Conceptual Design for the National Water Information System

U.S. DEPARTMENT OF THE INTERIOR

March 3, 1849

Conceptual Design for the National Water Information System

By MELVIN D. EDWARDS, ARTHUR L. PUTNAM, and NORMAN E. HUTCHISON

An overview of the online relational data base
that is scheduled to replace and augment
existing water-resource information services
operated by the Survey's Water Resources Division

U.S. GEOLOGICAL SURVEY BULLETIN 1792

DEPARTMENT OF THE INTERIOR

DONALD PAUL HODEL, Secretary

U.S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

CONTENTS

FIGURES

TABLE

Conceptual Design for the National Water Information System

By Melvin D. Edwards, Arthur L. Putnam, and Norman E. Hutchison

Abstract

The Water Resources Division (WRD) of the U.S. Geological Survey began to design and develop a National Water Information System (NWIS) in 1983. The NWIS will replace and integrate the existing data systems of the National Water-Data Storage and Retrieval System, National Water-Data Exchange, National Water-Use Information Program, and Water Resources Scientific Information Center. The NWIS is planned as an interactive, distributed data system. The software has been designed as a modular structure that integrates existing functions and allows multiple use of the various facilities. The data base incorporates a relational data model that allows integrated storage of the existing water data, water-use data, and water-data-indexing information within a common relational database-management system. The NWIS will be operated on microcomputers located in each of the WRD's district offices and in many of its State, subdistrict, and field offices; these microcomputers will be linked together through a national telecommunication network maintained by the U.S. Geological Survey. The NWIS is scheduled to go online in 1990.

1. INTRODUCTION

In September 1982, the Water Resources Division (WRD) of the U.S. Geological Survey began to install a nationwide, network of minicomputers for the storage, retrieval, analysis, and dissemination of its hydrologic data. This network, known as a Distributed Information System (DIS), is now fully operational and consists of nearly 70 minicomputers located in the WRD's regional, State, district, and subdistrict offices, a central laboratory, the Hydrologic Instrumentation Facility (HIF), and the national headquarters. These systems are linked through a national, Government-owned telecommunication network, known as GEONET.

The use of minicomputers in a distributed environment introduced new data-processing technology to the WRD. To fully use this technology, the WRD recognized that it must reorganize, redesign, and integrate its large, centralized data systems for operation in the distributed environment. A special task force was established in March 1983 to develop the plans, designs, and procedures for a National Water Information System (NWIS), so as to implement the WRD's existing water-related data and information systems of the following programs as part of the DIS:

- The National Water-Data Storage and Retrieval System (WATSTORE)—The primary system used by the WRD for the storage, retrieval, and dissemination of its water data.

- The National Water-Data Exchange (NAWDEX)—An interagency program that indexes the water data available from both Federal and non-Federal agencies and coordinates a nationwide user-service program for the identification and acquisition of available water data.

- The National Water-Use Information Program (NWUIP)—A national program for gathering site-specific water-use information through a network of State Water-Use Data Systems and for aggregating water-use information in a National Water-Use Data System.

- The Water Resources Scientific Information Center (WRSIC)—A program for abstracting water-resources publications worldwide and for providing a bibliographic user-service program.

The task force completed the original conceptual design for the NWIS in October 1983 on the basis of requirements determined through an analysis of the existing water-data systems of the WRD (M.D. Edwards and others, written commun., 1983). This report updates the original conceptual design developed by the task force to include new concepts, design changes, and technologies that have been identified or developed since 1983.

1.1. Objectives

The NWIS involves a complete redesign and modernization of the existing water-data systems that it will replace. Because of the rapid changes and advances in today's hardware and software technologies, future software and data systems are expected to have significantly shorter life cycles if these systems are to keep pace with modern technology. The NWIS must, therefore, be highly flexible, such that it can be easily changed and expanded in a rapidly changing technological environment. On the basis of these general requirements and additional requirements derived from analysis of the existing systems, the following major objectives were established for the design of the NWIS:

- Increase local processing for the WATSTORE, NAWDEX, NWUIP, and WRSIC programs by using hardware available through the Division's DIS.

- Develop the NWIS as a distributed data system to take advantage of the nationwide telecommunication capabilities of the DIS, to store the data at the points of highest need and use, and to make all the data and software resources of the NWIS more readily and conveniently available to all users.

- Provide the basic data storage and processing capabilities needed to meet all of the WRD's data-handling needs, including those for research and special projects, allowing scientists to focus on scientific applications.

- Provide a single flexible and expandable system that is easy to use, with fewer structures than the existing systems, so as to improve productivity and use.

- Provide standardization and uniformity of data handling, data storage, and software procedures, thus increasing the integrity of the data bases and software systems.

- Provide modular data-base and software systems that can be developed in a distributed environment by using application prototyping, to allow the development of thoroughly tested and efficient software and the ability to easily change, enhance, expand, and maintain implemented data bases and systems.

1.2. Scope

The NWIS is being designed and developed for the storage, handling, analysis, and dissemination of all water data and information about water data within the WRD. The major focus of the conceptual design of the NWIS is on the data systems of the WATSTORE, NAWDEX, NWUIP, and WRSIC programs. The NWIS is also intended to include the administrative subsystems required to operate and manage it.

1.3. Purpose of the Conceptual Design

The conceptual design is provided to present an overview of the NWIS' objectives, scope, network architecture, data-base content and structure, and hardware and software architectures. This design will allow the user community to review and evaluate the proposed system for relevance and completeness, deficiencies, and needed enhancements before its development. Users will also have an opportunity to provide input to the system's structure, functions, and requirements by suggesting changes or enhancements that will make it more effective and useful.

1.4. Acknowledgments

2. OVERVIEW

The NWIS is designed as a distributed data system that will integrate and replace the existing, centralized software and data-base systems of the WATSTORE, NAWDEX, NWUIP, and WRSIC programs. The NWIS will be operated within the GEONET telecommunication network installed and operated by the U.S. Geological Survey. Local nodes of this network will be further supported by local-area networks (LAN's) for the interoffice and intraoffice sharing of data, hardware, and software.

The NWIS is designed to operate on a wide range of computer hardware, including microcomputers. Several 32-bit microcomputers will be evaluated for use as central processing units at each node for data-base storage and management, and for NWIS general-purpose computing. Less powerful, 16-bit or smaller 32-bit microcomputers will be used as workstations linked to the central processors via LAN's. Various peripheral hardware, including laser printers, high-speed impact printers, digitizers, plotters, and file and communication servers, will be used and shared at each node. Mass-storage devices may be used at nodes with large data-storage requirements of more than 1 gigabyte (GB). Two types of mass-storage devices will be available for use: (1) data-base servers using magnetic-disk technology, for highly volatile data; and (2) optical-disk storage devices that use laser-disk technology, for static data requiring infrequent updating.

Most NWIS data will be stored in an online environment, using a relational data-base-management system (RDBMS). Data with special characteristics, such as text data, may be stored in other types of data-base-management systems to provide more efficient manipulation.

A total of 13 major files and data bases of the existing data systems will be integrated into 6 discrete data components in the NWIS. In addition, five new data components will be added to the NWIS that do not exist in the current systems: (1) Areal-Data Index, (2) Data-Elements Dictionary (DED), (3) Digitized/Line Data, (4) Administrative Data, and (5) Support Data. The following types of data will be stored in the NWIS:

- Index Data:
  — Site Index
  — Organization Index
  — Areal Index

- Data-Elements Dictionary (data-element definitions, precision of reporting, column headings, derivation algorithms, and other related data)

- Support Data (State/county codes, hydrologic-unit codes, agency codes, geologic-unit codes, rating curves, and other similar data)

- Text Data (station headings, bibliographic citations, and water-resources abstracts)

- Digitized/Line Data (State/county boundaries, hydrologic-unit boundaries, river traces, and other geographic boundaries and features)

- Administrative Data (user accounting, computer charges, and data-base statistics)

- Hydrologic Data:
  — Continuous Data (unit values, daily values)
  — Discrete Data (quality-of-water [QW] analyses, peakflow, ground-water [GW] levels, and other data that are time dependent)
  — Descriptive Data (GW site inventories, basin characteristics, and other characterists that are not time dependent)

Most of the basic hydrologic data collected by the WRD will be stored, managed, and disseminated by its district offices, except for specialized research and special-project data collected and stored by its regional and project offices. National headquarters will store a national index of all the data stored in the distributed network, as well as aggregate data, such as national water-use data and bibliographic data, that are not amenable to geographic distribution.

Software for the NWIS will be interactive and support batch-processing applications. The software is designed as a modular structure to maximize sharing of software subroutines and procedures and to minimize redundant code. Entry to the software will be through an interactive controller that will consist of a family of system and user-invoked menus, subroutines, and procedures to perform various control and administrative functions, including validating access privileges, recording access time and computing user charges, developing user profiles, computing data and software-usage statistics, maintaining and managing online documentation, and transferring control to and from selected software packages. A command mode will allow more experienced users to access the NWIS software directly, with only minimal interaction with the controller.

Use of the NWIS software will be supported by online documentation in the form of an "information desk" that will contain descriptions of NWIS data, software subsystems, and individual software packages, as well as prompts and tutorials that can be invoked by the user at any time by the use of "HELP" commands.

The NWIS software will be structured around relational and other data-base-management systems that will provide significant system support, in addition to full data-base management and maintenance. Software functions include data-base-design aids, screen generation and management facilities, data-base security, data-dictionary capabilities, an interactive query language, a report generator, an electronic spreadsheet, graphics and plotting capabilities, and application-code-generation capabilities. Software modules within the NWIS will interact directly with the data-base-management systems and provide all the basic data-processing requirements. These modules will include: data-indexing software for the self-indexing of data stored in the NWIS; edit and data-entry software for the common editing of data, using criteria stored in the DED; storage of data values in the data base; and generation of an audit trail to be used for data-indexing purposes. Retrieval software that will generate output in standardized formats, perform required computations of data values on output, and sort output to fit user specifications will also be included.

Various standardized interface and processing software will be provided to support the processing and entry of data from the WRD's central laboratory and automated-instrumentation systems. Several general-application packages will also be provided to perform commonly used applications, including tabling, plotting, and statistical analysis. The NWIS will also interface with several external data and software systems through modules provided to perform the data and format transformations needed to achieve compatibility with these external systems.

Much of the special-application software for the NWIS is expected to be provided by users of the system, including hydrologic models, Geographic Information System (GIS) applications, special analytical procedures, and data-display applications customized to fit user needs. Thus, software developed by researchers, district-office personnel, and others may be more widely shared with the NWIS user community. The conceptual design of the NWIS is intended to be dynamic and will continue to be changed and expanded as new requirements are identified for the service.

3. CONCEPTUAL DESIGN

3.1. Network Architecture

The NWIS will operate within the GEONET system installed by the U.S. Geological Survey (fig. 1). GEONET is a private packet-switching network that uses the X.25

telecommunication protocol within the TYMNET public network. Through this network, all offices of the Division and a large community of outside users will have access to the NWIS.

3.1.1. NWIS Nodes

The NWIS will be operated on various computer nodes located throughout the DIS network. Eight major types of nodes have been identified:

● National node

● State, district, subdistrict, field, and project nodes

● Regional nodes

● Central-laboratory node

● Hydrologic Instrumentation Facility (HIF) node

● Special-application nodes

● Interface nodes

● External users and systems

As discussed below, each node may operate and maintain a full or partial configuration of the NWIS, depending on its data storage, processing, and application needs. Varying configurations of the NWIS at each node may complicate the overall management of the system, but it is not practical to require a node to manage and maintain large volumes of data and software that are not needed. Thus, the hardware
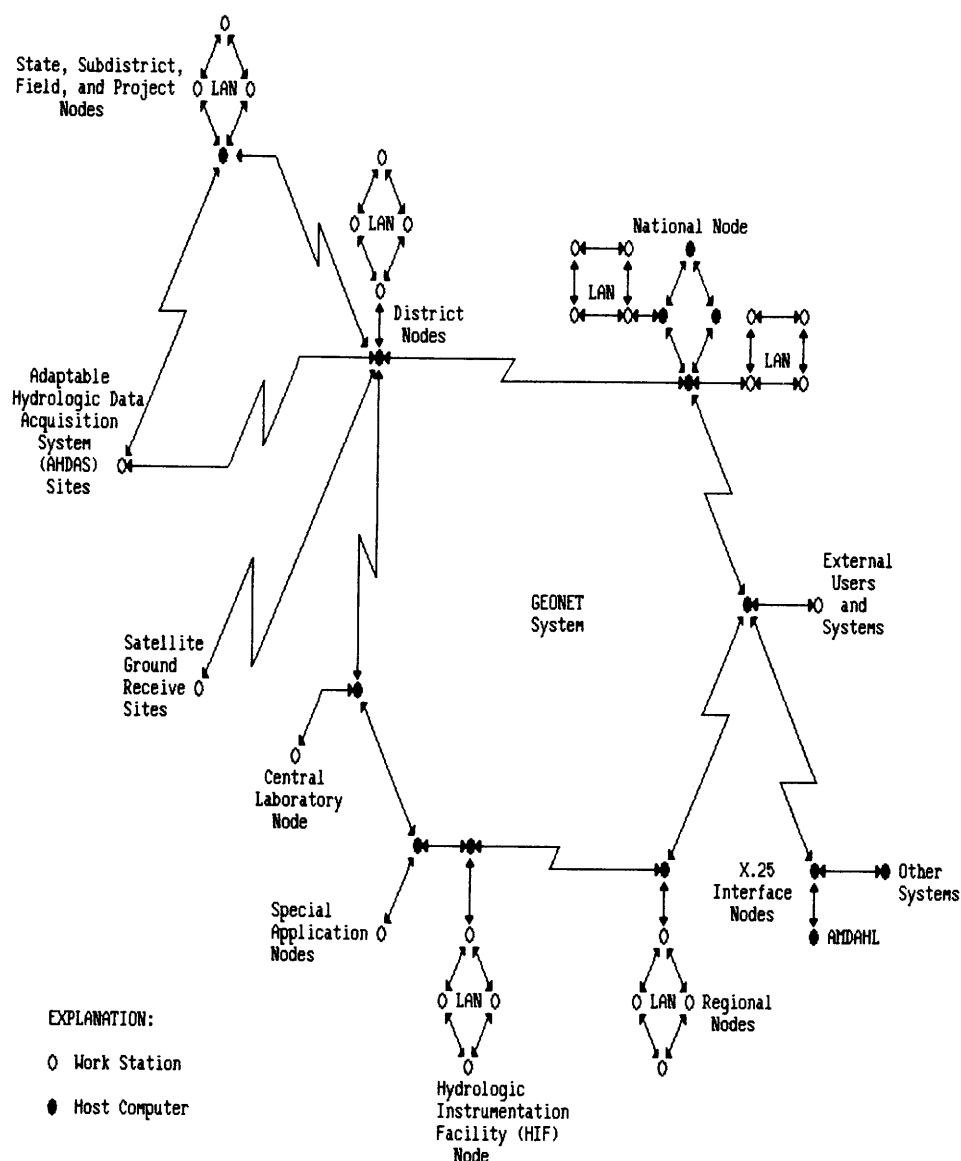
Figure 1. Network architecture of the National Water Information System.

configurations for each node will be scaled to its actual computing needs.

3.1.1.1. The National Node

The national node of the NWIS will be located at the U.S. Geological Survey's National Center in Reston, Va., where it will be managed and operated by the Office of the Assistant Chief Hydrologist for Scientific Information Management. The node configuration will consist of multiple computers that may be used independently or linked together to allow better use of available resources. The national node will perform various functions, including:

- Systems and network management

- Systems research, evaluation, development, and testing

- Software archiving and dissemination

- Central indexing of the NWIS distributed-data network

- Global access to the NWIS network

- Processing and dissemination of water and water-related data requiring centralized storage

3.1.1.2. State, District, Subdistrict, Field, and Project Nodes

The State, district, and subdistrict nodes will form the bulk of the NWIS network and will be distributed nationwide. They may consist of single or multiple computers that will process, store, and disseminate all detailed, site-specific water and water-related data collected by the WRD. They can be subdivided into three major categories:

- Nodes with NWIS data

- Nodes without NWIS data

- Nodes with NWIS data that additionally support Direct-Readout Ground Stations (DRGS) for the receipt and processing of data relayed by satellite from remote locations and for the dissemination of these data to WRD and cooperator offices

Each district node will operate and maintain a complete configuration of the NWIS system and will be responsible for processing, storing, indexing, managing, and disseminating the water data collected within its geographic area of responsibility. Larger districts may elect to delegate part of their NWIS responsibilities to State or subdistrict nodes, each of which may optionally operate and maintain a full or partial configuration of the NWIS, as required, to meet its data processing and storage requirements.

Many subdistrict, field, and project nodes will not store NWIS data but will remotely access the NWIS host node. These nodes will operate and maintain a partial configuration of the NWIS tailored to their needs that will allow them to receive and temporarily store data for local application and use.

Some district nodes will be additionally responsible for receiving data transmitted by data-collection platforms (DCP's) via satellite to DRGS's. Each receiving node will be responsible for disseminating these data to several designated State, district, and subdistrict nodes for processing and storage. Each receiving node will also be responsible for providing backup support to at least one other receiving node.

All the district nodes and many of the State and subdistrict nodes will eventually receive and process data collected by the Adaptable Hydrologic Data Acquisition System (AHDAS), which is under development by the WRD. The AHDAS is being developed to collect data at remote locations by using microcomputer technology. Data collected by each AHDAS unit may be (1) relayed by satellite or telephone linkages to a host computer for processing, or (2) taken from the unit in the field by a portable field interrogator that will temporarily store the data in an electronic or magnetic medium for subsequent entry into a host computer for processing.

3.1.1.3. Regional Nodes

Each of the WRD's regional offices, located in Reston, Va., Atlanta, Ga., Denver (Lakewood), Colo., and Menlo Park, Calif., may optionally operate and maintain a full or partial configuration of the NWIS system, as needed. These nodes may consist of multiple computers and will use the NWIS for processing, storage, analysis, and dissemination of the research-related data and data collected by regional projects.

3.1.1.4. Central Laboratory Node

The central laboratory, located in Denver (Arvada), Colo., which processes water samples collected by WRD offices nationwide, will serve as a major source of data for the NWIS. This laboratory will maintain and operate a partial configuration of the NWIS sufficient to interface with its internal processing systems and temporarily store data before their transfer to other nodes for subsequent processing, validation, and storage.

3.1.1.5. Hydrologic Instrumentation Facility Node

The HIF node, located in Bay St. Louis, Miss., may consist of multiple computers that will be shared by the Gulf Coast Hydroscience Center, which is involved in research and modeling activities. This node may optionally operate and maintain a full or partial configuration of the NWIS, as needed to meet its research and modeling data requirements.

3.1.1.6. Special-Application Nodes

Several computer nodes are expected to be configured and dedicated to special applications, for example, for operating and maintaining GIS applications, and for local or regional research projects and modeling activities. These nodes may optionally operate and maintain a full or partial configuration of the NWIS, as needed to meet their water-data acquisition, processing, and storage requirements.

3.1.1.7. Interface Nodes

There is a need to continue to access, and provide access by, computers external to the NWIS network that are incompatible with the X.25 telecommunication protocol of GEONET. One or more computer nodes will be maintained as a part of the network to provide the software emulations needed to interface with these external systems. These interface nodes will not operate and maintain the NWIS system but will allow access to the NWIS by, or access by the NWIS to, incompatible external systems.

3.1.1.8. External Users and Systems

The WRD currently provides access to its NAWDEX and WATSTORE data systems to more than 170 organizations at more than 260 remote locations nationwide. This service will continue to be provided by the NWIS. These organizations will not operate or maintain the NWIS system but will access it for data-retrieval purposes. Most users will continue to be serviced by dialup lines to GEONET. Some Federal and State cooperative agencies and organizations that serve as NAWDEX Assistance Centers may, however, require direct interfaces between their computer systems and the NWIS.

3.1.2. Local-Area Networks

Use of microcomputers, word processors, graphics devices, laser printers, and other hardware and software is rapidly increasing within the WRD. These systems can be most effectively used in a shared environment. LAN's are expected to be used extensively at the NWIS nodes for interoffice and intraoffice sharing of data, hardware, and software.

A LAN is a system for the interconnection of two or more communicating devices (Datapro Research Corp., 1985). Within a LAN, many types of applications, such as data processing, word processing, and electronic mail, can operate over a single communications cable. Each device on the LAN is able to communicate with every other device. A LAN allows numerous "intelligent" devices to share such resources as software, data bases, storage devices, file servers, printers, and connections to outside communications. The existence of these capabilities is an important factor in the distributed and modular design of the data-base and software systems of the NWIS, as discussed below. Each LAN will allow access to

GEONET and any workstation on the LAN to communicate with computers and other devices throughout the NWIS network, so as to share and exchange data and software (fig. 1).

3.2. Hardware Configurations

The original concept for the NWIS developed in 1983 was directed at operating on the existing DIS network of minicomputers. The NWIS has a projected development life cycle of 5 to 7 years, with nearly full implementation scheduled for 1990; the existing DIS minicomputer hardware has a projected life cycle of 8 years, which will also be reached in 1990. Therefore, the conceptual design of the NWIS has been modified to utilize computer hardware expected to be state of the art in 1990, not the current DIS hardware, and so it was decided to begin development of the NWIS by using microcomputer technology. The use of microcomputers provides several significant advantages during the development life cycle:

● Microcomputers are the lowest cost option for hardware during development.

● The lower cost advantage will allow testing and evaluation of diverse types of hardware to keep pace with rapid advances in computer technology.

● Diversity of computers during the development life cycle will permit testing and evaluation of several types of software systems and peripheral hardware devices.

● Progressive expansion of the prototyped NWIS over time will allow accurate redefinition of the sizes and computing requirements for each component of the system. Software, hardware, sizing, and computing requirements will be clearly defined and documented by the end of the development life cycle. Inadequate hardware selections will be avoided, and hardware can be selected to fit specific requirements of the NWIS software and data-base systems, to prevent costly retrofitting and modifications after hardware procurement.

The first prototype NWIS is currently being developed, using microcomputer workstations interfaced with the existing minicomputers; subsequent prototypes will be developed using only microcomputers. By the end of the development life cycle, the feasibility of operating the NWIS in a microcomputer environment, a minicomputer environment, or a hybrid environment using both scales of hardware will be more clearly defined.

Various computer and peripheral hardware may be used in the operation of the NWIS (fig. 2), as discussed below. This hardware will be linked and shared over the LAN's that will be a part of the DIS network.

3.2.1. Microcomputer Central Processors

One or more microcomputers will be the central processor(s) at each NWIS node; the number used will depend on the computing and data-storage requirements of each node. These microcomputers will be used for both data-base management and general data processing. The following are desirable features for any single microcomputer to be tested and evaluated for these purposes:

- Microcomputer:
  — Full 32-bit-word memory transfer
  — Memory expandable from 4 to 16 megabytes (MB)
  — Connectable to selected LAN's
  — Connectable to the X.25 GEONET network
  — Support as many as 32 concurrent users
  — Multitasking capability
  — Distributed-data-base capability
  — Shared-data-base capability
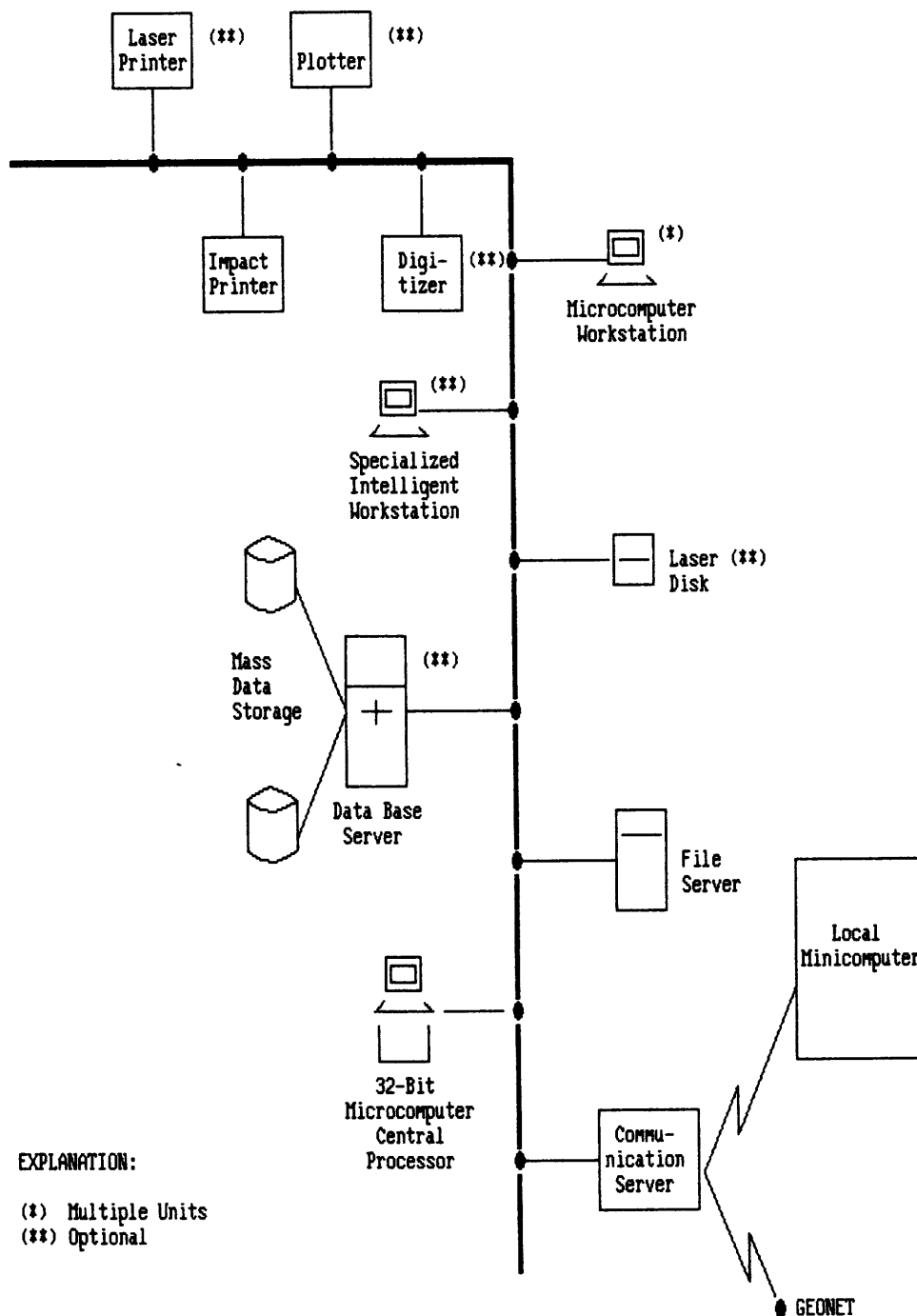
Figure 2. Potential computer and peripheral hardware for use in operation of the National Water Information System.

- Peripheral devices:
  — Fixed-disk storage of 500 to 2,000 MB
  — Floppy-disk storage
  — Streamer and nine-track-tape storage

- Software Support:
  — RDBMS
  — UNIX or UNIX-based operating system
  — High-level-language compilers:
    □ Fortran 77
    □ C
    □ Cobol
    □ Pascal
    □ BASIC
  — Structure Query Language (SQL)
  — Indexed Sequential Access Method (ISAM) utility, with interface to Fortran 77
  — Full-screen editor
  — Text editor
  — Electronic spreadsheet

3.2.2. Microcomputer Workstations

Smaller, less powerful microcomputers will be used as workstations inter-connected with the microcomputer central processor for both general data processing and minimal data-base-management purposes. Specialized, "intelligent" microcomputer workstations may also be used to support special-purpose peripheral hardware that may be integrated into local NWIS nodes. The following are desirable features of these microcomputer workstations:

- Full 16- or 32-bit-word memory transfer

- Memory expandable from 1 to 4 MB

- Connectable to LAN's used by the U.S. Geological Survey

- Fixed-disk storage of from 20 to 100 MB

- Low (360 KB)- and high (1.2 MB)-density floppy-disk storage

- High-resolution graphics monitor

- Software support of:
  — RDBMS
  — UNIX or UNIX-based operating system
  — High-level-language compilers for:
    □ Fortran 77
    □ C
    □ Pascal
    □ BASIC
  — SQL

— Electronic spreadsheet
— Word processor/text editor

3.2.3. Mass-Data-Storage Devices

Some NWIS nodes, specifically the national node, will have large data-storage requirements (more than 1 GB), with a need for frequent, high-speed access. Two types of mass-data-storage devices will be tested and evaluated for efficient management of these large data bases:

- Data-base servers that utilize magnetic-disk technology for high-speed storage, retrieval, and management of large data volumes, ranging from 1 to 3 GB. These devices will be considered for data, such as central indexes of data stored in the NWIS, that have a high degree of volatility and require frequent updating.

- Optical-disk storage devices that utilize laser-disk technology, with very high storage capacities and high-speed access rates. Two types exist: (1) read-only drives that have no disk-writing capability, where master disks are created and duplicated for distribution by using specialized equipment; and (2) write-once-read-many (WORM) drives, which allow data to be written on the disk but have no inplace updating capability. These devices will be considered for use with static data that require no, or infrequent, updating. Examples of these types of data are: support data, such as State, county, and hydrologic-unit codes; line data, such as State, county, and hydrologic-unit boundaries and stream traces; and raw data received via satellite or other electronic media that must be permanently archived in their original forms.

Both types of devices have applicability for data storage in the NWIS and are expected to be used within it.

3.2.4. Peripheral Hardware

Various peripheral hardware will be required in support of the computer and software systems of the NWIS, including:

- Impact and laser printers with graphics capability

- Line (vector) and raster plotter

- Digitizers

- File servers that support the filing requirements of workstations on LAN's

- Communication servers to provide gateways for remote access to other systems on the telecommunication network

3.3. Data-Base Architecture

More than 6 GB of water data and related information are stored in the existing data systems of the WRD. These data must be converted and transferred for distributed storage in the NWIS data base. Nearly an equal number of additional water and water-related data have been identified in project and other special data systems throughout the WRD that are desirable for conversion to and storage in the NWIS data base. Major objectives in the design of a data base to support these data are to:

- Provide ease and speed in accessing data

- Integrate the data, to the highest degree possible, into a single data base

- Minimize redundancy of data storage

- Provide for the integrated use of all types of data stored in the data base

3.3.1. Data Integration

The existing water-data and information systems of the WRD consist of 13 major files and data bases managed by the NAWDEX, WATSTORE, NWUIP, and WRSIC programs. Each of these files or data bases is managed separately by the individual programs and has its own associated data management, processing, and application-software systems. The design of the NWIS data base integrates these 13 major files and data bases into 6 discrete data-base components (fig.3). All components, except the Selected Water Resources Abstracts component of the WRSIC program, which is a text-based system, will be managed by a single data-base-management system. Under this structure, each component may still be managed separately but will be integrated with common software systems. The Selected Water Resources Abstracts component will be managed under a separate data-base-management system for more efficient use of its text-type data.

The following integrations of data have been accomplished in the NWIS data-base design (fig. 3):

- The Water Data Sources Directory and Water Supply Computerized Information Directory data bases of the NAWDEX program are integrated into a single Organization Index component.

- The Master Water Data Index of the NAWDEX program, the Station Header File of the WATSTORE program, and the site-descriptive portion of the

Ground-Water Site Inventory File of the WATSTORE program have been integrated into a single Site Index component.

- The data in the Ground-Water Site Inventory File and the Basin Characteristics File of the WATSTORE program that are not time dependent have been integrated into a single Descriptive Data component.

- The Peak Flow File and Water-Quality File of the WATSTORE program, and the National Water-Use Data System and State Water-Use Data System of the NWUIP program, have been integrated into a single Discrete Data component.

- The Unit Values File and Daily Values File of the WATSTORE program have been integrated into a single Continuous Data component.

- The Selected Water Resources Abstracts data base of the WRSIC program will be moved to a Text Data component. This component may also be used for the future integration of other types of text-based data, such as station headings and descriptions, system documentation, and user instructions.

3.3.2. Data-Base Structure

The NWIS data base has been designed on a relational data model (A.L. Edler, written commun., 1986), consisting of 11 major components, each of which contains data of different characteristics and usage requirements (fig.4):

- Site Index: This component will store information about individual sites for which data are stored in the NWIS. In addition, it will store information about sites for which data are available from other agencies through the NAWDEX program, including information about each site's geographic location, types of data available, frequency of measurement of the data available, time periods for which data are available, history of data collected at the site, and the data base or computer-node location of the data.

- Organization Index: This component will store information about organizations that store data in the NWIS or are sources of water or water-related data and services made available through the NAWDEX program. It will include information about each organization's location, its data-collection programs, services available, types of data available, office locations and addresses, and contacts for the acquisition of data or services.

- Areal Index: This component will store information about data stored in the NWIS that are collected or aggregated on an areal basis, such as by county or river basin.

- Data-Elements Dictionary: This component will store identifiers for all data elements stored in the NWIS data base, along with definitions, column headings, edit criteria, reporting precisions, derivation algorithms, and other items necessary for processing the data values associated with each data element.

- Support Data: Various support data will be stored in the NWIS in conjunction with the DED, including:

  — Rating Data: These data include rating curves, calibration curves, and other relations required to convert the relative values recorded by instrumentation to engineering units, or to perform other data transformations.
  — Station Computation Data: These data include datum corrections, shift corrections, and other data used to support computations within the system.
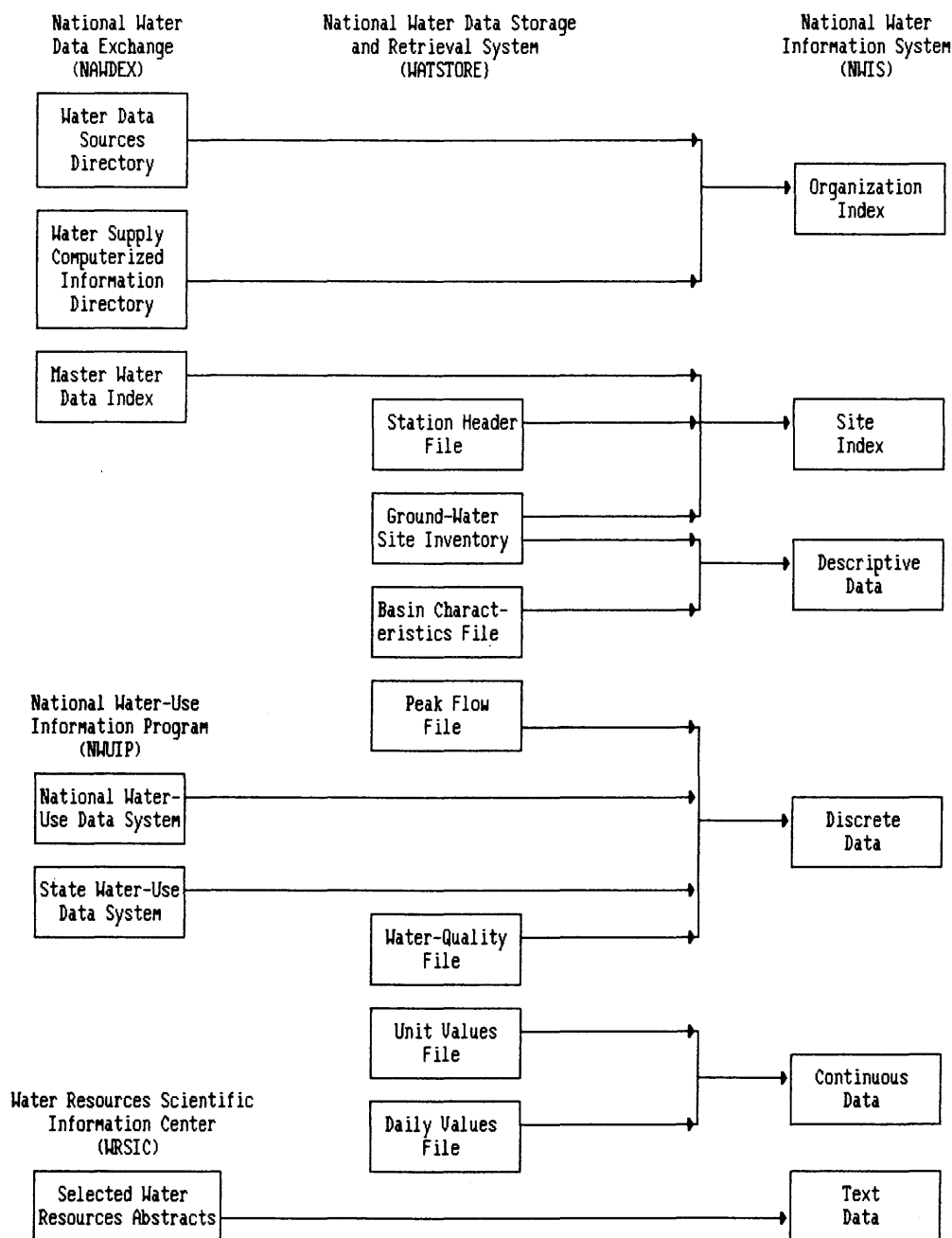
Figure 3. Conversion of existing data systems to the National Water Information System.

● Validation Codes: Several code tables will be stored in the NWIS to support the validation of data-element values and to provide names to be substituted on output for codes and mnemonics stored as data values. These tables will include:
  ❑ State and county codes and names
  ❑ City and place codes and names
  ❑ Country codes and names
  ❑ Standard Metropolitan Statistical Area (SMSA) codes and names
  ❑ WRD district codes and names
  ❑ Hydrologic-unit codes and names
  ❑ Geologic-unit codes and aquifer names
  ❑ Agency codes and names
  ❑ Parameter codes and names

● Text Data: Various text-type data will be stored in the NWIS, including online system documentation and user instructions, descriptive headings for hydrologic sites, and abstracts and bibliographic citations of water-related literature.

● Digitized/Line Data: Various data may be stored in the NWIS in the form of digitized areal representations or as digital line representations of geographic boundaries and features. These data will be used for GIS applications, data verification, data derivation (State, county, and hydrologic-unit codes), data retrieval, data mapping, and other applications. The data will include:
  — Satellite-imagery data
  — Land-use data
  — Land-cover data
  — State and county boundaries
  — Hydrologic-unit boundaries
  — River traces
  — Project boundaries
  — User-provided polygon boundaries defining geographic boundaries for mapping and other purposes

● Administrative Data: Various administrative data will be required to support the operation and management of the NWIS, including:
  — User accounting and billing data: Data relating to requests received for data, routing of requests to other sources for response, referrals of users to other sources and services, data provided, and charges billed to the user.
  — Computer charges: Accounting data applicable to account numbers assigned and charges made and billed to outside users of the NWIS computer, data, and software systems.
  — User profiles: Statistics relating to system components used by individual users, frequency of use, and types of applications.
  — Data-base statistics: Data relating to volume of storage, growth rates, frequency of use, type of use, and source of data stored in the NWIS.
  — Addresses: Mailing addresses for data dissemination, user billing, and other activities related to management and operation of the NWIS.

● Continuous Hydrologic Data: This component will store data values measured once daily, measured continuously and reduced to once-daily values, or measured more frequently than daily and stored at fixed or irregular time intervals. Examples of these data are daily and unit values of river stage or stream discharge, and daily maximum, minimum, or mean values of temperature and other QW parameters.

● Discrete Hydrologic Data: This component will store data collected or measured at irregular or fixed time intervals less frequent than daily (that is, weekly, quarterly, monthly, annually, or at other periodic intervals). These could be single-parameter data, such as GW levels, or multiparameter data, such as chemical analyses of water samples. This component will also store other types of time-dependent data, including water-use data and extreme events, such as peak streamflows, low streamflows, and QW alert values.
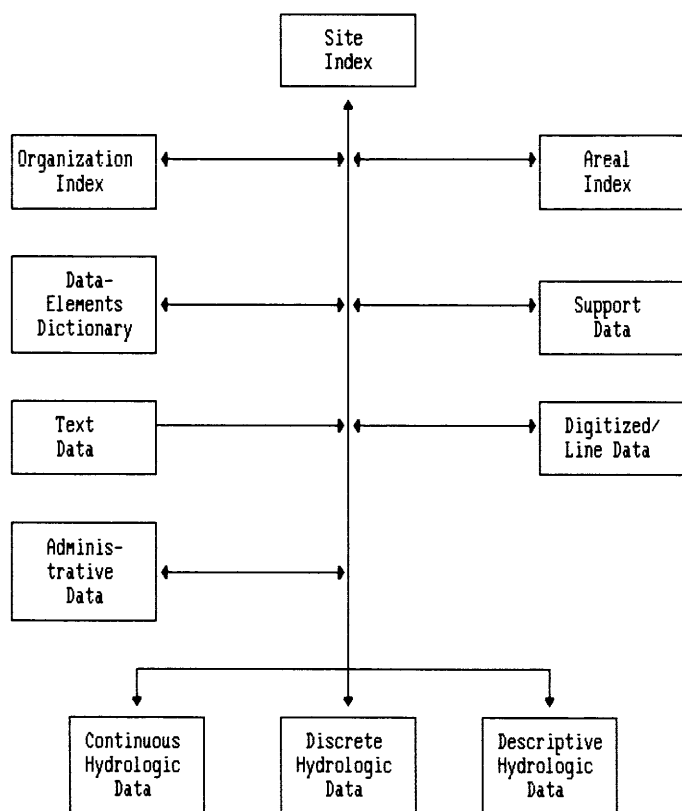
Figure 4. Simplified structure of the National Water Information System data base.

● Descriptive Hydrologic Data: This component will store hydrologic or hydrologic-related data values that are not time dependent. Examples of these data are basin characteristics and GW site-geologic characteristics and coefficients, equipment installed, names of owners, and other pertinent information.
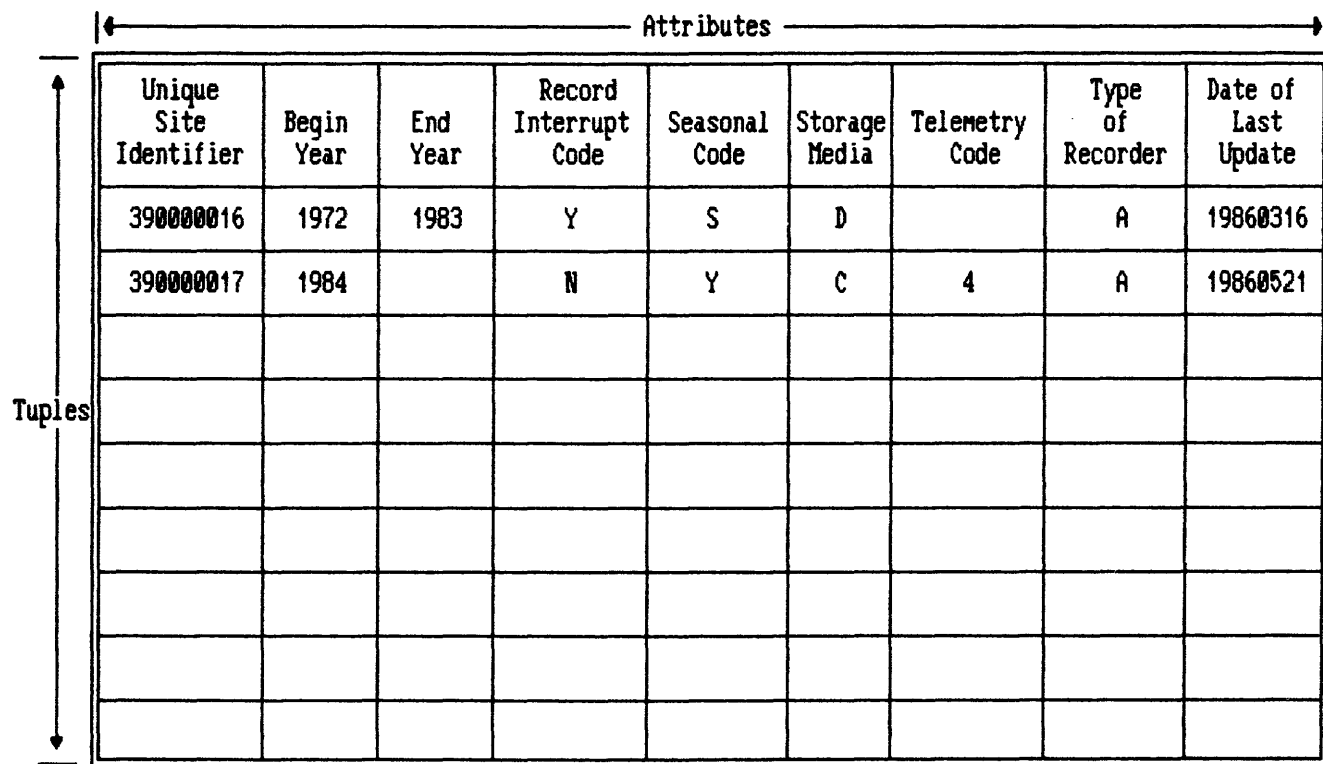
3.3.2.1. Data Storage

All data will be stored in the NWIS in an online environment. Data will be stored in the NWIS data base in the form of two-dimensional tables, technically referred to as relations (fig. 5). The data stored within each table are organized into rows and columns; a row is technically called a tuple, and a column is called an attribute. This form of data representation is similar in concept to the traditional sequential-file method of storage. The table can be compared to a data file, with its tuples corresponding to records within the file and its attributes corresponding to data elements within a record. In general, existing data-base-management systems will allow as many as 100 tables per data base, 250 attributes per table, and 150 million tuples per table. This mode of storage allows data to be grouped and managed in separate tables, allows the tables within the data base to be related to each other, and allow users of NWIS data to continue to logically view the stored data in much the same manner as in the current, diverse data files. As new data entities are identified for future storage,

new tables may be easily added to the data base, and additional attributes may be added to existing tables.

3.3.2.2. Data Relations

Relations between the data in tables in the NWIS data base are established by defining data attributes within the tables as keys. As many as 15 keys may be assigned within each table. There are two basic types of keys, primary and foreign: A primary key is an attribute or combination of attributes that uniquely identifies each tuple within a table; foreign keys are those attributes that appear as the primary key in one or more tables, but also appear as a nonprimary key in at least one other table. Both primary and foreign keys may be assigned within a table. For example, an attribute such as the site identifier may be defined as a primary key to uniquely identify each tuple in a table. Another data attribute, such as the date, may be assigned as a foreign key in the table, which, in turn, may be used as a primary key for uniquely identifying tuples in yet another table. The resulting tables are related by these two keys. Assigned keys may also be used to create indexes that identify the existence and positional storage of attributes in the data base, to increase the speed and efficiency of access to the stored data.

Two primary keys will be generated by the NWIS processing software and used to establish relations between tables in the data base: (1) unique site identifier and (2) cross-reference

	Unique Site Identifier	Begin Year	End Year	Record Interrupt Code	Seasonal Code	Storage Media	Telemetry Code	Type of Recorder	Date of Last Update
	390000016	1972	1983	Y	S	D		A	19860316
	390000017	1984		N	Y	C	4	A	19860521
Tuples									

Figure 5. Example of a relational data table. Headings shown are for illustrative purposes only.

identifier. These keys are discussed in the subsections that follow.

3.3.2.2.1. Unique Site Identifier

All data stored in the NWIS data base for a site will be identified and related by a unique site-identifer attribute. In the existing systems, each site is designated by a site identifier that may range in size from 8 to 15 characters; this identifier will continue to be used as the primary key for access by the user to data stored in the data base (fig. 6). Internally, however, all data will be identified, managed, and accessed by using a unique site identifier that will be computed and assigned by the site-indexing processing software each time a site is added to the data base. This identifier will consist of a nine-digit numeric value developed by concatenating a seven-digit sequence number to the two-digit numeric Federal
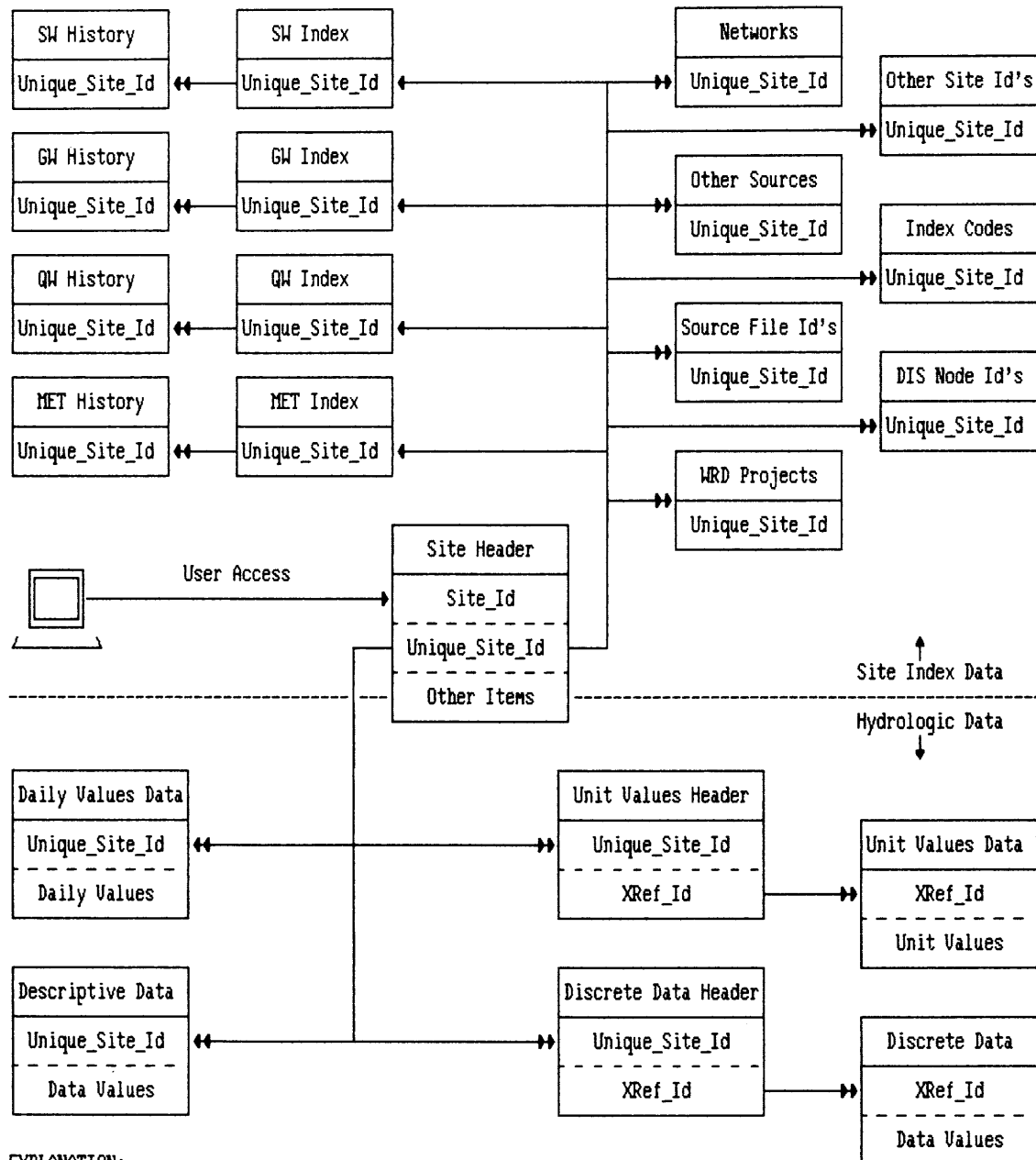
Figure 6. Simplified relational data model of the National Water Information System data base.

Table 1. Distribution of data in the National Water Information System.

[HIF, Hydrologic Instrumentation Facility; SMSA, Standard Metropolitan Statistical Area. Symbols: dots, all network data; circles, geographic areas of responsibility; solid squares, network summary; open squares, current data only; triangles, as needed]

Type of data	National	Regional	District	State	Sub-district	Central laboratory	HIF	Special application
Index data:								
Site index	●	○	○	○	○	▲	▲	▲
Organization index	●	○	○	○	○	▲	--	--
Areal index	●	○	○	○	○	▲	▲	▲
Data-elements dictionary	●	●	●	●	●	●	●	●
Support data:								
State/county codes	●	○	○	○	○	●	▲	▲
City/place codes	●	○	○	○	○	●	▲	▲
Country codes	●	▲	▲	▲	▲	▲	▲	▲
SMSA codes	●	○	○	○	○	--	▲	▲
District codes	●	○	○	○	○	●	▲	▲
Hydrologic-unit codes	●	○	○	○	○	●	▲	▲
Geologic-unit codes	●	○	○	○	○	●	▲	▲
Agency codes	●	●	●	●	●	●	▲	▲
Parameter codes	●	●	●	●	●	●	●	●
Index classification codes	●	●	●	●	●	●	▲	▲
Rating curves	--	--	○	○	○	--	▲	--
Datum/shift corrections	--	--	○	○	○	--	▲	--
Text data:								
Water-resources abstracts	●	--	--	--	--	--	--	--
Station headings	▲	▲	○	○	○	--	▲	▲
Digitized/line data:								
State/county boundaries	●	○	○	○	○	--	▲	▲
Hydrologic-unit boundaries	●	○	○	○	○	--	▲	▲
River traces	●	○	○	○	○	--	▲	▲
Satellite imagery	▲	▲	▲	▲	▲	--	▲	▲
Land use/land cover	▲	▲	▲	▲	▲	--	▲	▲
User generated	▲	▲	▲	▲	▲	--	▲	▲
Administrative data:								
User accounting data	○ ■	--	○	○	○	--	--	--
Computer charges	○ ■	○	○	○	○	--	○	○
User profiles	○ ■	○	○	○	○	--	○	○
Data-base statistics	○ ■	○	○	○	○	○	○	○
Addresses	▲	▲	▲	▲	▲	▲	▲	▲
Hydrologic data:								
Continuous data	▲	▲	○	○	○	--	▲	▲
Discrete data	▲	▲	○	○	○	□	▲	▲
Descriptive data	▲	▲	○	○	○	--	▲	▲

Information Processing Standards (FIPS) code that represents the State in which the site is physically located (U.S. National Bureau of Standards, 1970). All the tables in the data base except those requiring irregular time relations, will be related with this identifier (fig. 6). Figure 6 also shows that either a one-to-one (only one occurrence of the table in the data base) or a one-to-many (multiple occurrences of the table in the data base) relation can exist between tables.

The use of a unique site identifier simplifies the maintenance and management of data in the data base for the following reasons:

● When an agency-assigned site identifier is changed, a change is required in only one table in the data base.

● Duplicate site identifiers assigned by other agencies, for which data are indexed, can be supported.

● Because other agencies for which data are indexed frequently use nonnumeric site identifiers, these identifiers must be stored in the data base in character form, which requires 15 characters for each identifier. The unique site identifier is numeric in form and minimizes redundant storage of site keys by requiring only four bytes of storage for each occurrence in the data base.

3.3.2.2.2. Cross-Reference Identifier

Many of the data stored in the data base must be related to irregular time intervals, for example, discrete data, such as chemical analyses of water samples and unit-value continuous data. Such data may be stored for irregular date and time intervals within a date. For the storage of these data, two types of tables have been established in the data model: (1) Header Table and (2) Data Table. The Header Table is related to other data for the site via the unique site identifier (fig. 6); data attributes are also established in the Header Table that define the date and time intervals for the data to be stored. The data for each date and time interval are stored in a Data Table; an attribute designated as a cross-reference identifier (XREF) is used to relate data in the two tables to each other (fig. 6). Each tuple in the Header Table will contain a unique, arbitrarily assigned numeric XREF that will be computed by the hydrologic data-processing software; the XREF is repeated in the Data Table for all tuples containing data values for the related date and time interval. Figure 6 shows that there can be a one-to-many relation between the Header Table and the Site Header Table, and a one-to-many relation between the Header Table and the Data Table. Again, significant storage savings result from the use of the XREF because date values (year, month, and day) and time intervals (begin time and end time) need be stored only once for each set of discrete or continuous data associated with a date and time interval. The XREF is stored as a redundant data key and requires only four bytes of storage for each occurrence.

3.3.3. Data Distribution

The NWIS has been designed as a distributed data system. Under this concept, most of the basic hydrologic data collected by the WRD will be stored, managed, and disseminated by the district nodes of the NWIS network. These data may be further distributed to multiple computer nodes within each district or to State and subdistrict nodes, to better optimize data storage, utilization, and dissemination within districts. Research and special-project data will be collected and stored at regional and special-application nodes. The national node will store national indexes of all data stored in the distributed data base, national water-use data of the NWUIP program, and bibliographic citations and abstracts of the WRSIC program that are not amenable to geographic distribution. Each NWIS node will store data and maintain and operate only those software components of the system required to support that node's data-processing and application needs. The anticipated distribution of data in the NWIS is summarized in table 1.

3.3.4. Data Security

The maintenance and operation of the NWIS data base in a distributed environment requires an adequate program of security to assure the integrity of the distributed data. Adequate security is necessary to:

● Prevent unauthorized access to the data

● Restrict access to provisional data until validated for public release

● Restrict data entry and update functions to authorized personnel

● Allow the creation and maintenance of "private" data sets

Security will be provided for the NWIS by various techniques, including:

● User authorization codes and (or) account numbers assigned by site managers or system administrators

● Access codes assigned by data-base administrators at each node at the data-base entry and table levels

● User-assigned passwords at the tuple and attribute levels for both read and write access

User-assigned password authorities will be allowed for finite time periods to be established by the local data-base administrators. This procedure is necessary to prevent data from becoming permanently "locked" in the data base.

3.4. Software Architecture

In 1983, the existing data systems identified for inclusion in the NWIS used nearly 300 software packages consisting of more than 1,400 modules or subroutines and approximately 600,000 lines of inhouse-developed source code. In addition, at least 4 major data-base-management systems and 10 major proprietary software packages were used. Because the existing software was written to operate on large mainframe computers and was designed to fit the requirements of the 13 major files of the existing data bases, the software system for the NWIS is currently being redesigned. The new software system, however, must retain nearly all the functions of the old systems. Major objectives in the design of a software system for the NWIS were to:

● Provide both interactive and batch-processing capability

● Develop a menu-driven system that is simple to use

● Develop a modular system to minimize redundant source code and reduce the volume of source code

● Develop software packages that would be executable on microcomputers and easily transportable to different hardware

Ten basic functions were identified that must be supported by the NWIS software:

● Data entry and update

● Editing

● Retrieval

● Tabling

● Plotting

● Mapping

● Inventory and indexing

● General computations

● Statistical computations

● Maintenance utilities

These functions and the minimum software requirements developed during an analysis of the existing water-data systems of the WRD served as the basis for development of the conceptual design of the NWIS software architecture.

An overview of the NWIS software system is presented in figure 7. This software system will be developed and operated as a modular structure; many modules will be callable subroutines supporting common functions throughout the system. This modular structure will improve operating and maintenance efficiencies, the multiple use of software modules, and the integration of software functions. The multiple use of modules will also significantly reduce the amount of source code. The modular structure will also facilitate the distributed development of the software and the evolutionary development of the system through application prototyping.

Major features of the software will include:

● Interactive and batch-processing capabilities

● Integrated "HELP" facilities for users

● Online user documentation

● Integration of user-supplied software

3.4.1. Interactive Controller

Access to the NWIS software will be through an interactive controller (fig. 7). A family of user- and system-invoked menus, subroutines, and procedures used to perform various control, administrative, and housekeeping functions within the NWIS. This controller will provide a set of standardized, structured menus to assist users in selecting software and data within the NWIS for performing specific processing functions. These menus will provide information about data and software availability and guide the user through the selection process. The menus will be supported by internal prompts to assist the user in providing and selecting the information required to perform selected transactions. Additional user assistance will be provided through the use of prompts and tutorials that may be invoked by "HELP" commands at any time within a menu. Help will be available in the form of documentation explaining the functions, processes, and data options listed in the menu. This documentation will be provided in terse forms for more experienced users and in verbose forms for novice users. After execution of a "HELP" command, the user will be returned to the point of execution with no loss of position in the menu. Several methods may be used for selecting options within menus, including cursor movement, use of function keys, and key entry of option codes. Where complex or lengthy responses are required by a prompt, labeled screen forms may be used to simplify the process. The use of prompts and menus with their user-assistance capabilities will allow users with only minimal experience to access and use the NWIS. A command mode will be provided within the controller to allow more experienced users direct access to the software, with only minimal interaction with the controller.
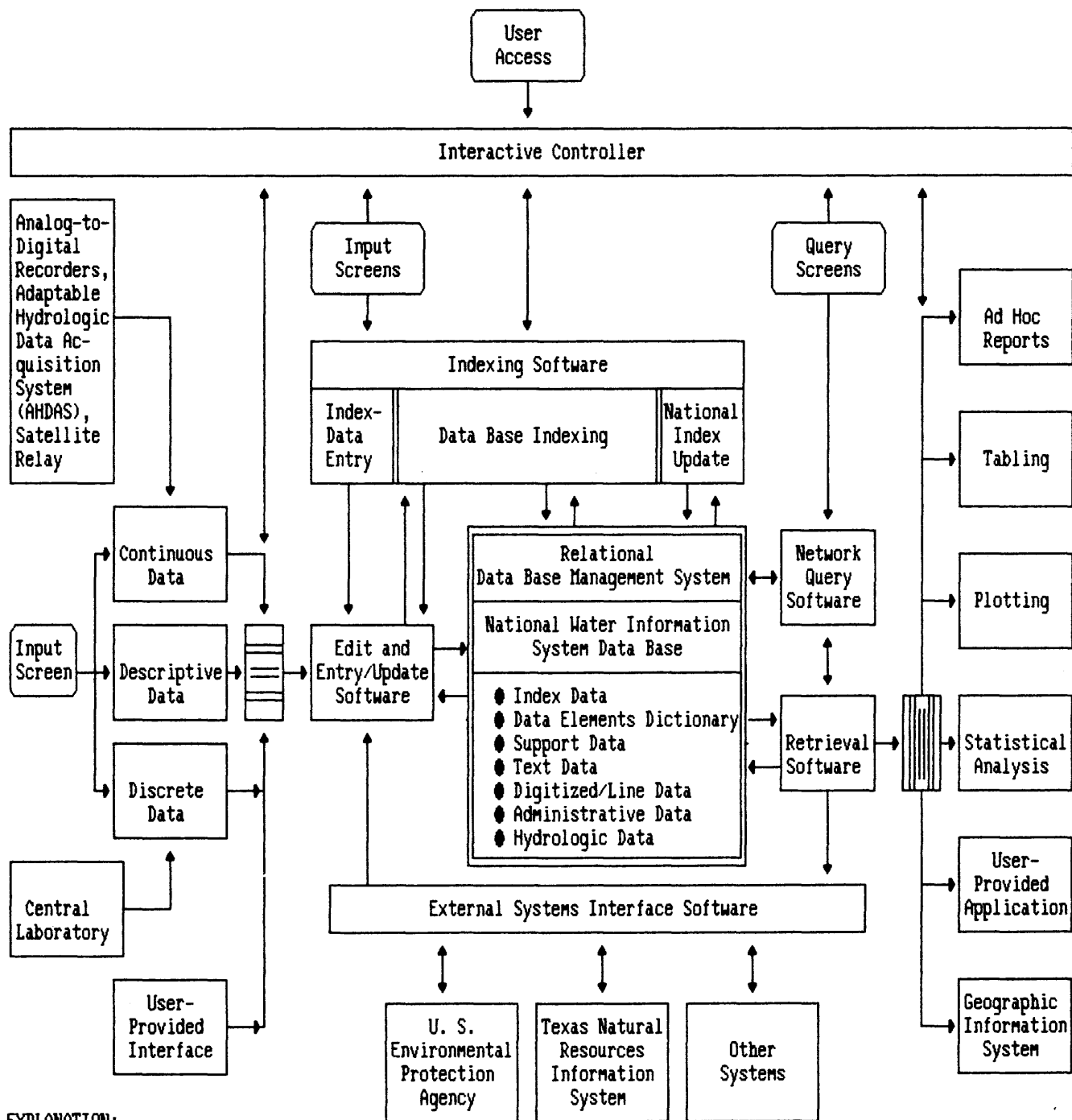
Figure 7. Overview of structure of the National Water Information System software.

The following major functions are proposed for the interactive controller:

- Monitoring user sessions
  — Validating user-access privileges
  — Recording the begin and end times of sessions
  — Developing and maintaining user profiles
  — Developing and maintaining NWIS-usage statistics
  — Transferring control to user-authorized NWIS-entry points

- Processing user selections
  — Providing menus for user selection of available software
  — Determining and invoking access paths to selected software
  — Transferring control to and from selected software

- System-administration functions
  — Maintaining user accounts
  — Maintaining user access privileges
  — Calculating user charges
  — Cataloging user-supplied software

- Data-base-administration functions
  — Backup/restore/recovery processes
  — Maintenance of data-base tables and attributes

- User services
  — Cataloging user-defined synonyms for NWIS software
  — Maintaining and cataloging user responses for future reuse
  — Maintaining and cataloging user-defined processes for sequential and (or) batch execution of multiple software modules

- User documentation and assistance
  — Maintaining an NWIS information desk
    □ Descriptions of NWIS subsystems
    □ Descriptions of NWIS applications and procedures
    □ Descriptions of NWIS software modules

- Maintaining and managing "HELP" requests
  — Explanations of error messages
  — Explanations of corrective actions to be taken
  — Brief and detailed user instructions
  — Code and mnemonic lists for value selections and assignment
  — Maintenance of the "HELP" message file

3.4.2. Data-Base-Management Systems

An RDBMS will serve as the central kernel of the NWIS software (fig. 7). All data in the NWIS will be stored and

managed by this RDBMS; the RDBMS will also provide software support to the NWIS. Various commercially available systems are currently being tested and evaluated for use in the NWIS. The RDBMS selected should have most, or all, of the following features:

- Full data-base management and maintenance capability, including load, backup, restore, recover, copy, and audit-trail facilities

- Extensive data-base-administration tools, including data-base design and development aids, a data-definition language, and a data-manipulation language

- Data-base security at the table, tuple, and attribute levels

- Full-screen-generation capability and screen-management facilities

- Menu selection of the RDBMS functions

- Callable interfaces with the Fortran and C programming languages

- Data-dictionary capabilities

- Data-edit facilities

- Interactive query language, with both update and retrieval capabilities

- Report generator

- Electronic-spreadsheet capabilities

- Indexing capabilities at the table, tuple, and attribute levels

- Plotting and graphics capabilities

- Minimal word-processing and text-editing capabilities

- Internal programming capability with either the available query language or a high-level programming language

- Code-generation or application-generation capability in the available query language or a high-level programming language

- Command-storage and cataloging capabilities

Current availability of most of the above features in commercial RDBMS's will contribute to a significant reduction in the amount of source code that must be developed in house.

3.4.3. Interface and Processing Software

The NWIS software will consist of various modules that will interface with the various types of data being entered for storage in the data base, including processing systems for data entered by the WRD's central laboratory; data collected by the AHDAS, analog-to-digital recorders (ADR's), and other automated instrumentation; and data relayed by satellite to DRGS's for processing and entry to the data base. Additional processing systems will be provided by using interactive screens for the manual key entry of data. All of these systems will provide data to the NWIS edit and entry/update software.

3.4.4. Edit and Entry/Update Software

All data to be stored in the NWIS data base will be processed by a common data edit and entry/update module (fig. 7) that will interact directly with the RDBMS and perform the following functions:

- Validation of data values by using edit criteria stored in the DED and support data tables, or by using special edit criteria and procedures included in the software

- Derivation and (or) computation of data values required before storage, by using derivation and computation algorithms stored in the DED

- Generation of site-indexing transactions for entry to the NWIS data base and (or) the NWIS indexing software

- Storage and (or) updating of data values in the NWIS data base

3.4.5. Indexing Software

The NWIS is designed to be self-indexing for data stored in the data base. Indexing software will consist of three major modules (fig. 7): (1) Index-Data Entry, (2) Data-Base Indexing, and (3) National-Index Update.

3.4.5.1. Index-Data Entry Software

The index-data-entry software performs the important function of assigning a unique site identifier to all sites in the data base. It will also process site-descriptor data and other site-index attributes manually entered into the data base before the initial processing of hydrologic data for a site. This software may also be used to manually update any index-data attributes previously stored in the data base, and will contain facilities for making global changes to index-data attributes stored in the data base.

3.4.5.2. Data-Base-Indexing Software

The data-base-indexing software will interact directly with the RDBMS. It will perform indexing transactions for all sites for which hydrologic data have been processed and stored since the last time the data-base-indexing software was executed. An audit trail generated by the edit and entry/update software will be used to search all applicable tables in the data base so as to generate indexing transactions for subsequent storage in the index tables of the data base. Also, the data-base-indexing software will be invoked to perform indexing functions that cannot be performed by using the indexing audit trail, such as determining whether or not data are still being collected for a site. At this time, summaries of data indexed in the data base will be computed for subsequent storage in its organization index. The data-base-query software will allow local data-base administrators to update local indexes as frequently as needed to assure that the indexing information remains current.

3.4.5.3. National-Index-Update Software

A national index will be maintained at the national node of the NWIS; this index will consist of a concatenation of all the local indexes maintained throughout the NWIS network, as well as information about water data that are available from other organizations throughout the country. The national index will serve as the central locator for determining the existence and computer-node location of data stored on the network. Because most indexing data will be generated at local nodes, these data must be periodically passed to the national index for updating and storage. The national-index-update software will be invoked at scheduled intervals to perform this function. The software will search the local index tables, determine which tuples are new or modified, and pass copies of these tuples to the national index for storage or replacement. Likewise, the software will be operated at the national node in a reverse function: Copies of tuples added to or updated in the national index from organizational sources that do not store data in the NWIS data base will be passed to local indexes.

3.4.6. Retrieval Software

All requests for data to be retrieved from the NWIS data base will be processed by a common retrieval software module (fig. 7) that will interact directly with the RDBMS and perform the following functions:

- Retrieval of specified tuples and attributes for tables in the data base, on the basis of criteria supplied by the user

- Derivation or computation of data values, as requested by the user, using derivation and computational algorithms stored in the DED

- Integration of data, as specified by the user, from specified tables in the NWIS data base

- Sorting and (or) ranking of the retrieved data in the order specified by the user

Output from the retrieval software will be provided in standardized formats for use by application software, or provided in machine-readable form to users for their personal use.

3.4.7. Network-Query Software

The national site index of the NWIS will serve as the central locator for determining the availability of data meeting specific user criteria and the computer-node location of the data in the NWIS distributed network. The network-query software will be operated at the national node of the NWIS and will perform the following functions:

- Search of the national site index to determine the existence and locations of requested data

- Generation of the identifiers and commands necessary to retrieve data from the NWIS data base

- Broadcast of requests to all applicable computer nodes in the NWIS distributed network to obtain the requested data

- Routing of the retrieved data to the requestors at their host computer node or to the national node for dissemination to the requestors

The network-query software will also be operated at district or other nodes that have redistributed data to multiple computer nodes within their operational jurisdiction.

3.4.8. General-Application Software

Various application-software modules will be provided in the NWIS architecture to support general, commonly used applications, including:

- Tabling

- Spreadsheets

- Plotting

- Mapping

- General computations

- Statistical analyses

- Numerical analyses

More specific application software is expected to be provided by the user community, as discussed below.

3.4.9. External-Systems-Interface Software

The NWIS software and data systems must interface directly with a wide array of data and software systems external to the NWIS. For example, it will need to exchange data with and receive indexing transactions from the Storage and Retrieval System (STORET) of the U.S. Environmental Protection Agency, the Texas Natural Information System (TNRIS), and many others. For external data and software systems that cannot use data generated in the standardized output formats of the NWIS retrieval software, many modules will be integrated into the NWIS software to perform the data and format transformations needed to achieve compatibility with those systems.

3.4.10. User-Supplied Software

A large amount of software required in the NWIS for specialized purposes and applications is expected to be developed and provided by the NWIS user community, including software systems and interfaces for the input and processing of special and unique types of data, and a wide array of special-application software, such as hydrologic models, special analytical procedures, geographic-information systems, and data-display applications customized to fit user needs. Upon completion of proper testing, documentation, and certification, user-supplied software may be integrated into the NWIS software. Thus, programs developed by researchers, district personnel, and others may be more widely shared with the NWIS user community. Likewise, software developed by other organizations that satisfy special requirements of the NWIS users can be integrated into the NWIS software.

3.4.11. Programming and Query Languages

Fortran 77 has been selected as the primary programming language for the NWIS software. Other programming languages, however, such as COBOL, C, Pascal, and BASIC, may be used to interface with proprietary software packages, to take advantage of the code-generation capabilities of data-base-management systems or other software packages, and to improve the computing efficiency of microcomputers and specialized hardware.

Because of the possible use of multiple data-base-management systems in the NWIS, it is highly desirable to have a

single query language that may be commonly used with all systems. SQL has been tentatively (1986) designated for this purpose. SQL is an English-like, high-level language designed for use by nonprogrammer personnel; experience to date has shown that it is easy to learn and use. Although versions of SQL may vary among vendors of the data-base-management systems selected for use in the NWIS, it is the query language most commonly used by these systems at the current time and is available on several of the data-base-management systems being tested and evaluated for use in the NWIS.

3.4.12. Design Data Dictionary

Software for the NWIS will be developed in a distributed manner by personnel located throughout the WRD. To assure adequate exchange of information pertinent to the NWIS among developers, a design-data dictionary will be maintained at the national node of the NWIS for common access by all development personnel. This dictionary will contain the following types of information:

● Data-element definitions (type, name, format)

● Data-base table and external file structures

● Record formats

● Input/output structures

● Screen formats

● Naming conventions (software, files, and data elements)

● Data and development standards

● Identification of existing programs and subroutines

The design-data dictionary will assure consistency in system terminology, naming conventions, data-base and data-element definitions, and a wide array of other items used in the daily design and development processes. It will also make developers more readily aware of previous work and thus reduce redundancy.

3.4.13. Software Documentation

Documentation for all the software subsystems and modules within the NWIS will be released as official U.S. Geological Survey publications. All subsystems will be documented as to functional requirements and conceptual design. Each software module will be documented by the following items:

● Program specifications used for development

● Test plan

● Test-analysis report

● User manual

● Operations manual

● Program-maintenance manual

FIPS Publication 38, "Guidelines for Documentation of Computer Programs and Automated Data Systems" (U.S. National Bureau of Standards, 1976), has been adopted by the U.S. Geological Survey as the standard for preparing this documentation. Copies of all design, specification, test, and maintenance documents will be stored at the national node of the NWIS and distributed as needed in the system design and development process. A comprehensive user and operations manual will be distributed to users of the NWIS. The various user and operations manuals will also be abstracted and stored online as part of the NWIS software in the forms of an NWIS information desk, prompts, and tutorials to be executed by "HELP" commands. These forms of online documentation will significantly reduce the need for users to work with the printed manuals.

3.4.14. Software Distribution

As discussed above in the subsection entitled "Nodes," it will not be necessary for each NWIS node to maintain a full configuration of the NWIS data-base and software systems. Instead, each node will acquire only the software necessary for its local operations. The components of the NWIS software (fig. 7) will be distributed in the following modules:

● The following components form the basic kernel of the NWIS software system and will be distributed to all NWIS computer nodes:
  — Interactive controller
  — Indexing software
  — RDBMS
  — Edit and entry/update software
  — Retrieval software

● The following components may be optionally distributed to NWIS computer nodes as requested:
  — Interfacing and processing software
  — General-application software
  — Network-query software
  — External-systems-interface software

All software modules will be fully tested and documented under the NWIS quality-assurance and configuration-

management program before they will be certified for release as part of the NWIS software. Once certified for release, modules will be integrated into the NWIS master software library maintained at the NWIS national node. Each optional module in the master library will be assigned identifiers for each computer node to which it is to be distributed. Nodes may request individual modules to be added or deleted from their distribution at any time. Descriptions of each software module will be maintained in the NWIS information desk of the national node for access by all users of the NWIS network to determine which modules may be useful to them.

All software distributions will be made from the national node of the NWIS. General distributions will be made to all nodes for new releases of the NWIS kernel software. New optional modules will be announced to all NWIS nodes at the time of their integration into the NWIS library. Nodes may optionally request to receive the announced modules. Distributions will be announced and made to all registered nodes each time an optional module is changed and certified for rerelease.

4. REFERENCES CITED

Datapro Research Corp., 1985, An introduction to local area networks: Delran, N.J., 16 p.

U.S. National Bureau of Standards, 1970, States and outlying areas of the United States: Federal Information Processing Standards Publication 5-1, 4 p.

—— 1976, Guidelines for documentation of computer programs and automated data systems: Federal Information Processing Standards Publication 38, 55 p.