J. ENVIRONMENTAL SYSTEMS, Vol. 8(4), 1978-79

MANAGEMENT OF GROUND-WATER QUALITY DATA

NORMAN F. HAMPTON
Aerospace Corporation
Germantown, Maryland

LORNE G. EVERETT
Manager, Water Resources Program
General Electric-TEMPO
Center for Advanced Studies
Santa Barbara, California

ABSTRACT

The growing concern for subsurface water resources will surely be accompanied by an expanding ground-water data base, a data base which is already quite large. This paper 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 ground-water data management and survey some of the available capabilities which may serve to satisfy these requirements. Additionally, recommended modifications to existing capabilities will be presented which will make those capabilities more responsive to ground-water quality data management needs.

INTRODUCTION

The intent of this paper is to identify the system requirements of a comprehensive ground-water quality management information system (MIS), to survey existing capabilities which may satisfy those requirements, and to recommend modifications to those existing capabilities which will make them more responsive to requirements. The inventory of existing data management systems

doi: 10.2190/JK1D-MH4K-C6X7-HGF9
http://baywood.com

which is described is not intended to be comprehensive. Rather, existing systems were selected for inclusion on the basis of their significance and relevance.

This discussion is concerned with the ground-water 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, a comprehensive computer system capability intended to satisfy these requirements is outlined. The system described will afford management of ambient ground-water quality information, percolate quality information, compliance monitoring information, and other data relevant to the management of ground-water quality, including citations of ground-water research documentation.

INFORMATION MANAGEMENT REQUIREMENTS

A complete MIS requirements analysis would call for an intensive survey of the potential users of the system to enable the development of system specifications. Instead, this paper employs the following assumptions:

- Information to be managed includes 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 ground-water research documentation;
- The ground-water surveillance data base is moderately large (expanding monotonically), consisting of millions of data elements requiring extensive storage capabilities. Once the initial data base is established, input data volume is relatively low and output volume, in response to user queries, somewhat greater;
- Frequency of interaction (updates and queries) with the data base is moderate;
- Updating and interrogating the ground-water data base does not require quick system response; several days turnaround is generally adequate. Interrogating information indexing files (water quality data file descriptions and document citations) does require quick system response, however, to allow for browsing;
- Source information is generated at locations distributed throughout the U.S. with concentrations in areas of high population density. In general, source information is

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 is alphanumeric text, tables of primary data and computed statistics, and pictorial presentations. Reports are generally 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, an effective ground-water monitoring MIS must 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 are discussed in the following paragraphs.

Station Descriptions

Station descriptive data consist 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, 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 with other special station-specific information which is not required for retrievals or computations.

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

- First character—sample extraction method: pump, bail or probe;
- Second character—type of data: quality, hydrogeologic, designated monitoring agency (DMA), and/or status data;

- Third character—type of site: municipal, industrial, or other;
- Fourth character—location: saturated zone, zone of aeration, or surface;
- Fifth character—monitoring justification: information, compliance, or other.

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

The designated monitoring agency responsible for monitoring the station should be stored as an "agency" code in a searchable field so that all stations being maintained by a particular DMA can be retrieved. In addition, the narrative text associated with a station could contain, for example, the names of specific individuals having responsibility for the station together with their telephone numbers. Each station should have an identification code that is maintained permanently within the MIS to provide access to station data even if the station becomes inactive. Station descriptive data to be maintained by the MIS must include information regarding political jurisdiction as well as a unique areal location that specifies the township, range, section, etc., or employs the conventional geographic coordinate system. Additionally, the depths of both the monitoring station hole and intake screen should be stored as station-specific information.

Quality Criteria

Information pertaining to established quality criteria which a ground-water 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 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 which reflects the significance of ground-water pollution in the environs of a monitoring station. This information, typically generated by local planning agencies, need not be used for retrieval or computational operations and consequently can be stored satisfactorily in the narrative text associated with each monitoring station.

The development of a comprehensive ground-water quality monitoring program entails systematic identification and inventorying 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 "Catalog of Aquifer Names and Geologic Unit Codes Used by the Water Resources Division." [1] Aquifer protected use designations should be codified and searchable. Protected use categories should include public water supply and agricultural and industrial use with allowance made for the possibility of subcategories of the latter two.

Permit data, other than discharge limitations imposed, should not be required for retrieval or computational operations and can therefore be stored in the narrative text associated with compliance monitoring stations. If permit numbers are required for search operations, they can be used as secondary station identifiers.

Permit-specified discharge limitations and 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 in "Quality Criteria for Water" with provision made for updating them as they are modified [2]. Although not likely to 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 addition to geographic coordinates, some geologic data are required to specify the aquifer from which the sample originated. Where a monitoring station taps more than one aquifer, aquifer identification is particularly essential and must be provided as sample-specific rather than station-specific data. The requirement for providing aquifer identification can be satisfied by storing the established aquifer name or the geologic formation name and age associated with the monitored aquifer. The latter form of identification is not preferred, however, since the 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 [1].

Additionally, information regarding the physical properties and chemical constituency of the water-bearing materials (aquifer, unsaturated zone, or topsoil) may be necessary, particularly if synergistic effects between these materials and pollutants are to be modeled. If a model is to be computer-accessible by the ground-water 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.

Hydrologic Data

A ground-water quality MIS must be capable of accommodating a wide range of hydrologic information. In general, hydrologic information will have previously been determined, particularly in areas of rigorous ground-water development. Hydrologic information is necessary to the monitoring program to predict the movement of pollutants, isolate sources of pollution, and interpret the relationship between ground water and surface waters.

Most hydrologic information will be station-specific and can be stored concurrently with the establishment of station descriptions in the data base. 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. Major hydrological data elements include water-bearing material depth, thickness, and area extent; permeability; aquifer transmissivity and storage coefficient; hydraulic gradient; sample-specific water table elevation; area and magnitude of natural and artificial recharge and discharge; and station sampling device operating characteristics. Hydrologic measurements required for computations such as determination of hydraulic diffusivity or specific flux must also be stored as searchable information.

Water Quality Parameter Identifiers

The selection of the water quality parameters to be maintained in a ground-water 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 ground-water monitoring, the data descriptions must be data inputs to the system. This allows the list of water quality parameters to 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 the Environmental Protection Agency (EPA), which subsequently presented them nearly intact [2]. The NAS criteria are intended to preserve water quality for public water supplies; agricultural uses; industrial uses; recreation and aesthetics; fresh-water aquatic life and wildlife; and marine aquatic life and wildlife. In general, only the first three of these would be affected by ground-water quality. The criteria proposed by the NAS for these three use categories and those imposed by U.S. Public Health Service (USPHS) water standards serve as a framework for identifying significant water quality information to be provided by a ground-water information management system [3]. A composite list of the NAS and USPHS quality criteria for public, agricultural, and industrial water use is presented in Table 1.

The set of quality parameters to be examined by any individual ground-water quality monitoring program would, for the most part, be a subset of most of Table 1. Additional parameters, not appearing in Table 1, might be included as dictated by specific situations. The list in Table 1 is by no means exhaustive, however. The inadequacy of the list for compliance monitoring purposes is reflected, for example, in "The Toxic Substances List" published in 1973 by USDHEW's National Institute for Occupational Safety and Health. This document identifies 11,000 "toxic," chemically unique substances [4]. It is reasonable to assume that significant amounts of any of these substances could be introduced either intentionally or unintentionally into a sub-surface water reservoir. An adequate ground-water information management system, therefore, must be flexible enough to accommodate a large and inconsistent set of variables.

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

Water quality parameter identifiers, because of the large number

Table 1. Candidate Water Quality Parameters for Ground-Water Monitoring

Alkalinity ($CaCO_3$)	pH
Ammonia	Phenolic compounds
Arsenic[a]	Phosphate[d]
Aluminum[b]	Phthalate esters[c]
Barium[a]	Polychlorinated biphenyls (PCB)[c]
Boron	Radioactivity
Cadmium[a]	Selenium[a]
Chloride	Silver[a]
Chromium (total)[a]	Silicon[d]
Color (e.g., platinum-cobalt color units)	Sulfate
Copper	Suspended solids[d]
Cyanide[a]	Temperature
Dissolved oxygen	Total dissolved solids (TDS)
Fluoride	Turbidity
Foaming agents (MBAS)	Viruses
Hardness	Zinc
Iron	Carbon chloroform (extractable)[a]
Lead[a]	Beryllium[b]
Manganese	Total coliform[a]
Mercury	Fecal coliform[a]
Nitrate-nitrogen[a]	Bicarbonates[b]
Nitrilotriacetate (NTA)[c]	Cobalt[b]
Nitrite-nitrogen[a]	Bicarbonates[b]
Odor	Cobalt[b]
Oil and grease	Lithium[b]
Organics-carbon adsorbable	Molybdenum[b]
Pesticides[a]	Nickel[b]
Insecticides-chlorinated hydrocarbons	Sodium[b]
Insecticides-organophosphate and carbamate	Vanadium[b]
Herbicides-chlorophenoxy	Calcium[d]
	Potassium[d]

[a] Significant health ramifications
[b] Agricultural impact only
[c] No criteria currently established
[d] Industrial impact only

involved, should be codified and system specific. Since water quality parameters are system specific, the system administrator rather than the DMAs would be responsible for depositing and maintaining this type of data in the ground-water MIS. An individual DMA could establish a special parameter identifier by petitioning the system administrator who would judge the validity, redundancy, and applicability of the new parameter before including it in the data base.

Each water quality parameter identifier entry would consist of two data elements. One would be an alphanumeric discriptor reflecting the common name of the parameter and its associated unit of measure. The second data element would be the system-administrator-assigned numeric code associated with that identifier.

Water Quality Measurements

The results of physical and chemical analyses of ground water, soil, and geologic material samples represent the bulk of the information to be managed by the ground-water MIS. This information is required for both retrieval and computational operations and must 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 data), 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 depth and reliability indicators.

Temporal Data

In order to provide reasonable system utility, water quality information must be maintained as a time series. Water quality data updates, therefore, need to be appending operations rather than destructive updates. When data are not collected at a constant frequency, which is most often the case with ground-water monitoring, the date of sampling must be recorded as sample-specific data with each new set of water quality measurements. Provision for storing dates as searchable information must be incorporated into a ground-water monitoring information system so that any data subset of the period of record can be retrieved.

Significant vertical stratification of water chemistry would also necessitate recording and storing the pumping time in hours prior to sampling. Additionally, composite samples taken over time necessitate recording and storing the duration (in hours) of the composite sampling period.

Information Qualification Data

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. 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 obtained by inputting "reliability indicators" 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 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 ground-water 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 ground-water pollution control functions, the "readiness status" of a control unit in terms of its ability to respond to a pollution incident could also be evaluated. An operational status index of a DMAs or pollution control units would be the responsibility of the their parameters at the frequencies required. 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.

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

Information Indexing

The ground-water quality MIS should provide indexing of two major categories of data sets: water quality data files in the MIS data bank, and ground-water research documentation.

Water quality file abstracts should be accessible by station identifier, geographical coordinates, aquifer code, political jurisdiction, station type, and agency code. Information contained in the water quality file abstract would be station-specific and include parameters monitored, monitoring frequency, and period of record.

All of the information required by the water quality data file index would exist elsewhere in the MIS so that this index can be system generated and does not require user input.

Research documentation indexing does require special user input. Data elements, all of which should be searchable, are document titles, author names, report numbers, performing and sponsoring organizations, report dates, abstracts, key words, and geographical areas of interest.

DATA COLLECTION

The primary factors considered in selecting data collection systems are purchase cost, operating cost, reliability, responsiveness, and data input speeds.

Total MIS expenditures are particularly sensitive to data collection costs since data entry typically accounts for 20 to 40 per cent of electronic data processing costs [5]. In addition, the data entry process represents the single greatest source of error in an MIS.

A wide variety of available capabilities 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 "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 costs and, correspondingly, increasing speed and reliability.

An additional particularly attractive category of devices applicable to ground-water monitoring is source data automation in order to capture primary data in machine readable form. Examples of such equipment are automatic digital recorders used in conjunction with Keck ground-water 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-usable 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.

User access to the ground-water surveillance data base should be in the batch mode whereas access to the information system indexing 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.

An ideal ground-water monitoring information system will provide flexible data flow procedures for both data submission to and data retrieval from the ground-water 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. Access to the ground-water quality data base should be provided to users with unsophisticated communication capabilities as well as to users with highly sophisticated capabilities.

Data collectors should be allowed to submit ground-water data for file creation and update in formatted nonmachine-readable, formatted machine-readable (i.e., punch cards, paper tape, or magnetic tape), and remote access batch modes. Data users should be allowed to obtain ground-water quality data by telephone, letter, and teletype batch inquiry.

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

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

Figure 1 is a diagram of user access to the proposed ground-water MIS as well as interfile data flow. Unary data are not subject to update except where errors necessitate corrections. Multiple (time series) data are subject to update and, therefore, multiple data flow channels are likely to support a high volume of data traffic.

DATA STORAGE

Three general classifications of hardware are available for data storage: internal, secondary, and external. Internal storage is best

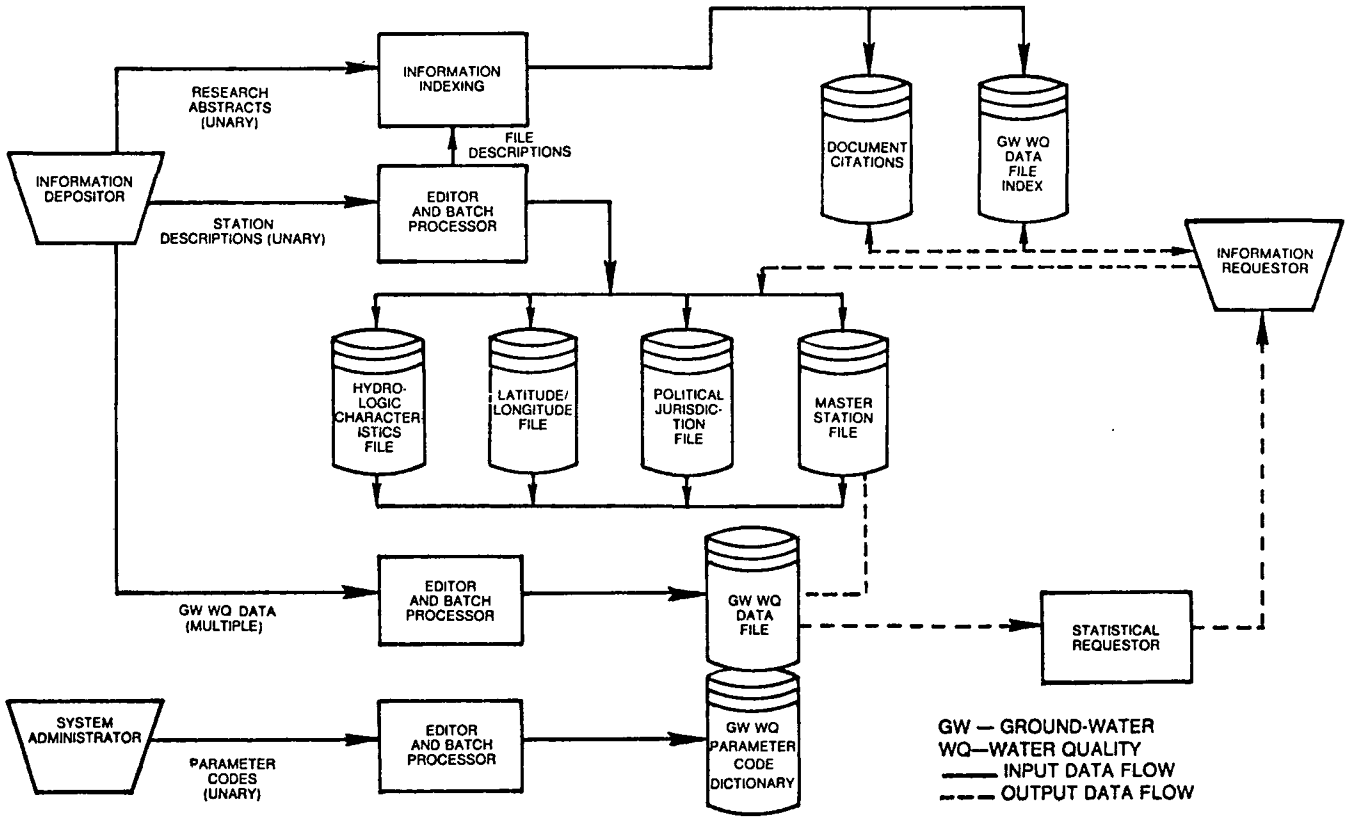

Figure 1. User access to ground-water data base.

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 internal 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 (off-line) storage media include removable disc packs, magnetic tape, punched cards, and paper tapes, all characterized by low access speeds and costs [6].

Figure 1 depicts all of the ground-water 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 ground-water data and provides 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.

As shown in Figure 1, unary station descriptive data should be stored in four separate, directly accessible disc files: the hydrologic characteristics file, the latitude/longitude file, the political jurisdiction file, and the master station file, which contains all station-specific data, including station-specific narrative text. The first three of these files can best be structured as inverted lists to allow rapid access, since they will likely be frequently accessed and infrequently updated. The master station file can be a random access file using station identifiers as record keys.

The water quality data file and parameter code dictionary shown in Figure 1 will also reside in disc storage. The water quality data file contains all sample-specific and measurement-specific ground-water surveillance data as well as DMA status data. Both files can be organized as random access files using station identifiers and parameter codes, respectively, as record keys.

The water quality data file index should be random and use station identifiers as record keys. The document citation file actually consists of a number of randomly accessible subfiles containing all information regarding each document and record keyed by report numbers, document title, author, and originating organization.

DATA PROCESSING

Only access to the ground-water information indexing components of the ground-water MIS requires real-time processing. This requirement is imposed by the users' need to interact intellectually (browse) with the information indexing data base. All other system components should be configured for batch processing because of its more efficient and economical hardware utilization.

Batch processing associated with access to the water quality 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, and compare water quality data with acceptable ranges. 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 and update data files. The retrieval module commands access to the appropriate data files, organizes the requested information, and formats output reports. The statistical processor functions 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

The data retrieval component of the proposed MIS which accesses the water quality data file must yield a wide range of both alphanumeric and pictorial outputs.

The capabilities that the ground-water monitoring MIS can offer the data user vary with the number of retrieval procedures available. Users 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

Users 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, they should be able to request various data manipulation and statistical operations to dictate, to some extent, the format of the output received.

EXISTING SYSTEMS

Several existing or proposed information management systems are relevant to the management of ground-water monitoring information. Table 2 lists some water resources data management systems which are currently operational and their more pertinent characteristics. Table 3 presents a selection of computerized information indexing systems, both operational and proposed, which provide data file or research documentation abstracts. The following discussion describes several existing systems selected for their significance and relevance.

Storage and Retrieval System

The Storage and Retrieval System (STORET) listed in Table 2 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. To date, forty-two 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 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 National Pollutant Discharge Elimination System (NPDES) permit numbers serving as station identification numbers.

The WQF measures the ambient quality of water bodies throughout the nation and the GPSF measures the quality of point source discharges throughout the nation. The software which updates, manipulates, and retrieves data from these files is coded in the

Table 2. Existing Environmental Data Management Systems

System name	Acronym	Administrator	Information	Storage location	Ground water	Computer system
Storage and Retrieval System[a]	STORET	USEPA	Water quality	Centralized	43,000 wells	IBM 371/58-OS/MVT
National Water Data Storage and Retrieval System[b]	WATSTORE	USGS	Surface and ground-water physical and chemical data	Centralized	25,000+ wells	IBM 370/155
ORSANCO Robot Monitor System[c]	—	Ohio River Valley Water, Sanitation Commission	Surface water quality	Centralized	None	IBM 1130
Ground-Water Quality System[d]	—	California	Ground-water quality and hydrographic	Centralized	1,400 wells	CDC 3300
Water Information System for Enforcement[e]	WISE	Michigan DNR	Water quality and discharge inventory	Centralized	22 WQ wells	Burroughs B5500
Tennessee State Ground-Water Data Retrieval System[f]	—	Tennessee Department of Conservation	Ground-water yield and quality	Centralized	75,000 wells 800 springs	IBM 370-OS
Well Hydrograph Data Storage and Retrieval System[g]	DSWELL	ERDA Hanford	Well hydrograph	Centralized	300 wells	PDP-9
Ground-Water Observation Well Network[h]	GOWN	Canada	Well logs, well data, hydrographs	Centralized	75,000 wells	IBM 360/165
Arizona Water Information System[i]	AWIS	Arizona Water Commission	Water resources	Centralized	2,500+ wells	DEC-10

[a] USEPA, 1971
[b] Edwards, 1974
[c] Klein et al., 1968
[d] Welsh, 1973
[e] Guenther et al., 1973
[f] Wilson et al., 1972
[g] Friedrichs, 1972
[h] Gilliland and Treichel, 1968
[i] Foster and DeCook, 1974

Table 3. Computerized Information Indexing Systems

System name	*Acronym*	*Administrator*	*File content*	*Retrieval options*	*Subject*	*File size*	*Computer system*
Remote Control System	RECON	U.S. Department of Energy	Document citations	Key words, 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[a]	80,000 citations	IBM 360/65[a]
Environmental Data Index[b]	ENDEX	NOAA	Data file descriptions	Geographic area (sq), institution, discipline	Environmental	3,000 file references	IBM 360/65
Oceanic and Atmospheric Scientific Information System[b]	OASIS	NOAA (Environmental Data Service, 1974)	Document citations	Title, key word, 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 queries	Water resources	70,000 citations[a]	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

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

PL/1 programming language. Updates and retrievals are done in the batch mode with input provided by card readers of 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 consists of information describing 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 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 five-character parameter identifier code. The parameter identifier codes are also stored in a cross reference (dictionary) file together with the alphanumeric discriptors 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 per cent 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 ground-water 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 ground-water monitoring are presented in Table 4.

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.

Originally, the input module of the STORET system was designed to store only numeric data in the water quality measurement field. The system has been modified, however, to 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 any one of 256 remark codes to be input along with water quality measurements.

Table 4. Established Ground-Water Specific STORET Parameter Codes (USEPA, 1971)

Code	*Output format*[a]	*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[b] well data)
72006	xxxxxxxx	Sampling condition code (BM well data)
72007	xxxxxxxx	Formation name code (BM well data) (AAPG[c] 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 (ft 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 (per cent)
72043	xxxxxx.x	Brake horsepower
72044	xxxxxx.x	Total dynamic pumping head (ft)
72045	xxxxx.xx	Pumping cost in dollars per thousand gallons
72050	xxxxxx.x	Withdrawal of ground water (millions of gallons/month)
72051	xxxxxx.x	Withdrawal of ground water (millions of gallons/year)
84000	xxxxxxxx	Geologic age code (USGS)
84001	xxxxxxxx	Aquifer name code (USGS)

[a] Bureau of Mines
[b] Can be modified at retrieval
[c] American Association of Petroleum Geologists

These remarks are used, for example, to indicate that the stored data element is not accurate, is a field measurement, is a lab measurement, is a lower limit, or is an upper limit.

National Water Data Storage and Retrieval System

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. Access to WATSTORE is through a telecommunication network which provides data services to forty-six 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 files, which contains physical and chemical data reported daily;
- The water quality file, which contains the results of analyses (chemical and physical) of all samples taken. This includes ground-water samples generally taken on an infrequent and irregular basis;
- The peak flow file, which contains annual maximum discharge and stage values for surface water sites;
- The ground-water site inventory file, which contains physical, topographic, aquifer hydraulic, and text data pertinent to ground-water monitoring sites.

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 ground-water 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 accommodate ground-water monitoring stations (wells) that penetrate more than one aquifer and draw samples from individual aquifers separately with the use of screen plugs; i.e., WATSTORE allows for the storage of aquifer identifiers along with the water quality analysis data for each sample.

The aquifer identifiers are stored as eight-character codes based on the stratigraphic coding system proposed by the American Association of Petroleum Geologists [1].

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 five-character code established in cooperation with the EPA STORET User Assistance Branch so that parameter codes are the same in both systems. WATSTORE is equipped with a module which generates STORET input corresponding to WATSTORE data file updates.

National Water Data Exchange

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 station 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 ground water and geographical area of interest (hydrologic basin code). The system allows additional information from the data requestor to further narrow the file search [17].

CONCLUSIONS

The management of ground-water 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). 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, STORET is not now used on a major scale for ground-water analyses and a major new

STORET user community will require a further evaluation and commitment of resources by the EPA.

Ground-water data indexing capabilities, which allow the data user to expeditiously locate pertinent ground-water 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 ground-water investigators and are generally sufficient to meet their needs.

Recommendations for Modifications to Existing Systems

1. The STORET parameter code dictionary should be appended to include the ground-water monitoring related parameters listed in Table 5.
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 ground-water 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 ground-water data file and avoid degradation of update times for the surface water data file.
4. The STORET ground-water file should be maintained on a detachable magnetic disc and placed on-line on the basis of some constant schedule. The periods during which the file will be on-line can be determined by a survey of potential users.
5. Some ground-water 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 or on the basis of its activity level.
6. The proposed STORET ground-water data file should accept compliance monitoring data as well as background information

Table 5. Proposed Additional Ground-Water-Specific 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 (per cent)
84112	Void ratio
84115	Soil bulk density (grams/liter)
84117	Soil moisture content (per cent)
84120	Soil exchangeable sodium (per cent)
84123	Soil specific gravity (g/cm^3)
84130	Soil gradation—per cent clay or silt fines
84131	Soil gradation—per cent fine sand
84132	Soil gradation—per cent medium sand
84133	Soil gradation—per cent coarse sand
84134	Soil gradation—per cent fine gravel
84135	Soil gradation—per cent coarse gravel
84136	Soil gradation—per cent cobbles
84138	Coefficient of soil uniformity
84140	Coefficient of curvature of soil gradation plot
84142	Capillary head (ft)
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 protected use made of aquifer
84500	Monitoring agency status index
84505	Pollution control readiness index
84600-84610	Alphanumeric, sample specific comments—10 fields, 4 characters each

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 ground-water 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 ground-water data file, the eight-character STORET station type code should be modified and interpreted as follows:
   - Column 1 (currently unused) 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 (current usage is other = 1; water quality = 2; flow, tide, well level = 4);
   - In column 5, a 1 would indicate information monitoring, a 2 would indicate compliance monitoring, and a 4 would indicate other (current usage is direct from stream, etc. = 1; intake = 2; outfall = 4);
   - 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 (current usage is ocean = 01; lake = 02; stream = 04; well = 10; land = 20; unused = 40).
8. The STORET ground-water data file should store water quality criteria (ambient or effluent) as sample data. The date of enactment of the criteria should be stored in the STORET sample date field and some exclusive value, such as 8888 for ambient criteria and 9999 for effluent limitation, should be stored in the STORET sample time field.
9. STORET retrieval options should be expanded to allow more extensive Boolean retrieval strategies. 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 nonmachine compatible user inferface with the data base on a routine basis.
11. Either the GIPSY or the RECON document citation retrieval system should be modified to accommodate polygon type

retrievals. This would allow the ground-water investigator to impose geographic delimiters and receive research documentation abstracts for his geographical area of interest.

REFERENCES

1. W. E. Price 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, Virginia, 306 pp., 1974.
2. U.S. Environmental Protection Agency, *Quality Criteria for Water*, Washington, D.C., 256 pp., July 1976.
3. U.S. Public Health Service, Drinking Water Standards, Federal Register, Government Printing Office, Washington, D.C., pp. 2152-2155, March 6, 1962.
4. U.S. Department of Health, Education and Welfare, *The Toxic Substances List, 1973 Edition*, National Institute for Occupational Safety and Health, Rockville, Maryland, 1001 pp., 1973.
5. R. Ferrara and R. L. Nolan, New Look at Computer Data Entry, *Journal of Systems Management*, Association for Systems Management, pp. 24-33, February 1973.
6. J. Lobel and M. V. Farina, Selecting Computer Memory Devices, *Automation*, Penton Publishing Co., Cleveland, Ohio, pp. 66-70, October 1970.
7. U.S. Environmental Protection Agency, *Storage and Retrieval of Water Quality Data. Training Manual*, PB-214 580, Washington, D.C., 302 pp., 1971.
8. M. D. Edwards, *The Processing and Storage of Water Quality in the National Water Data Storage and Retrieval System*, U.S. Geological Survey, Water Resources Division, Reston, Virginia, 85 pp., 1974.
9. W. L. Klein, D. A. Dunsmore, and R. K. Horton, An Integrated Monitoring System for Water Quality Management in the Ohio Valley, *Environmental Science and Technology*, American Chemical Society, *2*, pp. 764-771, October 1968.
10. J. L. Welsh, Ground-Water Quality Data for Planning, Monitoring, and Surveillance, *Proceedings at the Ninth Biennial Conference on Ground Water*, Goleta, California, September 1973.
11. G. Guenther, 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 pp., 1973.
12. J. M. Wilson, 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.
13. D. R. Friedrichs, *Information Storage and Retrieval System for Well Hydrograph Data—User's Manual*, Battelle Pacific Northwest Laboratories, Richland, Washington, 23 pp., 1972.

14. J. A. Gilliland and A. Treichel, GOWN—A Computer Storage System for Groundwater Data, *Canadian Journal of Earth Sciences, 5,* pp. 1518-1524, September 1968.
15. K. E. Foster 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 pp., 1974.
16. Environmental Data Service, *User's Guide to OASIS—Oceanic and Atmospheric Scientific Information System,* National Oceanic and Atmospheric Administration, Washington, D.C., 1974.
17. Planning Research Corporation, *Support in the Implementation of a National Water Data Exchange,* Second Quarterly Progress Report (September-November 1974), PRC-D-1863, 61 pp., December 1974.

Direct reprint requests to:

Norman F. Hampton
Aerospace Corporation
Germantown, Maryland