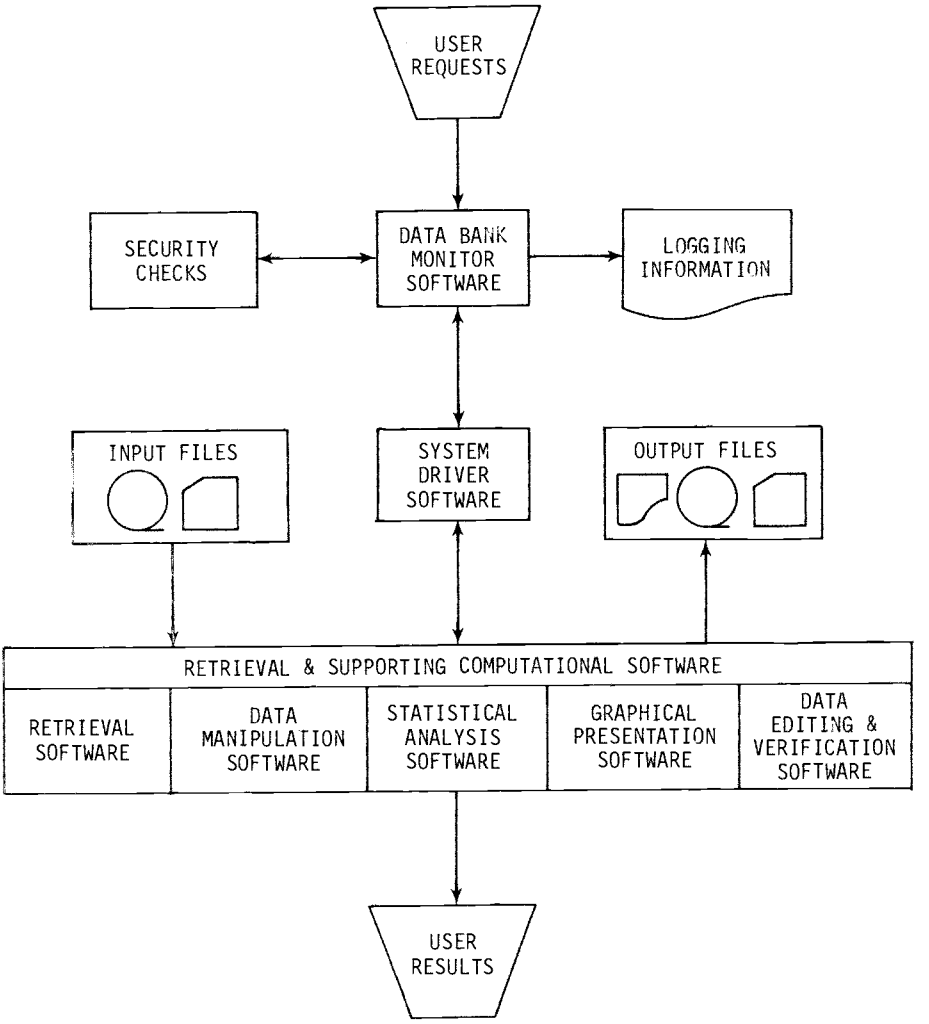

Figure 2. General Design Features of the Hard File Retrieval Sub-System.

ARIZONA WATER INFORMATION SYSTEM

WATER RESOURCES QUESTIONNAIRE

1. Project Code:		2. Title of Project:		
3. Principal Investigator or Personnel in charge:		4. Name and Address Where Activity conducted:		
5. Professional School	6. Period of Operation	7. Level of Funding		
8. Supporting Agency	9. Key Words:			
10. Brief Description of Project (include what data is being collected form, ie. charts, tables, punched card, tape; frequency of collection; will this data be available for introduction into a centralized water information system):				
11. Professional Papers of Articles Resulting from Project (Copies of Reprints would be helpful):				
12. Data Availability:	13. Province:	14. County:		
15. Basin:	16. Township:	17. Range:	18. Section:	19. Longitude-Latitude:

Figure 3, Water Resources Questionnaire