EPA-600/4-76-019
April 1976

Environmental Monitoring Series

MONITORING GROUNDWATER QUALITY: DATA MANAGEMENT

Environmental Monitoring and Support Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Las Vegas, Nevada 89114

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service. Springfield, Virginia 22161.

EPA-600/4-76-019
April 1976

MONITORING GROUNDWATER QUALITY:
DATA MANAGEMENT

by

Norman F. Hampton
General Electric Company— TEMPO
Center for Advanced Studies
Santa Barbara, California 93101

Contract No. 68-01-0759

Project Officer

George B. Morgan
Monitoring Systems Research and Development Division
Environmental Monitoring and Support Laboratory
Las Vegas, Nevada 89114

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
LAS VEGAS, NEVADA 89114

This report has been reviewed by the Environmental Monitoring and Support Laboratory-Las Vegas, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

TABLE OF CONTENTS

(continued)

CONTENTS - Continued

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGMENTS

Dr. Richard M. Tinlin, Dr. Lorne G. Everett, and the late Dr. Stephen Enke of General Electric-TEMPO were responsible for management and technical guidance of the project under which this report was prepared.

The cooperation of many individuals who provided information on existing and proposed data management systems is acknowledged with sincere thanks. In particular, special thanks are due to Mr. Phil Taylor of the U.S. Environmental Protection Agency and to Mr. Doug Edwards of the U.S. Geological Survey.

The following officials were responsible for administration and technical guidance of the project for the U.S. Environmental Protection Agency:

<u>Office of Research and Development</u> (Program Area Management)

Mr. Albert C. Trakowski, Jr.
Mr. John D. Koutsandreas

<u>EMSL-Las Vegas</u> (Program Element Direction)

Mr. George B. Morgan
Mr. Edward A. Schuck
Mr. Leslie G. McMillion
Mr. Donald B. Gilmore

The following personnel of the U.S. Environmental Protection Agency are to be thanked for their review and constructive criticism of the report: Mr. Edward M. Notzon, Monitoring and Data Support Division, Washington, D.C., Dr. Pong N. Lem, Monitoring Systems Research and Development Division, EMSL-Las Vegas, and Mr. Paul Thorpe, Data Services Branch, EMSL-Las Vegas.

SECTION I

INTRODUCTION

This is one of a series of five reports on the general subject of monitoring groundwater quality. The basic report in the group is Monitoring Groundwater Quality: Monitoring Methodology (Todd et al., 1976), which outlines a procedure for creating a monitoring program for groundwater quality under the general supervision of the U.S. Environmental Protection Agency. As an essential supplemental reference to the methodology volume, this report presents the information needed for development and utilization of an effective groundwater quality management program.

PURPOSE

The development of a management information system (MIS) entails the identification of system requirements, system design, organizational design, system procedures design, and if necessary programming, implementation, testing, debugging, documenting, and training. The intention of this report is to identify the system requirements of a comprehensive groundwater quality monitoring program MIS and to survey the existing capabilities which may serve to satisfy those requirements.

For those groundwater monitoring agencies whose needs are not met by existing capabilities, this report presents generic specifications and guidelines for the structuring of a computerized groundwater surveillance data management system. In addition, the inventory of existing data management capabilities (including generalized data base management packages offered by commercial vendors) presented by this report may provide the framework for developing the desired capabilities. The inventory of existing data management systems which is presented here is not intended to be comprehensive. Rather, existing systems were selected for inclusion on the basis of their significance and relevance.

It is hoped that the discussion herein will convey to the groundwater quality manager the scope and breadth of the field of information management systems and that it will expose him to the alternatives available to him in structuring an information management capability suitable to his needs.

SCOPE

Effective groundwater quality management requires that relevant information be available to the decision maker in a concise, comprehensive, timely, economical, and reliable manner. Realization of these goals can be accomplished with the assistance of any one of various tools including file drawers, microfilm, and digital computers. The choice of one of these alternatives will depend, for the most part, on the volume of data involved and the frequency of interaction with the data base.

The discussion here is concerned with the groundwater information management requirements of all levels of governmental monitoring agencies (Federal, State and local). In recognition of the volume of information which is likely to be generated by many of these agencies, this report is directed at outlining a comprehensive computer system capability intended to satisfy these requirements. The system described will afford management of ambient groundwater quality information, percolate quality information, compliance monitoring information, and other data relevant to the management of groundwater quality including citations of groundwater research documentation.

SECTION II

CONCLUSIONS

A nationwide groundwater monitoring program will produce a large volume of highly diversified information. The best use of this information can be realized only if efficient information management is exercised as an integral element of the overall monitoring program.

The prevalent proximity of the groundwater data user to the source of the data as well as the specialized needs of individual users indicates that decentralized (localized) groundwater data management systems are appropriate. A centralized (Federal) data management system is called for as well, however, for the coordination of the national effort (making data available for multiple users and uses and minimizing redundant data collection and analysis activities), the provision of interim groundwater data management support, the achievement of economies of scale, and the encouragement of local compliance with national groundwater monitoring requirements. Consequently, the development of comprehensive groundwater information management capabilities should be undertaken at the Federal, State, and, where necessary, local levels. Whereas the volume of data likely to be involved at the Federal level dictates the need for a computerized system, below the Federal level this is not necessarily so.

A comprehensive groundwater data management capability is composed of three major components: maintaining the data generated by groundwater surveillance, indexing that data so that it can be accessed expeditiously, and maintaining concise citations of relevant groundwater research documentation. At the Federal level these capabilities can be provided adequately by existing or proposed computerized information management systems with only minor modifications. Below the Federal level it may be necessary to develop a computerized capability to maintain groundwater surveillance data, a task which has already been accomplished by some States (e.g., California, Colorado, Tennessee, and Texas). Data indexing and the management of document citations are capabilities which can be provided, to agencies below the Federal level, by existing Federal systems.

SECTION III

RECOMMENDATIONS

The management of groundwater surveillance data at the Federal level can be satisfactorily achieved by application of the Storage and Retrieval (STORET) system currently operated by the U.S. Environmental Protection Agency (EPA). Suggestions for modifications to this system which will improve its effectiveness for managing groundwater data are presented in Section VI of this report. The STORET system is also available to State and local users whose participation is encouraged by the EPA. A system which is designed for a broad-based user population is characteristically not responsive to unique individual requirements, however, and State and local users should consider the merits of developing computerized systems designed specifically for their needs. In addition it should be noted that STORET is not now used on a major scale for groundwater analyses and a major new STORET user community will require a further evaluation and commitment of resources by the EPA.

Groundwater data indexing capabilities, which allow the data user to expeditiously locate pertinent groundwater data and examine its nature prior to accessing the data itself, can be provided to Federal, State, and local users, by the National Water Data Exchange (NAWDEX) proposed and currently being developed by the U.S. Geological Survey. The community of water data collectors and users should support and coordinate with this effort.

The Water Resources Scientific Information Center, U.S. Department of the Interior, provides computerized storage of and access to document citations through use of the Remote Control System (RECON) and the General Information Processing System (GIPSY). These capabilities are available to all categories of groundwater investigators and are generally sufficient to meet their needs.

SECTION IV

GROUNDWATER INFORMATION MANAGEMENT REQUIREMENTS

GENERAL

A complete MIS requirements analysis would call for a survey of the potential users of the system to enable the development of system specifications. Critical factors to be considered by this survey would include the following:

- Information to be managed
- Data volumes
- Frequency of interaction with the data base
- Responsiveness requirements
- Where and how the source information is to be generated
- Required data qualification procedures
- Required output documents.

For this report, an intensive user survey was superseded by the application of gross but reasonably utilitarian assumptions. The assumptions corresponding to the critical factors listed above are:

- Information to be managed will include monitoring station descriptions (i.e., location, hydrogeology, local water use, etc.), physical and chemical measurements of water samples together with sampling dates, and citations of groundwater research documentation.
- The groundwater surveillance data base will be moderately large (expanding monotonically) consisting of millions of data elements requiring extensive storage capabilities. Once the initial data base is established, input data volume will be relatively low and output volume in response to user queries somewhat greater.

- Frequency of interaction (updates and queries) with the data base will be moderate.
- Updating and interrogating the groundwater data base does not require quick system response with several days turnaround generally being adequate. Interrogating information indexing files (water quality data file descriptions and document citations) does require quick system response (i.e., real time), however, to allow for browsing.
- Source information will be generated at locations distributed throughout the U.S. with concentrations in areas of high population density. In general, source information will be generated at locations relatively close to the users of the information.
- Data qualification requirements include input data editing and provision for specific station, sample, and measurement comments to reflect special conditions.
- Output will be alphanumeric text, tables of primary data and computed statistics, and pictorial presentations. Reports will generally be generated on a demand basis with the possible exception of violation reports, associated with compliance monitoring, which may be triggered.

Within the framework established by these assumptions, this section will present a further discussion of the information content of the proposed groundwater MIS as well as a discussion of the fundamental functions to be performed by this system. These functions are data collection, data communication, data organization and storage, data processing, and information retrieval and display. The nature of these functions will be described as well as the alternative technologies available to accomplish them. Those technologies which are best suited to a groundwater monitoring program will be identified.

INFORMATION TO BE MANAGED

An effective groundwater monitoring MIS will be capable of maintaining the following types of data:

- Station Descriptions
- Quality Criteria
- Geologic
- Hydrologic
- Water Quality Parameter Identifiers
- Water Quality Measurements
- Temporal
- Information Qualification Data
- Monitoring Agency Status Data
- Information Indexing

The individual data elements comprising these information categories will be discussed in the following paragraphs. Each data element will be identified as system specific (i.e., applicable system wide), station specific, sample specific, or measurement specific. Further, those data elements which will be required for retrieval or computational operations will be specified as searchable, indicating that they must be stored as formatted data.

Station Descriptions

Station descriptive data consists of information which specifies the station type (i.e., pumped well, unpumped well, unsaturated zone, information monitoring, compliance monitoring, etc.), the party responsible for monitoring the station, a unique station identifier code(s), a unique location (three-dimensional), and directions for locating the station in the field. With the exception of the last item, all of this information should be searchable. Information providing instructions for locating stations in the field can be stored as narrative text along with other special station specific information which is not required for retrievals or computations (e.g., oil lubricated well subject to bearing leakage, continuous-slot stainless steel well screen, etc.).

The groundwater monitoring station type can be specified as coded information in a field of five characters or more. Station type data would be formatted as follows:

1st character

Sample extraction method - pump, bail, or probe

2nd character

Type data - quality, hydrogeologic, and/or DMA status data

3rd character

Type site - municipal, industrial, or other

4th character

Location - saturated zone, zone of aeration, or surface

5th character

Monitoring justification - information, compliance,and/or other

Combinations of attributes can be represented uniquely by coding individual attributes numerically with either a 1, 2, or 4, so that the combination 1 and 4, for instance, could be coded uniquely in one position as a 5.

The designated monitoring agency (DMA) responsible for monitoring the station should be stored as an "agency" code in a searchable field so that, for instance, all stations being maintained by a particular DMA can be retrieved. In addition, the narrative text associated with a station can contain,for example, the names of specific individuals having responsibility for a station together with their phone numbers.

Each station will require a unique identifier code. This identifier will be maintained permanently within the MIS and provide access to station data even if and when a station becomes inactive. Provision should be made for storing and retrieving multiple station identifier codes for the case where a DMA uses multiple codes for alternative retrieval schemes or where more than one DMA is monitoring the same station and station codes have not been standardized.

Station descriptive data to be maintained by the MIS must include information regarding political jurisdiction (e.g., state, county, city, irrigation district, park district, etc.) as well as a unique areal location. To specify a unique areal location, indication of either the township, range, section, etc. or the familiar conventional geographic coordinate system (latitude/longitude) will be most practicable. The degree of precision associated with the measurement of a station's coordinates should also be stored. Additionally, the depths of both the monitoring station hole and intake screen should be stored as station specific information. It should be noted that in cases where, for example, either a monitoring well is equipped with multiple intake screens or a thief sampler is used, individual sample depths may not correspond to either the well depth or existing water level.

Other major station specific information categories not discussed above are applicable quality criteria, geologic data, and hydrologic data.

Quality Criteria

Information pertaining to established quality criteria which a groundwater quality MIS should accommodate as station specific data includes current and projected land use, current and projected water use, demographic data, economic data, designated protected water uses, applicable permit data (compliance dates and monitoring requirements - parameters and frequency), and water quality criteria (either ambient or discharge limitations).

Demographic and economic data as well as current and projected land and water use in the neighborhood of a monitoring station is information, typically generated by local planning agencies, which reflects the significance of groundwater pollution in the environs of a monitoring station. This information need not be used for retrieval or computational operations and consequently can be satisfactorily stored in the narrative text associated with each monitoring station.

The development of a comprehensive groundwater quality monitoring program will entail the systematic identification and inventory of principal aquifers and, preferably, the designation of protected uses for these aquifers. In the process of developing the inventory of principal aquifers, full use should be made of the "Catalogue of Aquifer Names and Geologic Unit Codes" compiled by the Office of Water Data Coordination (OWDC), U.S. Department of the Interior (USDI) (Price and Baker, 1974). Aquifer protected use designations would be codified and searchable. Protected use categories would include public water supply, agricultural and industrial use with allowance made for the possibility of subcategories of the latter two.

Permit data, other than imposed discharge limitations, should not be required for retrieval or computational operations and can, therefore, be stored in the narrative text associated with compliance monitoring stations. Information content would be similar to that contained in NPDES applications and permits including permit numbers, compliance dates, and monitoring requirements. If permit numbers are required for search operations, they can be used as secondary station identifiers.

Permit specified discharge limitations and/or the water quality criteria associated with the designated protected uses established for an aquifer can be stored with the characteristics of each monitoring station as appropriate. Ambient quality criteria to be stored may be those published by EPA in "Proposed Criteria for Water Quality" (EPA, 1973) with provision made for updating these criteria as they are modified. Although it is not likely that they will be needed as record keys, the inclusion of discharge limitations and ambient water quality criteria within the monitoring data base as searchable information will allow efficient generation of exception reports.

Geologic Data

In order to uniquely identify the source of groundwater samples, some geologic data is required, in addition to geographical coordinates, to specify the aquifer from which the sample originated. In the case where a monitoring station taps more than one aquifer, aquifer identification is particularly essential and must be provided as sample specific (i.e., input in conjunction with each set of water quality analysis data) rather than station specific data. The requirement for providing aquifer identification can be satisfied by storing the established aquifer name, if available, or else the geologic formation name and age associated with the monitored aquifer (e.g., Mount Simon formation - Cambrian age or glacial drift - Pleistocene age). It should be pointed out that the latter form of identification is not preferred since aquifers and geologic formations do not necessarily coincide completely. Aquifer identification can be codified and standardized, and search operations facilitated by application of USGS proposed modifications to the stratigraphic coding system developed by the American Association of Petroleum Geologists (Price and Baker, 1974).

Additionally, information regarding the physical properties and chemical constituency of the water bearing materials (aquifer, unsaturated zone, or topsoil) may be necessary, particularly if the synergistic effects between these materials and introduced pollutants are to be modeled. This information may reflect material type and waste attenuation characteristics. If a model is to be computer accessible by the groundwater quality MIS, then the information required by the model should be searchable. Otherwise, it can be stored with the narrative text associated with each station description. Frequently, information regarding the characteristics of water bearing materials is generated by drillers during the installation of a well and is available in the form of well logs.

Hydrologic Data

An efficient groundwater quality monitoring system will require an MIS capable of accommodating a wide range of hydrologic information. In general, this type of information has previously been determined, particularly in areas of rigorous groundwater development, and a groundwater quality monitoring program will only demand gathering and storing it. Hydrologic information is necessary to the monitoring program to predict the movement of pollutants, isolate the source of the pollution, and interpret the relationship between groundwater and surface waters.

Most hydrologic information will be station specific and can, therefore, be stored concurrently with the establishment of station descriptions in the data base. In cases where many stations penetrate the same homogeneous medium, it may be possible to store the characteristics of that medium under only one station together with a list of the other stations common to that medium. Major hydrological data elements will include the following:

- Water bearing material depth, thickness, and areal extent
- Permeability
- Aquifer transmissivity and storage coefficient
- Hydraulic gradient (vector)
- Water table elevation (sample specific)
- Area and magnitude of natural and artificial recharge and discharge
- Station sampling device (e.g., pumped well, suction lysimeter, neutron probe, etc.) operating characteristics

Hydrologic measurements required for computations such as to determine hydraulic diffusivity or specific flux will be required to be stored as searchable information.

Water Quality Parameter Identifiers

The selection of the water quality parameters to be maintained in a groundwater monitoring MIS poses one of the principal design considerations related to the development of the system. This is because of the large number of candidate variables. In many information systems, the data description (i.e., the variable identification) is imbedded in the program logic. However, because of the large number of variables involved in groundwater monitoring, a generalized data storage system is more appropriate. This requires that data identification be independent of the programs, that is, the data descriptions must themselves be data inputs to the system. Consequently, the list of water quality parameters maintained can be virtually open ended.

Stipulating the types of quality measurements to be included in a monitoring system is extremely difficult, due to the large number of potential contaminants involved. In 1972 the National Academy of Sciences (NAS) published "Water Quality Criteria - 1972" at the request of and funded by the Environmental Protection Agency (EPA). Subsequently these recommendations were presented nearly intact by the EPA in "Proposed Criteria for Water Quality" (EPA, 1973).

The National Academy of Sciences report propounded criteria which would serve to preserve water quality for the following purposes:

- Public Water Supplies
- Agricultural Uses
- Industrial Uses
- Recreation and Aesthetics
- Freshwater Aquatic Life and Wildlife
- Marine Aquatic Life and Wildlife

In general, only the first three of these would be affected by groundwater quality. The criteria proposed by the NAS for these three use categories and those imposed by U.S. Public Health Service (USPHS) water standards would serve as a framework for identifying significant water quality information to be provided by a groundwater information management system (USPHS, 1962). A composite list of the parameters for which the NAS and USPHS have established criteria regarding public, agricultural and industrial use is presented in Table 1.

The set of quality parameters to be examined by any individual groundwater quality monitoring program would, for the most part, be a subset of Table 1, which could be considered as a menu of water quality parameters. The sample set to be surveyed at any one groundwater quality surveillance station could be selected, at least partially, from this menu. Additional parameters, not appearing in Table 1, might be included as dictated by specific situations.

The justification for presenting Table 1 as such a menu rests with the fact that the NAS and the USPHS deem these parameters to be significant, as it is these parameters for which criteria have been developed. The list in Table 1 is by no means exhaustive, however. The inadequacy of the list for compliance monitoring purposes is reflected, for example, by "The Toxic Substances List" published in 1973 by the U.S. Department of Health, Education and Welfare (HEW); National Institute for Occupational Safety and Health. This document identifies 11,000 "toxic," chemically unique, substances (HEW, 1973). It is reasonable to assume that any one of these substances could find its way to a subsurface water reservoir, either by intentional or unintentional introduction, and achieve significance. A groundwater information management system would be required, therefore, to be flexible enough to accommodate a large and inconsistent set of variables.

As stated previously, a centralized groundwater quality MIS is called for to provide support of local efforts. In general, however, a centralized data repository would require more succinct and less detailed information than would be required by decentralized (localized) data banks. Compendiousness can be accomplished by summarization, aggregation, and the use of status indicators. The Council on Environmental Quality has funded (jointly with EPA and USGS) an ongoing study entitled "Comparative Evaluation of Techniques for the Interpretive Analysis of Water Quality" which will provide methodologies for generating concise data and will help to satisfy the inherent requirements of the centralized system component.

TABLE 1. MENU OF CANDIDATE WATER QUALITY PARAMETERS FOR GROUNDWATER MONITORING.

Alkalinity($CaCO_3$)
Ammonia
*Arsenic
a Aluminum
*Barium
Boron
*Cadmium
Chloride
*Chromium (total)
Color (eg. platinum-cobalt color units)
Copper
*Cyanide
Dissolved Oxygen
Fluoride
Foaming agents (MBAS)
Hardness
Iron
*Lead
Manganese
*Mercury
*Nitrate-Nitrogen
+Nitrilotriacetate (NTA)
Odor
Oil and grease
Organics-Carbon Adsorbable
*Pesticides
  Insecticides-Chlorinated Hydrocarbons
  Insecticides-Organophosphate and Carbamate
  Herbicides-Chlorophenoxy
*Nitrite-Nitrogen

pH
Phenolic compounds
i Phosphate
+Phthalate Esters
+Polychlorinated Biphenyls (PCB)
Radioactivity
*Selenium
*Silver
i Silicon
Sulfate
i Suspended Solids
Temperature
Total dissolved solids (TDS)
Turbidity
Viruses
Zinc
*Carbon Chloroform(extractable)
a Beryllium
*Total Coliform
*Fecal Coliform
a Bicarbonates
a Cobalt
a Lithium
a Molybdenum
a Nickel
a Sodium
a Vanadium
i Calcium
i Potassium

---

i Industrial impact only
a Agricultural impact only
+ No criteria currently established
* Significant health ramifications

Water quality parameter identifiers will be codified and system specific. Since water quality parameters are system specific, the system administrator rather than the DMAs will have responsibility for depositing and maintaining this type of data in the groundwater MIS. An individual DMA can establish a special parameter identifier by petitioning the system administrator who will judge the validity, redundancy, and applicability of the new parameter before including it in the data base.

Each water quality parameter identifier entry will consist of two data elements. One will be an alphanumeric descriptor reflecting the common name of the parameter and the units of measure in which numeric measurements associated with that parameter will be reported. In order for the system to accommodate various units of measure, it will be necessary to assign different parameter identifiers for each one. The second data element comprising a parameter identifier will be the system administrator assigned numeric code associated with that identifier. Every effort should be made to organize these codes in a hierarchical fashion so that parameters of a similar nature will be grouped together.

Water Quality Measurements

The results of physical and chemical analyses of groundwater, soil, and geologic material samples will be stored as water quality measurement data which will represent the bulk of the information to be managed by the groundwater MIS. This information will be required for both retrieval and computational operations and must, therefore, be stored as searchable data. Each measurement data element is measurement specific and must be stored in conjunction with information which specifies the parameter measured (parameter code), the sample analyzed (sample date), and the station sampled (station identifier code). Efficient utilization of the fields set aside for analytical measurements can be realized by also using them to store sample specific data such as sample depth or sample specific reliability indicators.

Temporal Data

In order to provide reasonable utility, a water quality information system must be capable of reflecting trends. This would require maintaining water quality data as time series. Water quality data updates need, therefore, to be appending operations rather than destructive updates. Consequently, a water quality data base can be expected to grow monotonically and linearly (if fluctuations in the number of stations and parameters observed are disregarded).

If water quality data are collected at a constant frequency, it is only necessary to store the data collection rate and initial collection date once for each station (as station specific data). It would also be essential to make provision for entering information regarding interruptions in the period of record.

When data are not collected at a constant frequency, which is most often the case with groundwater monitoring, the date of sampling must be recorded as sample specific data with each new set of water quality measurements which is input. Provision for storing dates as searchable information must be incorporated into a groundwater monitoring information system so that any subset of the period of record data set may be retrieved. It should be noted that in contrast with surface water monitoring, recording calendar dates is usually sufficient to fix the location of a groundwater monitoring sample in time (i.e., clock times are not required). This is due to the far less dynamic nature of groundwater phenomenon.

In situations where significant vertical stratification of water chemistry is present it will also be necessary to record and store the pumping time in hours prior to the collection of either a simple grab or composite sample. Additionally, in the case of composite samples taken over time, it will be necessary to record and store the duration (in hours) of the composite sampling period.

Information Qualification Data

There is a cogent need for a data qualification capability in any water quality monitoring information system. To accomplish this the system should include, in addition to data verification, a comprehensive edit function, preferably computerized, which would operate prior to data storage. The edit check can be based on comparison of input data with previous trends, allowable data ranges, and established units of measure. Data failing any one of these checks should not be modified but rather flagged and reported as suspect. The capability to compare input data with allowable ranges imposes an additional data requirement which can be satisfied by storing these ranges as station specific, searchable data.

Improvements in the value of a data base can also be attained by allowing "reliability indicators" to be input and stored as nonsearchable data. These indicators could be of the type that reflect, for example, station performance anomalies, unusual sampling conditions, unusual methods of measurement, measurement precision, or which reflect qualitative judgments of the "goodness" of data. Reliability indicators should be stored either as station specific (in the narrative text), sample specific (as a water quality measurement) or measurement specific (in a special field) as appropriate.

DMA Status Data

A nationwide or statewide groundwater quality monitoring program may involve the periodic inspection of DMA facilities to determine the "operational status" of monitoring programs and equipment. In addition, where a DMA or other agency has groundwater pollution control functions, the "readiness status" of a control unit in terms of its ability to respond to a pollution incident may also be evaluated. Consequently, a comprehensive groundwater quality MIS should be capable of maintaining this type of information. Most efficiently, a DMA or pollution control unit would be regarded as a station by the MIS, an inspection tour regarded as a sampling iteration, and status data as water quality measurements with parameter codes being established accordingly.

The operational status of a DMA will be estimated based upon its ability to monitor the stations, parameters, and at the frequencies required. An operational status index could be established where, for example:

1 = 100 percent monitoring effectiveness

2 = 90 percent monitoring effectiveness

3 = 80 percent monitoring effectiveness

4 = 70 percent monitoring effectiveness

5 = 60 percent monitoring effectiveness or worse.

A "readiness index" could be formulated which would reflect the ability of a DMA or other pollution control unit to respond to a pollution incident. This index would be a function of personnel on hand, personnel training, equipment on hand, and equipment reliability. The "readiness index" could take the form of a numeric grade where, for example:

1 = able to respond effectively within 1 day

2 = able to respond effectively within 2 days

3 = able to respond effectively within 3 days

4 = able to respond effectively within 4 days

5 = able to respond effectively within 5 days or more.

Estimating the operational and readiness ratings of individual DMAs or pollution control units would be the responsibility of the national or state groundwater quality monitoring program administrator.

Information Indexing

Information indexing allows ready access to abstracts of existing data sets. The groundwater quality MIS should provide indexing of two major categories of data sets: water quality data files present in the MIS data bank, and groundwater research documentation.

Water quality data file abstracts will provide information regarding activities at each station in the monitoring program and should be accessible by station identifier, geographical coordinates, aquifer code, political jurisdiction, station type, or agency code. Information contained in the water quality file abstract will be station specific and would include parameters monitored, monitoring frequency, and period of record. All of the information required by the water quality data file index will exist elsewhere in the MIS so that this index can be system generated and will not require user input.

Research documentation indexing will require special user input. Data elements to be stored, all of which should be searchable, will be document titles, author names, report numbers (access numbers), performing and sponsoring organizations, report dates, textual abstracts, keywords, and geographical area of interest.

Summary

Table 2 presents a list of the data elements to be managed by the groundwater quality MIS.

DATA COLLECTION

Data collection, in the context of MIS design, is the process of translating information into machine readable form. The primary factors considered in selecting data collection systems are purchase cost, operating cost, reliability, responsiveness, and minimizing the bottleneck created by relatively high internal computer processing speeds and low input speeds.

Total MIS expenditures are particularly sensitive to data collection costs since data entry typically accounts for 20 to 40 percent of electronic data processing costs (Ferrara and Nolan, 1973). In addition, the data entry process represents the single greatest source of error in a MIS. The significance of the imbalance between input speeds and central processing

TABLE 2. SUMMARY OF INFORMATION TO BE MANAGED BY GROUNDWATER MIS.

1. System Specific Data

   - Water Quality Parameter Names
   - Units of Measure
   - Parameter Codes

2. Station Specific Data

   - Station Type
     - Sample Extraction Method
     - Type Data
     - Type Site
     - Location(i.e.,saturated zone, unsaturated zone, or surface)
     - Monitoring Justification
   - Responsible Monitoring Agency
   - Station Identifier Code(s)
   - Geographic Coordinates and Associated Measurement Precision
   - Station Location (township, range, section, etc.)
   - Station Depth (hole depth and screen depth)
   - *Field Location
   - *Responsible Individual
   - *Station Specific Information Qualification
   - Quality Criteria
     - *Demographic and Economic Data
     - *Land Use
     - *Water Use
     - Permits
       - *Stipulated Monitoring Program (Parameters and Frequency)
       - *Compliance Schedules
       - Discharge Limitations
     - Ambient Criteria
   - Political Jurisdiction Code
   - Geological Data
     - Aquifer Code (may be sample specific)
     - Geochemical Information
   - Hydrologic Data
     - Aquifer Depth, Thickness, and Areal Extent
     - Aquifer Transmissivity and Storage Coefficient
     - Hydraulic Gradient
     - Permeability
     - *Area and Magnitude of Natural and Artificial Recharge and Discharge
     - *Station Sampling Device Operating Characteristics

Table 2 - Continued

3. Sample Specific Data

    Sample Date
    Pumping Duration
    Composite Sample Duration
    Sample Depth
    Water Table Elevation
    *Sample Specific Information Qualification

4. Measurement Specific Data

    Physical-Chemical Analyses
    *Measurement Specific Information Qualification

5. DMA Status Data

    Monitoring Effectiveness Index
    Pollution Control Readiness Index

6. Information Indexing

    Water Quality File Abstracts

        Parameters Monitored
        Monitoring Frequency
        Period of Record

    Research Documentation Citations

        Titles
        Authors
        Report Numbers
        Performing and Sponsoring Agencies
        Report Dates
        Textual Abstracts
        Keywords
        Geographical Area of Interest

---
*Not searchable

unit (CPU) speeds can be illustrated by the fact that a keypunch operator can punch and verify roughly four cards a minute, a card reader can read about 1,000 cards a minute, and a moderately sized CPU can process about 100,000 cards a minute or more (Schwab and Sitter, 1969).

There is a wide variety of available capabilities which will provide automated support of the data collection phase of a computerized MIS. These include conventional keypunch, buffered keypunch, key-to-tape, key-to-disc, remote "dumb" terminals, remote "intelligent" terminals, mark sensing, magnetic ink character recognition, and optical character recognition (OCR) devices. These nine options are listed more or less in order of increasing implementation cost and, correspondingly, increasing speed and reliability. The devices listed all have applicability to groundwater data entry. Selection of equipment by each groundwater data depositor will depend primarily upon the magnitude of data flow. If necessary, to further minimize the bottleneck which can occur at the data input interface, buffered input units and overlapped input systems can be installed at the centralized groundwater computer data bank.

An additional category of devices available to the data depositor, which has particularly attractive applicability to groundwater monitoring, is source data automation. Source data automation is the process of capturing primary data in machine readable form. Examples of such equipment are automatic digital recorders used in conjunction with Keck groundwater level recorders, automatic laboratory chemical analysis equipment and robot water quality monitoring stations. The advantages of source data automation are that it produces data which are easily converted into other machine-useable form, reduces the opportunities for introducing errors, and lowers clerical costs.

DATA COMMUNICATIONS

User interaction with a management information system can be segregated into four major activities: 1) file creation; 2) file updating; 3) information requests; and 4) information reception. Computerized management information systems accomplish these functions in one of two modes: 1) batch; or 2) real-time/interactive.

User access to the groundwater surveillance data base should be in the batch mode whereas access to the information index system component should be in the real-time mode, at least for retrievals. Although user interaction with a batch processing system allows optional use of telecommunication links with the system, telecommunication is mandatory for real-time processing.

A telecommunication link requires a terminal to enter data, modems to encode (in a form acceptable to the transmission channel) and decode data, and a transmission channel. Transmission channels can be ordinary telephone services such as provided by WATS (best suited to widespread, high volume data flow), dial-up service such as provided by TWX or TELEX (best suited to widespread, low volume data flow which is likely to be the case for groundwater surveillance), or dedicated private line (best suited to high volume data flow concentrated between a few points)(House, 1974). The major factors to be considered in the selection of a transmission service will be responsiveness, reliability, and implementation and operating costs. Data security will not be a significant consideration for groundwater surveillance information.

An ideal groundwater monitoring information system will provide flexible data flow procedures for both data submission to and data retrieval from the groundwater surveillance data base. The requirement for flexible data flow procedures is imposed by the desirability of wide system usage and the likelihood that data depositors and data users will have variable transmitting and receiving capabilities. It is important to note that access to the groundwater quality data base should be provided to users with unsophisticated communication capabilities as well as to those users with highly sophisticated capabilities.

All information management systems benefit from the responsiveness of real-time access to computerized data bases. However, provision of real-time capabilities does not always result in the most efficient allocation of data management resources. Since the management of groundwater surveillance data does not necessitate dynamic information flows relative to many other information management functions, real-time systems are not included in the recommendations presented below.

Data collectors should be allowed to submit groundwater data for storage (both file creation and file update) in the following modes:

- Formatted nonmachine readable
- Formatted machine readable (i.e., punch cards, paper tape, or magnetic tape)
- Remote access batch (i.e., teletype of card reader)

Data users should be allowed to request groundwater quality data via the following modes:

- Telephone inquiry
- Letter inquiry
- Teletype batch inquiry

The system should be capable of transmitting data retrievals in any of the following modes:

- Nonmachine readable hardcopy
- Punch cards
- Dial-up remote teletype or remote printer (batch)
- Magnetic tape (to promote intermachine compatibility, options for number of tracks, bits per inch, parity convention and blocked or unblocked output should be provided)

Figure 1 is a diagram showing user access to the proposed groundwater MIS as well as interfile data flow. Unary data is information not subject to update except where errors necessitate corrections. Multiple data is information subject to update (time series data) and therefore multiple data flow channels are likely to support a high volume of data traffic. The content of the files depicted in Figure 1 is described in the following section which discusses data storage requirements.

DATA STORAGE

The development of the data storage component of a MIS entails the selection or fabrication of hardware devices, data organization schemes, and data base management software packages. Factors to be considered include cost, storage space, response time and current and future use of information stored.

Three general classifications of hardware are available for data storage: internal, secondary, and external. Internal storage is best utilized for holding programs and data being immediately executed. Internal storage media include magnetic core, thin films, magnetic rods, and plated wire devices all of which are characterized by high access speeds and costs. Secondary storage is not an integral part of but is directly connected (on-line) to the CPU. Secondary storage devices include magnetic disc, drum, card, and tape peripherals characterized by moderate access speeds and costs. External storage is not directly connected (off-line) to the CPU. External media include removable disc packs, magnetic tape, punched cards, and paper tapes all characterized by low access speeds and costs (Lobel and Farina, 1970).

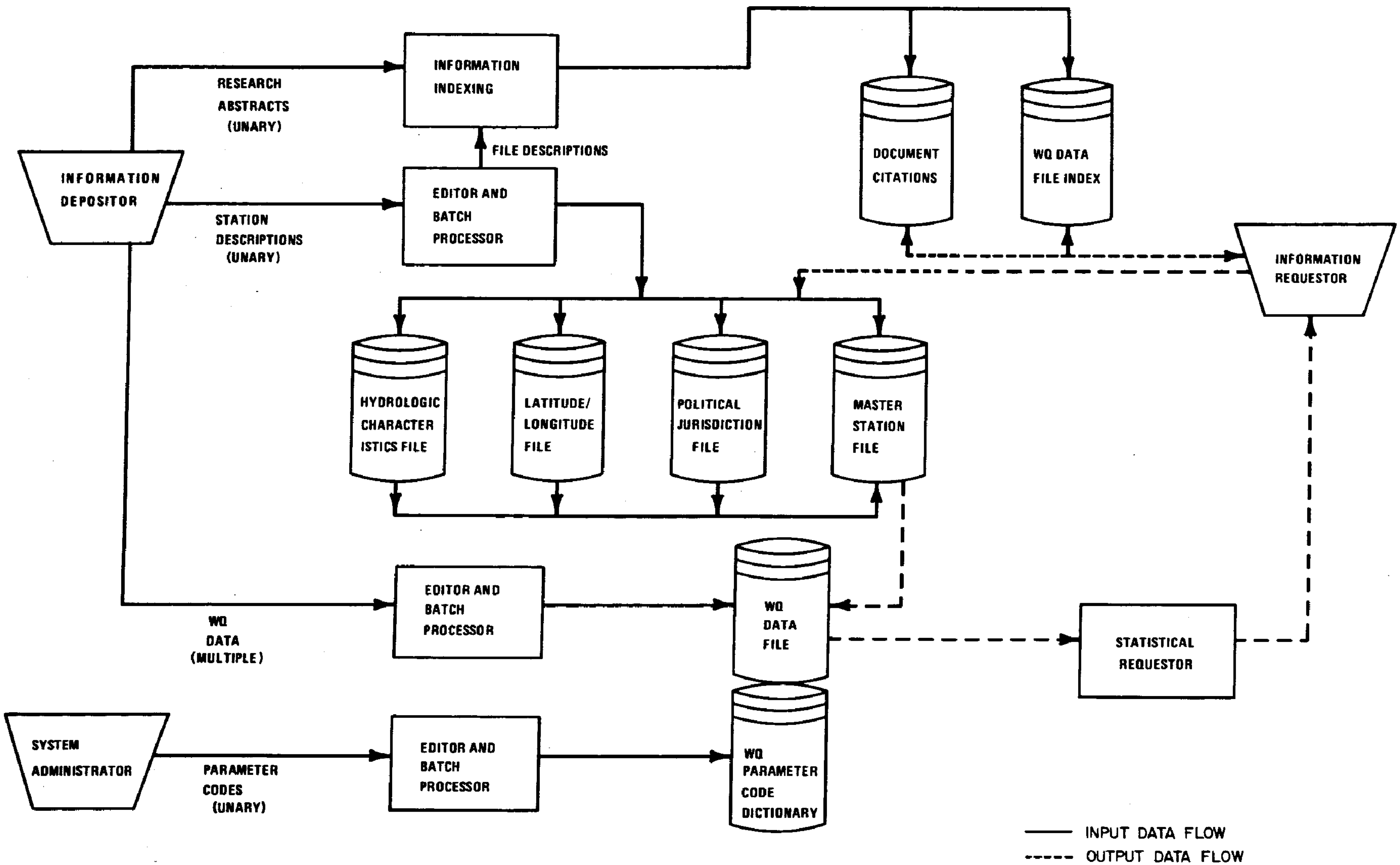

Figure 1. User access to groundwater data base.

Figure 1 depicts all of the groundwater data files as being in secondary storage and resident in on-line magnetic disc or drum, both of which provide random access. Magnetic cards could also be used but they are not widely compatible. Although drum storage allows access speeds nearly an order of magnitude greater than disc, disc storage is adequate for storing groundwater data and will provide significant storage cost savings compared to drum storage. Additional storage cost savings can be realized if removable disc packs are used (as external storage) and placed on-line only during certain time intervals and if certain low priority data sets (e.g., seldomly accessed water quality data) are structured for sequential access and archived on off-line magnetic tapes.

Data files are structured using one or a combination of three basic organizational concepts: sequential, random, and list. Sequential files store records in a specified sequence relative to other records so that the next logical record is also the next physical record. Sequential organization permits rapid access to a series of records logically related to one another but is cumbersome for updating and retrieving individual records out of sequence.

Random organization requires the establishment of a predictable relationship between a record key and the direct address of the location where the record is stored. In most cases this will require a "dictionary look-up" process as part of each record retrieval. Random access allows rapid retrieval of individual records or data items where only a small portion of the data file is affected but is not well suited to retrievals of multiple records.

List structures (simple, inverted or ring) incorporate pointers in each record which point to other records that are logically related to the first record. Of particular applicability to the management of groundwater data are inverted list structures which make every data element available as a record key. For instance, a station type code could be used as the key to a record which contained pointers to every station of that type. The inverted list approach allows very rapid (and therefore inexpensive) retrievals but requires a great deal of storage and does not foster easy file updates. Therefore, inverted list structures can best be used for files which are small and which are frequently accessed but infrequently updated.

As shown in Figure 1, unary station descriptive data should be stored in four separate, directly accessible disc files as follows:

- The Hydrology file will contain hydrologic characteristics of water bearing media and sampling devices as listed in Table 2.

- The Latitude-Longitude file will list each station number by its latitude and longitude.

- The Political Jurisdiction file will list each station number by its associated political jurisdiction code (state, county, city, etc.).

- The Master Station file will contain all station specific data including station specific narrative text.

The first three files described above can best be structured as inverted lists since they will likely be frequently accessed and infrequently updated. The Master Station file can be random using station identifiers as record keys.

The Groundwater Data File, shown in Figure 1, will reside in disc storage. This file will contain all sample specific and measurement specific groundwater surveillance data as well as DMA status data (see Table 2). The Groundwater Data File can be organized as a random file also using station identifiers as record keys.

The Parameter Code Dictionary should also reside in disc storage and be structured as a random file using parameter codes as record keys.

The Groundwater Data File Index should be random and use station identifiers as record keys. The Document Citation file will actually consist of a number of randomly accessible subfiles. The Master Document Citation file would contain all information regarding each document and would be accessible by report numbers which would serve as record keys. Additional files would list report numbers by and, correspondingly, be keyed by document title, author, agency, etc.

DATA PROCESSING

Computerized data processing is accomplished either in batches or on a continuous (real-time) basis. Batch processing requires the accumulation and preprocessing of a group of transactions all of which will be computer processed at one time. Real-time processing, on the other hand, accepts and processes transactions as they occur. Both processing modes can accept input data from either remote or local terminals. The basic difference between the two processing methodologies, as seen by the system user, is the difference in response time with the turn-around time for real-time processing being significantly quicker.

Real-time processing should be implemented only where rapid system response is really needed since batch processing permits more efficient and economical hardware utilization by requiring less system redundancy. Therefore, only accession to the groundwater information indexing components of the groundwater MIS requires real-time processing. This requirement is imposed by the users' need to interact intellectually (browse) with the information indexing data base.

Batch processing associated with access to the Groundwater Data File will be composed of editing, sorting, storing, retrieving, and statistical operations. Input editing will examine input data for format errors, check the validity of codes, (parameter codes, aquifer codes, etc.) and compare water quality data with acceptable ranges. For compliance monitoring the input editing module can also be used to compare water quality data with established water quality standards and prepare violation reports as necessary. The sorting and storing processes will organize the data and update the appropriate files. The retrieval commands, access the appropriate data files, organize the requested information, and format output reports. The statistical processor would function in conjunction with the retrieval routines to operate on raw data as designated by the information requestor. The statistical processor would be required to generate extreme values, first and second moments, regression and correlation coefficients, logarithms, daily loading (for source monitoring), and coordinates necessary to create plots.

DATA RETRIEVAL

Data retrieval is the process of translating information which is meaningful only to machines into a form which is meaningful to humans. Designing the data retrieval component of an MIS requires identifying the information to be output, specifying the retrieval procedures acceptable to the system, developing the required retrieval software, determining output formats, and selecting hardware.

The data retrieval component of the proposed MIS which accesses the Groundwater Data File will be required to yield both alphanumeric and pictorial output. The system should be capable of providing alphanumeric output which will include the following types of information:

- Measured parameters
- Number of observations
- Beginning and ending sampling dates
- Raw data
- Minimum and maximums
- Arithmetic means
- Standard deviations
- Regression coefficients
- Correlation coefficients
- Percentiles
- Confidence intervals
- Daily loadings
- Logarithms
- Station descriptive paragraphs

User requests for pictorial (graphic) displays may require the following types of plots:

- Physical-chemical variations with time
- Physical-chemical variations with sample depth
- Monitoring stations located geographically
- Physical-chemical variations with distance
- Vertical bar charts
- Circular diagrams
- Radial vector diagrams
- Pattern diagrams
- Trilinear diagrams

The last three information presentation techniques listed above, which may be unfamiliar to some readers, are described by Hem (1959).

The groundwater monitoring MIS can offer the data user the most powerful capabilities if it can provide a wide range of useful retrieval procedures. A retrieval procedure is characterized by the information required by that procedure as user input to the system to enable the system to locate data and generate output.

The groundwater monitoring MIS user should be able to request data from the system by specifying one or a combination of the following information elements:

- Station number
- Range of station numbers
- Latitude and longitude
- Polygon (specified by the latitude and longitude of its vertices)
- Political jurisdiction
- Sampling date
- Range of sampling dates
- Sampling depth
- Range of sampling depths
- Monitoring agency
- Maximum or minimum parameter values

The user should be able to implement a number of these procedures in conjunction with each other so that Boolean retrieval strategies can be applied. In addition he should be able to request that various data manipulation and statistical operations be performed and to dictate, to some extent, the format of the output which he receives.

Factors involved in the selection of data retrieval hardware include considerations of speed, cost, flexibility, reliability, noise, number of copies needed, and formatting (i.e., requirements for number of characters per line, number of lines per page, and plot sizes). Retrieval hardware can be categorized according to the following distinctions:

- Impact, non-impact, cathode ray tube (CRT), digital plotter, microfilm, or voice response.
- Serial, which produces 10 to 200 characters per second (cps) or parallel, which produces 300 to 10,000 cps (Lorber, 1972).
- Full character or dot matrix.

In general, impact printers produce full characters either one at a time (serially) or a line at a time (parallel). Impact printers provide good legibility and multiple copies (a constraining factor for many applications) but, in general, are noisy and subject to relatively frequent breakdowns because of the large number of moving parts which they require.

Non-impact printers will best satisfy the requirements of accessing groundwater monitoring data. Non-impact printers can be either serial or parallel printers which, in a majority of machines, produce dot matrix characters. Ink-jet and electrostatic printers are two types of non-impact printers which offer speed, reliability, portability, competitive purchase cost, and quiet operation. The disadvantages which are normally characteristic of these devices are high operating costs (e.g., electrostatic printers require special paper), the inability to produce multiple copies, and slightly poorer image quality than is provided by impact printers.

CRT displays produce dot matrix characters, either serially or in parallel as well as graphics. Although CRT displays, themselves, are unable to generate permanent records they are fast, reliable, and economical to purchase and operate. In addition, these devices afford great flexibility by virtue of the optional peripheral equipment which may be attached, such as hard copy output, light pens, and information storage capabilities. CRT terminals would be most appropriate for accessing groundwater information indexing files.

Digital plotters which produce permanent graphic displays, are available at a wide range of prices and, correspondingly, with a wide range of capabilities. Microfilm systems can receive output directly from a CPU, via either paper to film or CRT to film, and provide the advantages of a compact, inexpensive, external storage medium. Microfilm systems generate output in the form of microfilm (normally 16 mm film), aperture cards (normally 35 mm film), or microfiche (which records many pages of data on one frame of film).

With the exception of voice response units, which are used most extensively by operations which interface with the public, any of the above mentioned hardware options may find appropriate applications in a groundwater monitoring program. The selection of specific retrieval hardware components will depend upon the requirements of individual data requestors and of interfacing with the central system. The central system should be designed to be flexible so that it represents a minimal constraint on the selection of user output hardware.

SECTION V

EXISTING SYSTEMS

GENERAL

This section presents a survey of existing or proposed information management systems which are relevant to the management of groundwater monitoring information. Table 3 lists some of the water resources data management systems which are currently operational together with some of their more pertinent characteristics. Table 4 presents a selection of computerized information indexing systems, both operational and proposed, which provide data file or research documentation abstracts. Table 5 presents several generalized data base management packages, offered by various commercial vendors, which may afford capabilities suited to the needs of specific groundwater data management efforts.

The discussion which follows describes in further detail some of the more pertinent systems listed in Tables 3 and 4. Readers with particular interest in one of these systems are referred to the associated users and systems documentation.

STORET

The Storage and Retrieval System (STORET) was developed initially by the U.S. Public Health Service and is currently operated by the U.S. Environmental Protection Agency, where it is undergoing further development. This system is intended to provide federal assistance to the states in the performance of water quality management and to insure compliance with PL 92-500. Table 6 presents a list of those sections of PL 92-500 which are supported by STORET. Providing the states access to a centralized information retrieval system realizes economies primarily in the areas of system maintenance and user assistance. To date, 42 of the states are utilizing STORET.

The STORET system consisted of two basic files: the Water Quality File (WQF) and the General Point Source File (GPSF). Primarily because of high operating costs, the GPSF was deactivated during February of 1975 and is to be replaced by a less expensive but also less powerful generalized information retrieval system called the "Interim Enforcement System." One aspect of this interim measure will be the provision of the capability to store self-monitoring and compliance data in the WQF with each discharger being treated as a station and SPDES permit numbers serving as station identification numbers.

TABLE 3. EXISTING ENVIRONMENTAL DATA MANAGEMENT SYSTEMS.

System Name	Acronym	Administrator	Information	Storage Location	Groundwater	Computer System
Storage and Retrieval System (1)	STORET	EPA	Water quality	Centralized	43,000 wells	IBM 371/58-OS/MVT
National Water Data Storage and Retrieval System (2)	WATSTORE	USGS	Surface and ground water physical and chemical data	Centralized	25,000+ wells	IBM 370/155
ORSANCO Robot Monitor System (3)	-	Ohio River Valley Water, Sanitation Commission	Surface water quality	Centralized	None	IBM 1130
Groundwater Quality System (4)	-	California	Groundwater quality and hydrographic	Centralized	1,400 wells	CDC 3300
Water Information System for Enforcement (5)	WISE	Michigan DNR	Water quality and discharge inventory	Centralized	22 WQ wells	Burroughs B5500
Tennessee State Groundwater Data Retrieval System (6)	-	Tennessee Department of Conservation	Groundwater yield and quality	Centralized	75,000 wells 800 springs	IBM 370-OS
Well Hydrograph Data Storage and Retrieval System (7)	DSWELL	ERDA Hanford	Well hydrograph	Centralized	300 wells	PDP-9
Groundwater Observation Well Network (8)	GOWN	Canada	Well logs, well data, hydrographs	Centralized	75,000 wells	IBM 360/165
Arizona Water Information System (9)	AWIS	Arizona Water Commission	Water resources	Centralized	2,500+ wells	DEC-10

1. EPA, 1971.
2. Edwards, 1974.
3. Klein et al., 1968.
4. Welsh, 1973.
5. Guenther et al., 1973.
6. Wilson et al., 1972.
7. Friedrichs, 1972.
8. Gilliland and Treichel, 1968.
9. Foster and DeCook, 1974.

TABLE 4. COMPUTERIZED INFORMATION INDEXING SYSTEMS

System Name	Acronym	Administrator	File Content	Retrieval Options	Subject	File Size	Computer System
Remote Control System	RECON	ERDA	Document citations	Keywords, publishers countries, authors, etc.	Energy/ Environmental	700,000 citations	IBM 360/75
General Information Processing System	GIPSY	University of Oklahoma	Document citations	Author, any word(s) in abstract, title	Selected water resources abstracts*	80,000 citations*	IBM 360/65*
Environmental Data Index**	ENDEX	NOAA	Data file descriptions	Geographic area (sq), institution, discipline	Environmental	3,000 file references	IBM 360/65
Oceanic and Atmospheric Scientific Information System**	OASIS	NOAA (Environmental Data Service, 1974)	Document citations	Title, keyword, author, publication, etc.	Atmospheric, water and earth resources	10,000,000 citations, 33 files	IBM 360/65 plus others
National Water Data Exchange	NAWDEX	USGS	Type and sources of water data	Station code, WRC Basin code, Lat/Long.	Surface and ground water	Developmental	IBM 370/155
World Science Information System	UNISIST	UNESCO	Type and source of Global Research Documentation	Developmental	Scientific	Developmental	Developmental
International Referral System	IRS	U.N. Environmental Program, Nairobi	Type and source of Global Research Documentation	Developmental	Environmental	Developmental	Developmental
Water Resources Information Program		University of Wisconsin - Madison	Document citations	Free form questions	Water Resources	70,000 citations+	IBM 360/75
Smithsonian Science Information Exchange	SSIE	Smithsonian Science Information Exchange, Inc.	Research in progress	Free form queries	Scientific	170,000 research projects	IBM 370/135

* Department of the Interior, Water Resources Scientific Information Center information base.
** GIPSY is also used to access some modules of the ENDEX and OASIS data bases.

TABLE 5. GENERALIZED DATA BASE MANAGEMENT PACKAGES.

System Name	Vendor	Purchase Price	Minimum Core Req'd	Compatability	Applicability
DYL-250/260	Dylakov Computer Systems, Inc.	$8 K+***	32 K	IBM 360/370	Index sequential files, report writing
IMS	IBM	$1316/mo.***	128 K	IBM 360/370	Extremely flexible but complex
MARK IV	Informatics, Inc.	$7.5-35 K**	20 K	IBM 360/370, Univac 70/90/ 9400	Infrequent, large retrievals from large data base
SYSTEM 2000	MRI System Corp.	$1 -4 K/mo.*	128 K	IBM 360/370, CDC 6000, Univac 1106, 1108, 1110	On-line direct access, inverted files
TOTAL	Cincon Systems, Inc.	$26,500+***	31 K(avg.)	IBM 360/370, CDC Cyber, H 200/2000, Univac 70	Complex interrelation-ships between data files
ADABAS	Software AG	$120 K***	30 K	IBM 360/370, Univac 9000	Extremely large data bases, many files
PANVALET	Pansophic Systems, Inc.	$5 K+***	50 K(avg.)	IBM 360/370	Library maintenance
RAMIS	Mathematica, Inc.	$28 K***	120 K	IBM 360	Hierarchial structured data bases
RSVP	Honeywell Infor-mation Systems, Inc.	$ 4 K+***	22 K	IBM 360/370	User oriented

* ([illegible]elke, 1972)
** (Steig, 1972)
*** (Datapro Research Corporation, 1974)

TABLE 6. STORET SUPPORTED SECTIONS OF PL 92-500 (after Conger, 1975).

Title I - Research and Related Programs

[1,3]Sec. 104 - Research, Investigations, Training and Information

[1,3]Sec. 104(a)(5) - National Water Quality Surveillance System (NWQSS)

[2,3]Sec. 105 - Grants for Research and Development

[3]Sec. 106 - Grants for Pollution Control Programs

[3]Sec. 107 - Mine Water Pollution Control Demonstrations

Sec. 108 - Pollution Control in the Great Lakes

[3]Sec. 113 - Alaska Village Demonstration Projects

Sec. 114 - Lake Tahoe Study

Title II - Grants for Construction of Treatment Works

[3]Sec. 201 - Construction Grant Facility Plan

[3]Sec. 208 - Areawide Waste Treatment Management Plan

[3]Sec. 209 - Basin Planning

[2]Sec. 210 - Annual Operation and Maintenance Survey

Title III - Standards and Enforcement

[3]Sec. 303 - Water Quality Standards and Implementation Plans

Sec. 303(e) - River Basin Water Quality Management Plans

Sec. 305(b) - Water Quality Inventory

[3]Sec. 308 - Inspections, Monitoring and Entry

[3]Sec. 311 - Oil and Hazardous Substance Liability

Sec. 314 - Clean Lakes

[3]Sec. 315 - National Commission on Water Quality

Sec. 316 - Thermal Discharges

Sec. 318 - Aquaculture

Table 6 - Continued

Title IV - Permits and Licenses

[3]Sec. 402 - National Pollutant Discharge Elimination System

Sec. 403 - Ocean Discharge Criteria

Sec. 404 - Permits for Dredged or Fill Material

Title V - General Provisions

[2,3]Sec. 516 - Reports to Congress

[3]Sec. 516(b) - Economics of Clean Environmental Report

---

[1]Requires Federal information management support.

[2]Requires dissemination of information.

[3]Groundwater implications.

The WQF measures the ambient quality of water bodies throughout the nation and the GPSF measured the quality of point source discharges throughout the nation. The software which updates, manipulates, and retrieves data from these files is coded in the PL/1 programming language. Updates and retrievals are done in the batch mode with input provided by card readers or low to medium speed remote terminals. Output reports are generated on a demand basis only.

The WQF contains information which can be segregated into three categories. The first of these categories consists of information which describes the source of water quality samples (i.e., water quality monitoring stations). This descriptive information is required only when the stations are established, in or deleted from the STORET system data base or when the descriptive information is changed. The input data content and format for station descriptions is presented in Figure 2. Header cards 1, 2, 3, 4 and 5 are optional inputs. A detailed description of the procedure for using all of these station storage cards can be found in available STORET documentation. Only a brief description of the mandatory agency and station cards is provided here.

The agency header card contains general information pertaining to a station or group of stations involved in a single station storage or retrieval operation. The agency header card is used in the following manner:

- The agency identifier which associates data with the contributing organization must be provided in columns 1 through 8.

- An "unlocking key" is an alphanumeric code which is input via columns 17 through 24 of the agency card and which is mandatory for all station storage and, if requested by the data contributor, for all retrieval operations.

- Columns 25 through 61 are provided to accommodate the name, location and telephone number of the individual responsible for storing the station description. Information in this field, though required as input, is not stored as a part of the STORET data base.

  Column 62 is used to record the units in which the sample depths are to be reported and allows the entering of either an F (feet) or an M (meters).

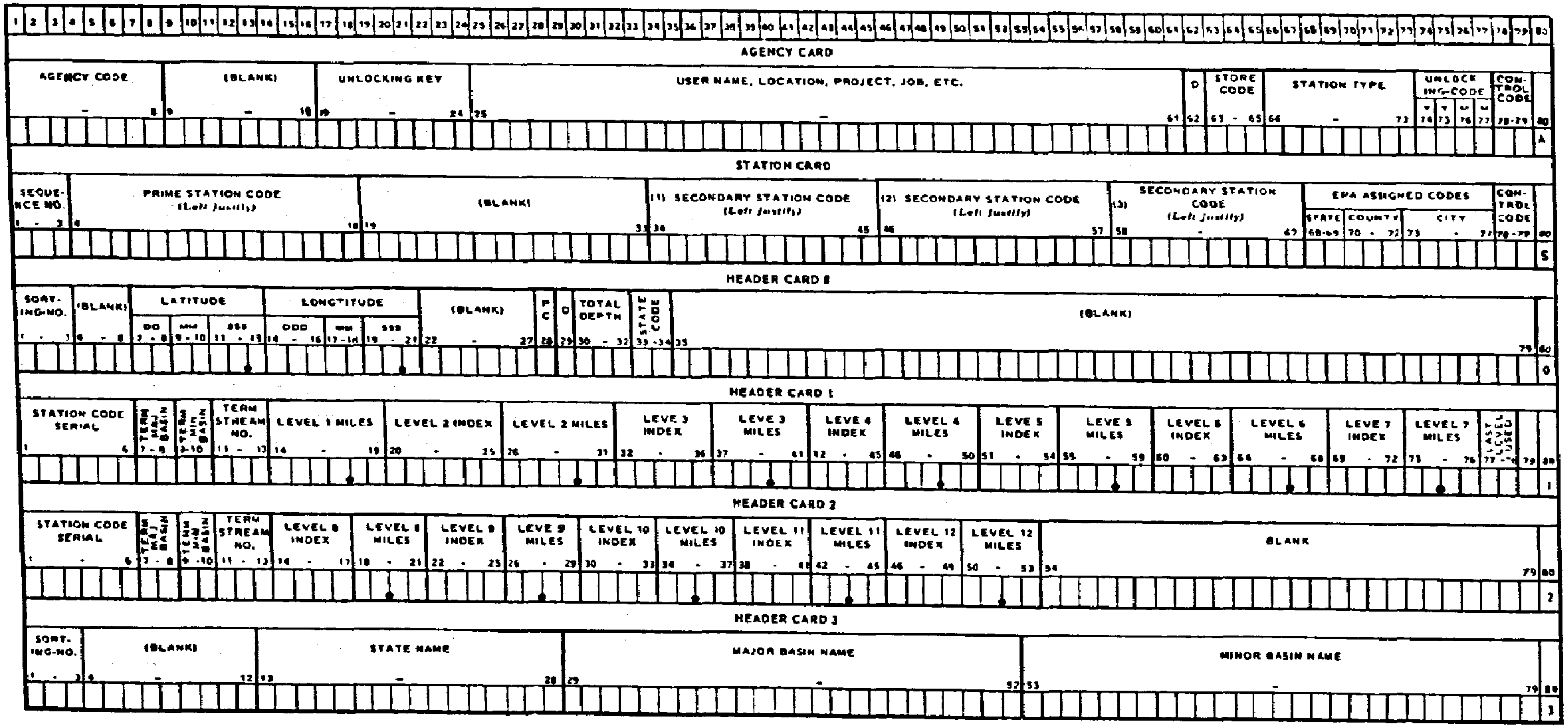

Figure 2. STORET system-station storage format.

HEADER CARD 4

SORTING NO. (1 - 3)	(BLANK) (4 - 6)	MINOR, BASIN NAME (Continued from card 3) (7 - 19)	LOCATION NAME (20 - 51)	(BLANK) (52 - 79)	80
					4

HEADER CARD 5

STATION CODE (1 - 6)	DESCRIPTION (7-78)	79	80
		0	5
		1	5
		2	5
		3	5
		4	5
		5	5
		6	5
		7	5
		8	5
		9	5
		J	5

Figure 2 - Continued

Columns 63 through 65 may be used (optionally) to stipulate (by inputting a 1 in the appropriate column) that it is desired to store latitude-longitude, RMI code, and/or the state-county-city code as a secondary station number(s).

Columns 66 through 73 must be used to provide a station type code. Station type codes are constructed as shown in Figure 3.

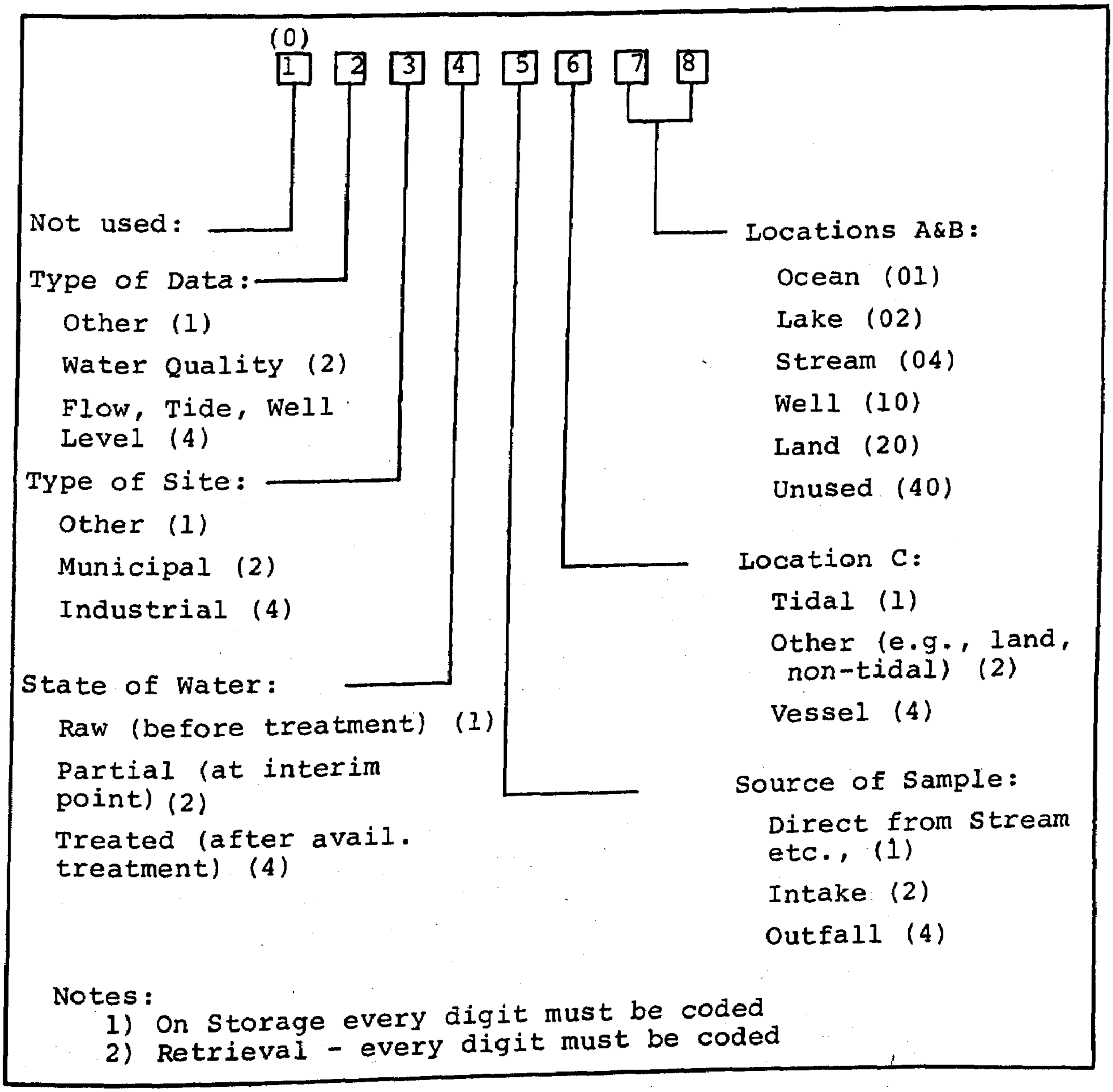

Figure 3. STORET system-station type codes.

- Columns 74 through 77 are used to stipulate the date <u>after</u> which data cannot be retrieved without providing an unlocking key.

- Columns 78 and 79 are used as a card use control and are coded with a CD to change the unlocking date, a CT to change the station type, or left blank for other types of operations.

- For the agency card an A is required in column 80.

The station card is also mandatory and provides a vehicle for inputting information specific to individual stations. In addition to its use in the establishment of a station in the STORET data base, it is also used to delete a station, to change a station location, or to update water quality data. The station card is completed as follows:

- The first field of the card, columns 1 through 3, will contain a sequence number which corresponds to the entries in the same field of all other location and water quality cards for the same station. This field is not stored but rather used for resorting in the event the card deck becomes disarranged.

- The second field, columns 4 through 18, is used to enter the primary station code (alphanumeric) into storage. In general, only the first 6 characters of this field are used.

- The three fields consisting of columns 34 through 45, 46 through 57, and 58 through 67 are used to store secondary station codes if required. Secondary station codes are used in the event, for example, that several organizations are storing water quality information derived from the same station but have assigned different codes to that station.

- The next three fields (columns 68 through 69, 70 through 72, and 73 through 77) are numeric and are used to store state, county, and city codes respectively. State and county codes are those adopted by the National Bureau of Standards. City codes are based upon codes adopted by the U.S. Postal Service.

- Columns 78 and 79 are for card use control. This field is coded with an "NS" if an original storage is being executed, with "DD" to delete all data associated with a station, with "DS" to delete both the station and all data associated with the station, with "CN" to change secondary station numbers, with "CC" to change or delete station descriptive data, and with blanks for water quality data updates.

- Column 80 of a station card is coded with an S.

Sampling stations are located areally by stipulating either geographical coordinates (header card 0), hydrologic index (header cards 1 and 2) or both. Locating sampling stations by geographical coordinates allows the retrieval of data from all stations located within a polygon simply by specifying the vertices of that polygon. Hydrologic indexing, referred to as River Mile Index (RMI) coding, offers an extremely powerful tool since it defines the hydrologic relationship between a sampling point and the rest of the river system. A complete RMI requires between 15 and 112 numeric characters and is composed of the following codes:

- Major basin code (2 characters)
- Minor basin code (2 characters)
- Terminal stream number (3 characters)
- Indexes defining direction and level of flow
- Mileages between confluences
- Stream level code (2 characters)

As of November 1975, use being made of the various station locating schemes is presented below (Conger, 1974).

	Stations
Total	197,000
RMI	28,000
Geographic	160,000
Both (RMI and Geo.)	15,000
Neither	24,000
Political	193,000

Although RMI coding represents a useful tool, relatively little use is being made of it undoubtedly because of the level of effort required to generate the code. Most of the stations which have been stored in the WQF using both RMI and geographic coordinates are located in only two areas, the Tennessee and Columbia River basins.

The second category of information stored in the WQF data base is water quality parameter identification. Each water quality measurement which is stored in the file must be accompanied by a numeric 5 character parameter identifier code. The 5-character water quality parameter identifier code is stored in a 3-byte field in packed decimal format which allows the storage of 2 numeric characters per byte. The parameter identifier codes are also stored in a cross reference (dictionary) file together with the alphanumeric descriptors which the codes represent.

The WQF can store up to 100,000 parameter identifiers but only about 2,000 identifiers are currently stored. Eighty-five percent of the water quality data in the WQF is stored under only 187 of the existing identifiers, however. An effort has been made to commit specific ranges of parameter codes to sets of parameters with similar characteristics. For example, the range of codes 00300-00365 has been dedicated to measurements of oxygen demand.

Of particular interest is the fact that the range of codes from 84,000 to 84,999 has been set aside for identifiers pertinent to groundwater monitoring. To date, the code 84,000 has been designated as a geologic age code and 84,001 as an aquifer name code. The remainder of the range is uncommitted. Additional parameter codes which have been established specifically to accommodate groundwater monitoring are presented in Table 7.

The third category of information in the WQF is the water quality measurements themselves together with the depth of the sample and the date and time the sample was taken. The water quality measurements are stored in 4-byte words in standard IBM 370 floating point format (single precision).

Originally the input module of the STORET system was designed to store only numeric data in the water quality measurement field. System modifications have been accomplished, however, that allow the storage of alphabetic characters, required for aquifer descriptions, in the fields associated with parameter codes 84,000 through 84,999.

The STORET Water Quality File also allows remarks to be input along with water quality measurements. The system accepts remark codes into a 1-character (1-byte) field, one of which has been set aside for each water quality measurement field. The remarks are stored in "Extended Binary Coded Decimal Interchange Code" (EBCDIC) which allows any one of 256 alternative remark codes. These remarks are used, for example, to indicate that the stored data element is not accurate, a field measurement, a lab measurement, a lower limit, or an upper limit.

TABLE 7. ESTABLISHED STORET PARAMETER CODES - GROUNDWATER SPECIFIC (EPA, 1971). Celsius (BM*)

Code	Output Format**	Parameter Description
72000	xxxxxx.x	Elevation of land surface datum (ft.*** above MSL)
72001	xxxxxx.x	Total depth of hole (ft. below land surface datum)
72002	xxxxxx.x	Depth to top of water-bearing zone sampled (ft.)
72003	xxxxxx.x	Depth to bottom of water-bearing zone sampled (ft.)
72004	xxxxxx.x	Pump or flow period prior to sampling (minutes)
72005	xxxxxxxx	Sample source code (BM* well data)
72006	xxxxxxxx	Sampling condition code (BM* well data)
72007	xxxxxxxx	Formation name code (BM* well data) (AAPG**** code)
72008	xxxxxx.x	Total depth of well (ft. below land surface datum)
72009	xxxxxx.x	Elevation of land surface in feet (BM*)
72010	xxxx.xxx	Resistivity (ohm-meters) (BM* well data)
72011	xxxx.xxx	Acids, organic (Mg/l) (BM* well data)
72012	xxxxx.xx	Specific gravity, temperature, degrees Celsius (BM*)
72013	xxxx.xxx	Specific gravity (BM* well data)
72014	xxxxx.xx	Resistivity, temperature, degrees Celsius (BM*)
72015	xxxxxx.x	Depth to top of sample interval (ft. below LSD)
72016	xxxxxx.x	Depth to bottom of sample interval (ft. below LSD)
72017	xxxxxxxx	Series code (BM* well data)
72018	xxxxxxxx	System code (BM* well data)
72019	xxxxx.xx	Depth to water level (feet below land surface)
72020	xxxxx.xx	Elevation in feet above MSL
72040	xxxxx.xx	Observed drawdown (ft.)
72041	xxxxx.xx	Specific capacity in gpm/ft. of drawdown
72042	xxxxxx.x	Pump efficiency (percent)
72043	xxxxxx.x	Brake horsepower
72044	xxxxxx.x	Total dynamic pumping head (ft.)
72045	xxxxx.xx	Pumping cost in dollars per thousand gallons

TABLE 7 - Continued

Code	Output Format**	Parameter Description
72050	xxxxxx.x	Withdrawal of groundwater (millions of gallons/month)
72051	xxxxxx.x	Withdrawal of groundwater (millions of gallons/year)
84000	xxxxxxxx	Geologic age code (USGS)
84001	xxxxxxxx	Aquifer name code (USGS)

*BM - Bureau of Mines
**Can be modified at retrieval
***See Appendix for conversion to metric units
****American Association of Petroleum Geologists

Recent cost and use data for the WQF are presented below (Notzon, 1975):

Annual operating costs excluding EPA personnel	$1,100,000
Federal, state and local users	240
Cost per user per year	$3,667
Observations stored annually	8-10 million
Observations presently in system	30,000,000
Data acquisition cost	$150-300 million
Annual storage cost per observation	$.01
Processing cost per observation	$.011
Retrievals/analysis per year	46,000
Retrieval/analysis cost per job (avg.)	$7.58

The General Point Source File consisted of an inventory of dischargers and abatement plans. More specifically, the GPSF contained the following information:

1. Inventory of municipal dischargers
2. Inventory of industrial dischargers
3. Inventory of municipalities
4. Fish kills
5. Agricultural permits
6. Mine drainage permits
7. Deep well injection survey
8. Municipal drinking water supplies
9. Construction needs survey
10. Ocean dumping permits
11. Federal government discharges
12. Grant information

WATSTORE

The National Water Data Storage and Retrieval System (WATSTORE) was implemented in 1971 with the objective of providing the Water Resources Division of the USGS with a comprehensive water data management capability. The system is computerized and operated at the facilities of USGS in Reston, Virginia. Access to WATSTORE is through a telecommunication network which provides data services to 46 district offices throughout the country. Data are input to WATSTORE by remote entry from laboratories and data centers.

The system data base consists of a "Station Header File" which maintains an index of stations and provides access to the following files:

- The "Daily Values File" contains physical and chemical data reported daily.
- The "Water Quality File" contains the results of analysis (chemical and physical) of all samples taken. This includes groundwater samples generally taken on an infrequent and irregular basis.
- The "Peak Flow File" contains annual maximum discharge and stage values for surface water sites.
- The "Groundwater Site Inventory File" contains physical, topographic, aquifer hydraulic and text data pertinent to groundwater monitoring sites. Parameters maintained in this file are presented in Table 8.

WATSTORE retrieval capabilities enable the output of text, tabular, and graphic reports. Retrieval options include individual station, station type (e.g., wells), specific periods, polygon, political, aquifer code (for groundwater sites), and individual parameter retrievals. In addition, data for a particular parameter which falls within a specified range may be retrieved.

The WATSTORE system is designed to recognize the possibility that a groundwater monitoring station (well) can penetrate more than one aquifer and that samples can be drawn from individual aquifers separately with the use of screen plugs. Therefore, WATSTORE allows for the storage of aquifer identifiers along with the water quality analysis data for each sample.

TABLE 8. PARAMETERS MAINTAINED IN WATSTORE GROUNDWATER SITE INVENTORY FILE (Baker, 1975)

Site Id
Site Type
Record Classification
Source Agency
Project Number
District
State
County
State County*
Latitude
Longitude
Coordinate Accuracy
Local Number
Land Net Location
Location Map Id
Location Map Scale
Altitude
Altitude Method
Altitude Accuracy
Topographic Setting
OWDC Hydrologic Unit
Date Constructed
Date Const. Acc.*
Site Use
Water Use
Second Water Use
Third Water Use
Hole Depth
Well Depth
Well Depth Source
Water Level
Water Level Date
WL Date Accuracy*
Water Level Source
Water Level Method
Site Level Status
Pump
Geohydro Data Source
Last Update*
Verified*

Lift
- Lift Type
- Lift Date
- Lift Date Accuracy*
- Intake Setting
- Power Type
- Horsepower
- Major Pump
  - Manufacturer
  - Serial Number
  - Power Company
  - Account
  - Meter
  - Consumption
  - Pump Maintainer
- Standby
  - Standby Power Type
  - Standby Horsepower

Geohydrologic Units
- Geohydro Top
- Geohydro Bottom
- Geohydro Unit
- Lithology
- Lithology Modifier
- Aquifer
  - Aquifer Date
  - Aquifer Date Acc.
  - Aquifer Static Level
  - Aquifer Contribution

Remarks
- Remarks Date
- Remark

Construction
- Const. Sequence No.
- Date Completed
- Const. Date Accuracy*
- Contractor
- Const. Data Source
- Const. Method
- Finish
- Seal Type
- Seal Bottom
- Development Method
- Development Duration
- Special Treatment
- Holes
  - Hole Top
  - Hole Bottom
  - Hole Diameter
- Casings
  - Casing Top
  - Casing Bottom
  - Casing Diameter
  - Casing Material
  - Casing Thickness
- Openings
  - Opening Top
  - Opening Bottom
  - Opening Type
  - Screen Material
  - Opening Diameter
  - Opening Width
  - Opening Length

Site Visits
- Inventory Date
- Inventory Person

Table 8 - Continued

Hydraulic Data
- Hydraulic Seq. No.
- Hydraulic Unit Id.
- Test Interval Top
- Test Interval Bottom
- Hydraulic Unit Type
- Hydraulic Remarks
- Coefficients
  - Coef. Seq. No.
  - Transmissivity
  - Horizontal Cond.
  - Vertical Cond.
  - Storage Coef.
  - Leakage
  - Diffusivity
  - Specific Storage

Quality Network(QN)
- QN Begin Year
- QN End Year
- QN Data Source
- QN Frequency
- QN Analysis Type

Level Network (LN)
- LN Begin Year
- LN End Year
- LN Data Source
- LN Frequency

Pumpage Network (PN)
- PN Begin Year
- PN End Year
- PN Data Source
- PN Frequency
- PN Data Method

Flow Data
- Flow Seq. No.
- Flow Meas. Date
- Flow Date Acc.*
- Flow Discharge
- Flow Discharge Source
- Flow Discharge Method
- Flow Prod. Level
- Flow Static Level
- Flow Level Source
- Flow Level Method
- Flow Period

Pump Production
- Pump Seq. No.
- Pump Meas. Date
- Pump Date Acc.*
- Pump Discharge
- Pump Discharge Source
- Pump Discharge Method
- Pump Prod. Level
- Pump Static Level
- Pump Level Source
- Pump Level Method
- Pump Period

Owners
- Ownership Date
- Ownership Date Acc.*
- Last Name
- First Name
- Middle Initial

Minor Repairs
- Repair Seq. No.
- Repair Nature
- Repair Date
- Repair Date Acc.*
- Repair Contractor
- Performance Changes

Springs
- Spring Name
- Spring Type
- Permanence
- Discharge Sphere
- Improvements
- Number Spring Openings
- Flow Variability
- Flow Var. Accuracy

Other Data
- Type Data
- Data Location

Other Ids
- Other Id
- Other Id Assigner

Field Water Quality
- FWQ Sample Date
- FWQ Date Acc.*
- FWQ Geohydro Unit
- FWQ Parameter
- FWQ Measurement

Logs
- Log Type
- Log Top
- Log Bottom
- Log Source

Well Group(WG)
- Number Wells
- WG Deepest
- WG Shallowest
- WG Method

Pond Tunnel Drain
- PTD Length
- PTD Width
- PTD Depth

Cooperator Data
- Cooperators Id
- Contractor Reg. No.
- Inspection Status
- Reason Unapproved
- Date Inspected
- Cooperator Remarks

Laterals
- Lateral Number
- Lateral Depth
- Lateral Length
- Lateral Diameter
- Lateral Mesh

*System-Generated

The aquifer identifiers are stored as 8-character codes based on the stratigraphic coding system proposed by the American Association of Petroleum Geologists.

The 8-character code consists of three parts. The first 3 characters are numeric and identify the geologic age (Erathem, System and Series, respectively) of the aquifer as shown in Table 9. The next 4 characters constitute an alphanumeric mneumonic code which specifies the name of the rock-stratigraphic unit. The rock-stratigraphic unit name code is generated by the use of an algorithm which specifies the order in which characters are to be eliminated from the original term until only 4 remain. The last character of the 8-character is optional and provides for modifiers of the rock-stratigraphic unit name. For example, the complete code for the Pliocene Upper Pico Formation in California is 121PICOU.

The WATSTORE system currently stores data for several hundred different water quality parameters. The list of water quality parameters is open ended and is expanded as necessary. The water quality parameters stored in WATSTORE are coded with a 5-character code established in cooperation with the EPA STORET User Assistance Branch so that the parameter codes are the same in both systems. WATSTORE is equipped with a module which generates STORET input corresponding to WATSTORE data file updates. The input formats for storing data in the WATSTORE Water Quality File are presented in Figure 4.

NAWDEX

The National Water Data Exchange (NAWDEX) is a developmental computerized information indexing capability being implemented by the Water Resources Division of the U.S. Geological Survey. This effort resulted from a determination by the U.S. Department of the Interior that accessibility to water data on a national scale required upgrading.

NAWDEX will consist of a centralized data inventory file and communications links, not necessarily automated, with management information systems maintained by the various data depositors that subscribe to NAWDEX. The centralized data file will contain monitoring stations descriptions as well as sources and types (parameters and monitoring frequency) of available water data. Access to this file is provided by requiring the user to stipulate his interest in either surface or groundwater and geographical area of interest (e.g., hydrologic basin code). The system allows additional information, as available, from the data requestor to further narrow the file search (Planning Research Corporation, 1974).

TABLE 9. USGS NUMERIC CODES FOR GEOLOGIC AGE IDENTIFICATION (Price and Baker, 1974).

Age	Code
Unknown Age	000
Cenozoic	100
Quaternary	110
Holocene	111
Pleistocene	112
Tertiary	120
Pliocene	121
Miocene	122
Oligocene	123
Eocene	124
Paleocene	125
Mesozoic	200
Cretaceous	210
Upper	211
Gulfian	212
Lower	217
Comanchean	218
Coahuilan	219
Jurassic	220
Upper	221
Middle	224
Lower	227
Triassic	230
Upper	231
Middle	234
Lower	237
Paleozoic	300
Permian	310
Upper	311
Ochoan	312
Guadalupian	313
Lower	317
Leonardian	318
Wolfcampian	319
Pennsylvania	320
Upper	321
Virgilian	322
Missourian	323

Age	Code
Paleozoic (cont'd)	
Middle	324
Des Moinesian	325
Atokan	326
Lower	327
Morrowan	328
Mississippian	330
Upper	331
Chesterian	332
Meramecian	333
Lower	337
Osagean	338
Kinderhookian	339
Devonian	340
Upper	341
Chautauquan	342
Senecan	343
Middle	344
Erian	345
Lower	347
Ulsterian	348
Silurian	350
Upper	351
Cayugan	352
Middle	354
Niagaran	355
Lower	357
Ordovician	360
Upper	361
Cincinnatian	362
Middle	364
Champlainian	365
Lower	367
Canadian	368
Cambrian	370
Upper	371
Middle	374
Lower	377
Precambrian	400

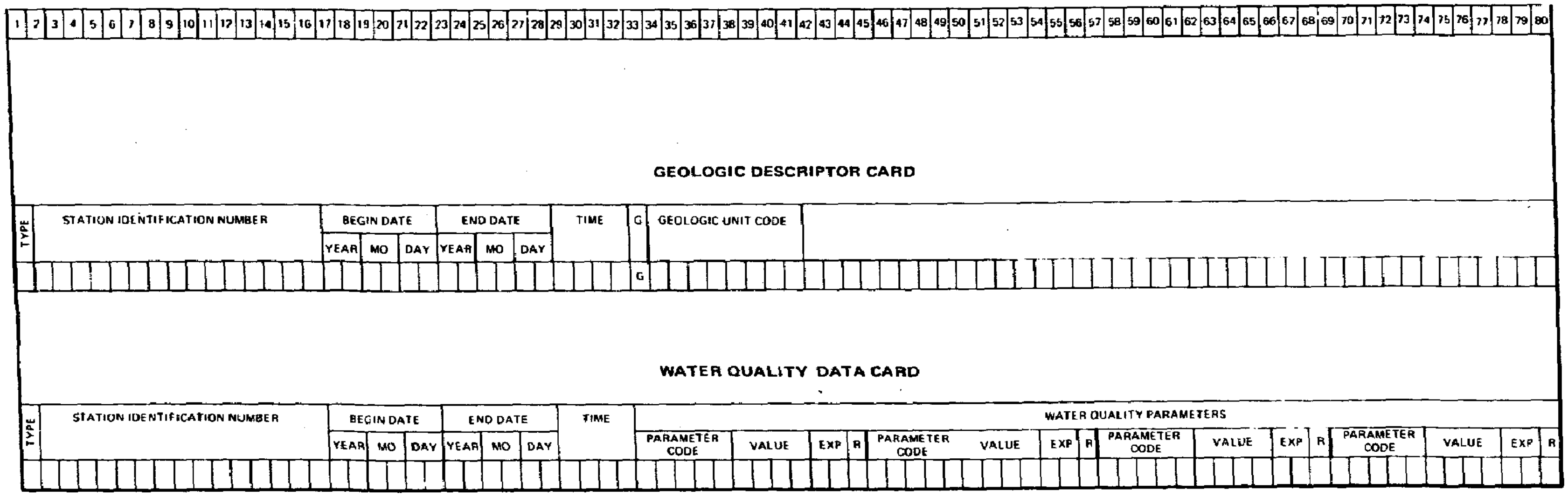

Figure 4. WATSTORE Water Quality File - data storage format.

SECTION VI

PROPOSED MODIFICATIONS TO EXISTING SYSTEMS

1. The STORET parameter code dictionary should be appended to include those groundwater monitoring related parameters listed in Table 10.

2. The STORET system should be modified to accept multiple remark codes with individual measurements. It is recognized that a modification of this type would represent a major commitment of resources.

3. The STORET groundwater data file should be developed separately from the existing STORET surface water data file (i.e, the WQF). This will promote faster updates of the groundwater data file and avoid degradation of update times for the surface water data file.

4. The STORET groundwater data file should be maintained on a detachable magnetic disc and placed on-line on the basis of some constant schedule (e.g., Tuesdays and Fridays from 2:00 p.m. to 6:00 p.m.). The periods during which the file will be on-line can be determined by performing a survey of potential users.

5. Some groundwater data should be archived off-line on magnetic tape. The data set to be archived can be defined either on the basis of its age (e.g., data over two years old) or on the basis of its activity level (e.g., stations not accessed or updated within the preceding year).

6. The proposed STORET groundwater data file should be allowed to accept compliance monitoring data as well as background information monitoring data. Discharge permit numbers may be used as station identifier codes. The fact that a monitoring station is generating compliance data can be indicated in the station type code. In addition, the groundwater data file should be able to accept DMA status data, with the DMA treated as a station and the DMA code used as a station code.

7. For the groundwater data file, the eight character STORET station type code should be modified and interpreted as follows:

TABLE 10. PROPOSED ADDITIONAL STORET PARAMETER CODES.

Code	Parameter Description
84100	Horizontal permeability (gpd/ft^2)
84105	Vertical permeability (gpd/ft^2)
84107	Specific yield (dimensionless)
84110	Effective porosity (percent)
84112	Void ratio
84115	Soil bulk density (grams/liter)
84117	Soil moisture content (percent)
84120	Soil exchangeable sodium (percent)
84123	Soil specific gravity ($grams/cm^3$)
84130	Soil gradation - percent clay or silt fines
84131	Soil gradation - percent fine sand
84132	Soil gradation - percent medium sand
84133	Soil gradation - percent coarse sand
84134	Soil gradation - percent fine gravel
84135	Soil gradation - percent coarse gravel
84136	Soil gradation - percent cobbles
84138	Coefficient of soil uniformity
84140	Coefficient of curvature of soil gradation plot
84142	Capillary head (feet)
84200	Hydraulic gradient
84205	Hydraulic gradient direction (degrees from North)
84210	Transmissivity (gpd/ft)
84215	Storage coefficient (dimensionless)
84220	Leakage - downward (gpd/sq. mi.)
84222	Leakage - upward (gpd/sq. mi.)
84225	Diffusivity (gpd/ft)
84230	Specific flux (gpd/ft^2)
84300	Highest use made of aquifer (protected use)
84500	Monitoring agency status index
84505	Pollution control readiness index
84600-84610	Alphanumeric, sample specific comments - 10 fields, 4 characters each

- Column 1 which is not currently used should be allowed to accept a code to indicate the sample extraction method employed at the subject station (i.e., pump = 1, bail = 2, and probe = 4).

- In column 2 a 1 would indicate DMA status data, a 2 would indicate water quality data, and a 4 would indicate hydrogeologic data.

- In column 5 a 1 would indicate information monitoring, a 2 would indicate compliance monitoring, and a 4 would indicate other.

- In columns 7 and 8 a 10 would indicate monitoring directly in the saturated zone, a 20 would indicate surface monitoring, and a 40 would indicate monitoring of the zone of aeration.

8. The STORET groundwater data file should store water quality criteria (ambient or effluent) as sample data. The date of enactment of the criteria would be stored in the STORET sample date field and some exclusive value such as 8888 for ambient criteria and 9999 for effluent limitation would be stored in the STORET sample time field.

9. STORET retrieval options should be expanded to allow more extensive Boolean retrieval strategies. It is recognized that these additions would require setting up new index and cross-reference files and correspondingly entail a significant additional commitment of resources.

10. STORET user assistance capabilities and policies should be expanded to allow non-machine compatible user interface with the data base on a routine basis.

11. Either the GIPSY or the RECON document citation retrieval systems should be modified to accommodate polygon type retrievals. This would allow the groundwater investigator to provide geographic delimiters and receive research documentation abstracts regarding his geographical area of interest.

SECTION VII

REFERENCES

Baker, C., Written Communication, U.S. Geological Survey, Water Resources Division, Reston, Va., April 3, 1975.

Conger, C. S., Personal Communication, U.S. Environmental Protection Agency, Data Processing and User Assistance Branch, Washington, D.C., November 4, 1974.

Conger, C. S., Written Communication, U.S. Environmental Protection Agency, Data Processing and User Assistance Branch, Washington, D.C., March 4, 1975.

Datapro Research Corporation, A Buyers Guide to Data Base Management Systems, Delran, N.J., 12 pages, December, 1974.

Edwards, Melvin D., The Processing and Storage of Water Quality in the National Water Data Storage and Retrieval System, U.S. Geological Survey, Water Resources Division, Reston, Va., 85 pages, 1974.

Environmental Data Service, User's Guide to OASIS - Oceanic and Atmospheric Scientific Information System, National Oceanic and Atmospheric Administration, Washington, D.C., 1974.

Ferrara, R., and R. L. Nolan, "New Look at Computer Data Entry," Journal of Systems Management, Association for Systems Management, p 24-33, February, 1973.

Foster, K. E., and J. DeCook, Implementation of Arizona Water Information System (AWIS) Remote Terminal Accessible Hydrologic Data Sets on DEC-10 Computer, University of Arizona, Tucson, Arizona, 21 pages, 1974.

Friedrichs, D. R., Information Storage and Retrieval System for Well Hydrograph Data - User's Manual, Battelle Pacific Northwest Laboratories, Richland, Washington, 23 pages, 1972.

Gilliland, J. A., and A. Treichel, "GOWN - A Computer Storage System for Groundwater Data," Canadian Journal of Earth Sciences, Vol. 5, p 1518-1524, September 1968.

Guenther, G., D. Mincavage, and F. Morley, Michigan Water Resources Enforcement and Information System, U.S. Environmental Protection Agency, Office of Research and Monitoring, Socioeconomic Environmental Studies Series, EPA-R5-73-020, Washington, D.C., 161 pages, 1973.

Hem, J. D., Study and Interpretation of the Chemical Characteristics of Natural Water, U.S. Geological Survey, Water Supply Paper 1473, 269 pages, 1959.

House, W. C., ed., Data Base Management, Mason and Lipscomb Publishers, Inc., New York, New York, 470 pages, 1974.

Klein, W. L., D. A. Dunsmore, and R. K. Horton, "An Integrated Monitoring System for Water Quality Management in the Ohio Valley," Environmental Science and Technology, Vol. 2, American Chemical Society, p 764-771, October, 1968.

Lobel, Jerome, and M. V. Farina, "Selecting Computer Memory Devices," Automation, Penton Publishing Co., Cleveland, Ohio, p 66-70, October 1970.

Lorber, Matthew, "Evaluating Computer Output Printers," Automation, Penton Publishing Co., Cleveland, Ohio, p 64-67, March 1972.

Notzon, E.M., Written Communication, U.S. Environmental Protection Agency, Monitoring and Data Support Division, Washington, D.C., October 2, 1975.

Planning Research Corporation, Support in the Implementation of a National Water Data Exchange, Second Quarterly Progress Report (September-November, 1974), PRC-D-1863, 61 pages, December 1974.

Price, W. E., and C. H. Baker, Catalog of Aquifer Names and Geologic Unit Codes Used by the Water Resources Division, U.S. Department of the Interior, Geological Survey, Water Resources Division, Reston, Va., 306 pages, 1974.

Schwab, B., and R. Sitter, "Economic Aspects of Computer Input-Output Equipment," Financial Executive, Financial Executives Institute, p 75-87, September 1969.

Showen, C. R., and O. O. Williams, Index to Water Quality Data Available from the U.S. Geological Survey in Machine-Readable Form to December 31, 1972 - Western Region, PB-232-794, U.S. Geological Survey, Water Resources Division, Washington, D.C., 520 pages, 1973.

Steig, D. B., "File Management Systems Revisited," Datamation, Barrington, Illinois, p 48-51, October 1972.

Taylor, P. L., Written Communication, U.S. Environmental Protection Agency, Data Reporting Branch, Washington, D.C., November 27, 1974.

U.S. Department of Health, Education and Welfare, The Toxic Substances List, 1973 Edition, National Institute for Occupational Safety and Health, Rockville, Maryland, 1001 pages, 1973.

U.S. Environmental Protection Agency, Storage and Retrieval of Water Quality Data. Training Manual, PB-214 580, Washington, D.C., 302 pages, 1971.

U.S. Environmental Protection Agency, Proposed Criteria for Water Quality, Vol. 1, Washington, D.C., 425 pages, 1973.

U.S. Public Health Service, "Drinking Water Standards," Federal Register, Government Printing Office, Washington, D.C., p 2152-2155, March 6, 1962.

Wilson, J.M., M. J. Mallory, and J. M. Kernodle, Summary of Groundwater Data for Tennessee through May 1972, Miscellaneous Publication Number 9, State of Tennessee, Department of Conservation, Division of Water Resources, Nashville, Tennessee, 1972.

Welke, Larry, "A Review of File Maintenance Systems," Datamation, Barrington, Illinois, p 52-54, October 1972.

Welsh, J. L., "Ground-Water Quality Data for Planning, Monitoring, and Surveillance," Proceedings at the Ninth Biennial Conference on Ground Water, Goleta, California, September 1973.

APPENDIX

METRIC CONVERSION TABLE

Non-Metric Unit	Multiply by	Metric Unit
feet (ft)	0.3048	meters (m)
gallons (gal)	3.785	liters (l)
miles (mi)	1.609	kilometers (km)
gallons per day (gpd)	3.785412	liters/day (l/day)
gallons per minute (gpm)	3.785412	liters/minute (l/min)

LIST OF ABBREVIATIONS AND ACRONYMS

AWIS	Arizona Water Information System
BM	Bureau of Mines
CPU	Central processing unit
CRT	Cathode ray tube
DMA	Designated monitoring agency
DNR	Department of Natural Resources
DSWELL	Well Hydrograph Data Storage and Retrieval System
EBCDIC	Extended Binary Coded Decimal Interchange Code
EMSL	Environmental Monitoring and Support Laboratory
ENDEX	Environmental Data Index
EPA	Environmental Protection Agency
ERDA	Energy Research Development Agency
GIPSY	General Information Processing System
GOWN	Groundwater Observation Well Network
GPSF	General Point Source File
HEW	Department of Health, Education, and Welfare
IRS	International Referral System
MIS	Management information system
NAS	National Academy of Sciences
NAWDEX	National Water Data Exchange
NPDES	National Pollutant Discharge Elimination System
NOAA	National Oceanic and Atmospheric Administration
NWQSS	National Water Quality Surveillance System
OASIS	Oceanic and Atmospheric Scientific Information System
OCR	Optical character recognition

ORSANCO	Ohio River Valley Water, Sanitation Commission
OWDC	Office of Water Data Coordination
RECON	Remote Control System
RMI	River Mile Index
SSIE	Smithsonian Science Information Exchange
STORET	Storage and Retrieval system
TELEX	Telephone Exchange
UNESCO	U.N. Educational, Scientific and Cultural Organization
UNISIST	World Science Information System
USDI	U.S. Department of the Interior
USGS	U.S. Geological Survey
USPHS	U.S. Public Health Service
WATS	Wide Area Telephone Service
WATSTORE	National Water Data Storage and Retrieval System
WISE	Water Information System for Enforcement
WQF	Water Quality File

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/4-76-019	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE MONITORING GROUNDWATER QUALITY: DATA MANAGEMENT		5. REPORT DATE April 1976
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Norman F. Hampton		8. PERFORMING ORGANIZATION REPORT NO. GE75TMP-70
9. PERFORMING ORGANIZATION NAME AND ADDRESS General Electric Company—TEMPO Center for Advanced Studies 816 State Street Santa Barbara, California 93101		10. PROGRAM ELEMENT NO. 1HA326
		11. CONTRACT/GRANT NO. 68-01-0759
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Research and Development Environmental Monitoring and Support Laboratory Las Vegas, Nevada 89114		13. TYPE OF REPORT AND PERIOD COVERED
		14. SPONSORING AGENCY CODE EPA-ORD-Office of Monitoring and Technical Support

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The growing concern for subsurface water resources will surely be accompanied by an expanding groundwater data base, a data base which is already quite large. This report is intended to point the way towards the efficient management of this data base which will assure that pertinent information is available when and where it is needed. The discussion presented here will describe the requirements of groundwater data management, survey some available capabilities which may serve to satisfy these requirements and identify the means by which these capabilities can be used to accomplish the management of groundwater data.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Groundwater Data Management, Groundwater Quality Data, Water Quality Data, Monitoring Groundwater, Groundwater, Aquifers, Aquifer Characteristics	Groundwater Data Management, Water Quality Data	08H 09B 13B

18. DISTRIBUTION STATEMENT Release to public	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 70
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE