CONCEPTUAL DESIGN FOR THE

NATIONAL WATER INFORMATION SYSTEM

U.S. GEOLOGICAL SURVEY
Open-File Report 86—604

CONCEPTUAL DESIGN FOR THE NATIONAL WATER INFORMATION SYSTEM

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

---

U. S. GEOLOGICAL SURVEY

Open-File Report 86-604

Reston, Virginia

1986

DEPARTMENT OF THE INTERIOR

DONALD PAUL HODEL, Secretary

U. S. GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

Assistant Chief Hydrologist for
Scientific Information Management
U. S. Geological Survey
440 National Center
Reston, Virginia 22092

Copies of this report can
be purchased from:

U. S. Geological Survey
Books and Open-File Reports Section
Federal Center
Box 25425
Denver, Colorado 80225

CONTENTS

CONTENTS -- Continued

ILLUSTRATIONS

CONCEPTUAL DESIGN FOR THE NATIONAL WATER INFORMATION INFORMATION SYSTEM

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

ABSTRACT

The Water Resources Division of the U.S. Geological Survey began the design and development of 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 has been designed as an interactive, distributed data system. The software system has been designed in a modular manner which integrates existing software functions and allows multiple use of software modules. The data base has been designed as a relational data model that allows integrated storage of the existing water data, water-use data, and water-data indexing information by using a common relational data base management system. The NWIS will be operated on microcomputers located in each of the Water Resources Division's District offices and many of its State, subdistrict, and field offices. The microcomputers will be linked together through a national telecommunication network maintained by the U. S. Geological Survey. The NWIS is scheduled to be placed in operation in 1990.

1. INTRODUCTION

In September 1982, the Water Resources Division of the Geological Survey began the installation of a nationwide, distributed network of minicomputers to be used for the storage, retrieval, analysis, and dissemination of its hydrologic data. This network, known as a Distributed Information System (DIS), is now (1986) fully operational and consists of nearly 70 minicomputers installed in the Division's District offices, State and subdistrict offices, regional offices, a central laboratory, the Hydrologic Instrumentation Facility, 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 Division. To fully use this technology, the Division 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) to implement the Division's water-related data and information systems as part of the DIS. The objective of the task force was to analyze, redesign, and integrate the existing water-data and information systems of the following programs:

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

- The National Water Data Exchange (NAWDEX) - An interagency program which indexes 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 the aggregation of water-use information in a National Water-Use Data System.

- The Water Resources Scientific Information Center (WRSIC) - A program for abstracting water-resource publications worldwide and 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 Division (Edwards, M. D., compiler, and others, written commun., 1983). This document updates the original conceptual design developed by the task force to include new concepts, design changes, and new technologies that have been identified or developed since 1983.

1.1 Objectives of the National Water Information System

The NWIS involves a complete redesign and modernization of the existing water-data systems that it will replace. Today's hardware and software technologies are changing and advancing very rapidly. Because of these rapid changes and advancements, it is expected that future software and data systems will have significantly shorter life cycles if these systems are to keep pace with modern technology. The NWIS must, therefore, be a highly flexible system that can be easily changed and expanded in a rapidly changing technological environment. On the basis of these general requirements and the requirements derived from the 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 Distributed Information System.

- 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 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 Division's data-handling needs, including those for research and special projects, allowing scientists to focus on scientific applications with the data.

- Provide a single flexible and expandable system that is easy to use, with fewer structures than the existing systems which will 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. This will 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 of the National Water Information System

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 Division. The major focus of the conceptual design of the NWIS is directed at the data systems of the WATSTORE, NAWDEX, NWUIP, and WRSIC programs. The system is also intended to include administrative subsystems required to operate and manage the NWIS.

1.3 Purpose of the Conceptual Design

The conceptual design is provided to present an overview of the system's objectives, scope, network architecture, data base content and structure, and hardware and software architectures. This will allow the user community to review and evaluate the proposed system for relevance and completeness, deficiencies, and needed enhancements prior to 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

Acknowledgment is given to the following members of the National Water Data System Task Force who developed the original requirements and concepts upon which the current conceptual design for the NWIS is based:

Melvin D. Edwards, U.S. Geological Survey, Water Resources Division
Isabelle Halley des Fontaines, U.S. Geological Survey, Water Resources Division
David V. Maddy, U.S. Geological Survey, Water Resources Division
Charles F. Merk, U.S. Geological Survey, Water Resources Division
James C. Poole, U.S. Geological Survey, Information Systems Division
Curtis L. Sanders, U.S. Geological Survey, Water Resources Division
Gale A. Cooley, The MITRE Corporation
Audrey L. Edler, The MITRE Corporation
Amr ElSawy, The MITRE Corporation

Ahmed Nadeem, The MITRE Corporation
Robert H. Stratman, The MITRE Corporation

2. OVERVIEW OF THE NATIONAL WATER INFORMATION SYSTEM CONCEPTUAL DESIGN

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

The NWIS is designed to operate on a wide range of computer hardware, including microcomputers. Thirty-two bit microcomputers will be evaluated for use as central processing units at each node for data base storage and management, and NWIS general-purpose computing. Less powerful 16-bit microcomputers or smaller 32-bit microcomputers will be used as work stations linked to the central processors via LAN's. A variety of 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 in excess of one gigabyte. Two types of mass-storage devices will be available for use: (1) Data base servers using magnetic disk technology for data with a high degree of volatility, 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. Data with special characteristics, such as text data, may be stored in other types of data base management systems to provide more efficient storage and use.

Thirteen major files and data bases of the existing data systems have been integrated into six discrete data components in the NWIS. In addition, five new data components have been added to the NWIS that do not exist in the current systems: (1) An Areal Data Index, (2) A Data Elements Dictionary, (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 types of support data)

- o Text Data (Station headings, bibliographic citations, water resources abstracts)

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

- o Administrative Data (User accounting, computer charges, data-base statistics)

- o Hydrologic Data

  -- Continuous Data (Unit values, daily values)
  -- Discrete Data (Quality-of-water (QW) analyses, peak flow, groundwater (GW) levels, and other data that are time dependent)
  -- Descriptive Data (GW site inventories, basin characteristics, and other characteristics that are not time dependent)

Most of the basic hydrologic data collected by the Division will be stored, managed, and disseminated by the District offices. Exceptions will be specialized research and special-project data collected and stored by regional and project offices. National Headquarters will store a national index of all data stored in the distributed network and also store 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 will also support batch-processing applications. The software system is a modular design to maximize sharing of software subroutines and procedures and minimize the development of redundant code.

Entry to the software will be through an interactive controller. The controller will consist of a family of system and user invoked menus, subroutines, and procedures to perform a variety of control and administrative functions. These will include the validation of 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 experienced users to access the NWIS software directly with minimal interaction with the interactive 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, and prompts and tutorials that can be invoked by the user at any time by the use of "HELP" commands.

The NWIS software system will be structured around relational and other data base management systems that will provide significant software support to the system in addition to providing full data base management and maintenance for the system. Software support to be provided by these systems include data base design aids, screen generation and management facilities, data base security, data dictionary capabilities, an interactive query language, report

generation, 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 the basic processing requirements of the system. These modules include data-indexing software for self-indexing of data stored in the NWIS, edit and data entry software for the common editing of data using criteria stored in the data elements dictionary, storage of data values in the data base, and generation of an audit trail to be used for data indexing purposes. Retrieval software which 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.

A variety of standardized interface and processing software will be provided to support the processing and entry of data from the WRD Central Laboratory and automated instrumentation systems. A number of general application packages will also be provided to perform general, commonly used applications including tabling, plotting, and statistical analysis procedures. The NWIS will also interface with a number of external data and software systems. This will be accomplished through modules provided to perform data and format transformations needed to achieve compatibility with the external systems.

Much of the special application software for the NWIS is expected to be provided by the users of the system. This will include software such as hydrologic models, Geographic Information System (GIS) applications, special analytical procedures, and data-display applications customized to fit user needs. In this manner, 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 system.

3. THE NATIONAL WATER INFORMATION SYSTEM CONCEPTUAL DESIGN

The conceptual design for the NWIS is presented in the subsections that follow. Discussions are presented on the network architecture, hardware configurations, data base architecture, and software architecture of the system.

3.1 Network Architecture

As shown in figure 1, the NWIS will operate within the GEONET system installed by the Geological Survey. GEONET is a private, packet-switching network that uses the X.25 telecommunication protocol within the TYMNET 1/ public network. Through this network, all offices of the Division and a large community of outside users will have access to the NWIS.

1/ - Use of brand and trade names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

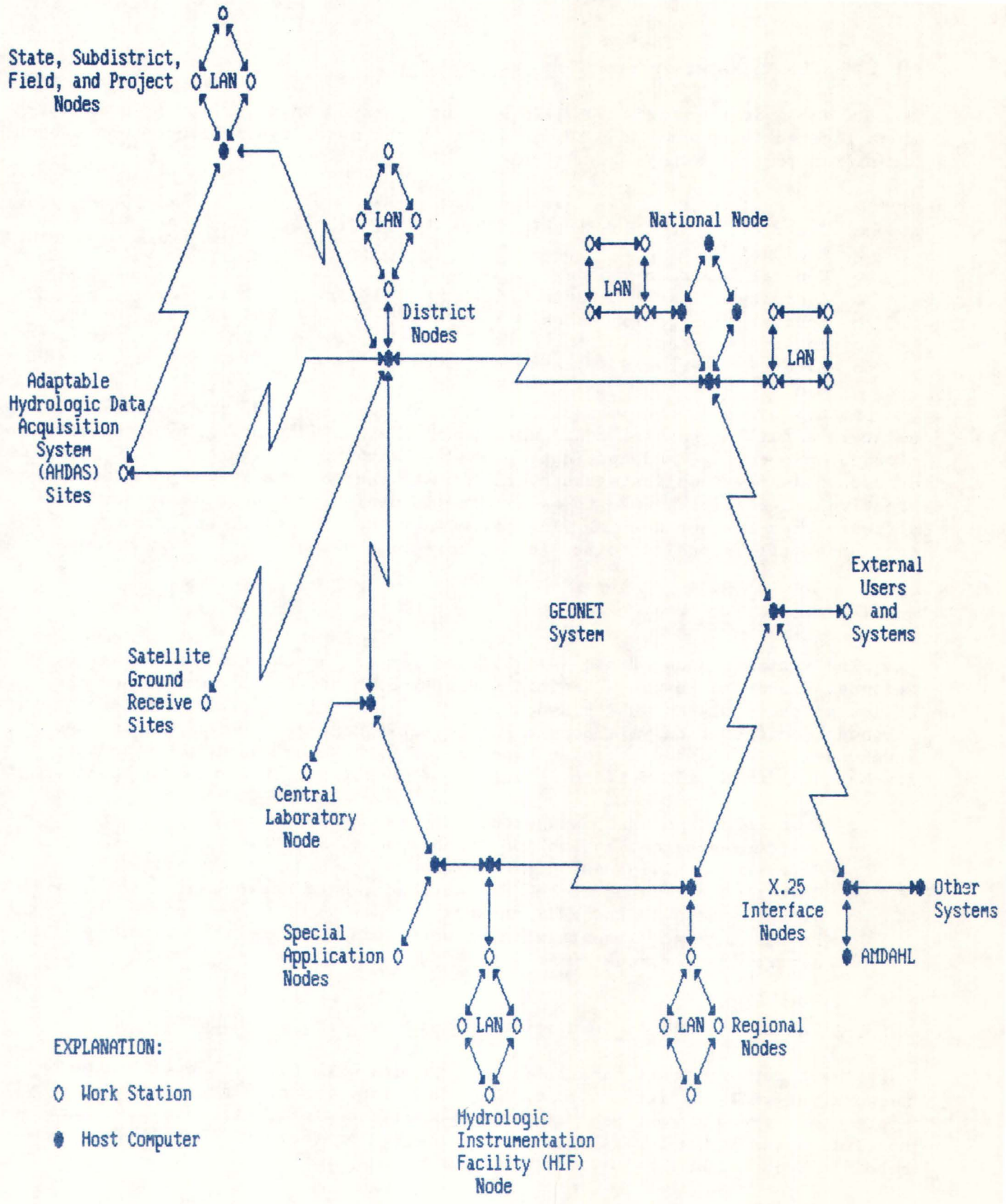


Figure 1.--Network architecture of the National Water Information System.

3.1.1 National Water Information System Nodes

As shown in figure 1, the NWIS will be operated on a variety of computer nodes located throughout the DIS network. Eight major types of nodes have been identified for the NWIS:

- o National Node
- o District, State, Subdistrict, Field, and Project Nodes
- o Regional Nodes
- o Central Laboratory Node
- o Hydrologic Instrumentation Facility (HIF) Node
- o Special Application Nodes
- o Interface Nodes
- o External Users and Systems

As discussed in the subsections that follow, each node may operate and maintain a full or partial configuration of the NWIS depending upon 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. In this manner, the hardware configuration 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 Geological Survey's National Center in Reston, Virginia and 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 which may be used independently or linked together to allow better use of available resources. The National node will perform a variety of functions, including the following:

- o Systems and network management
- o Systems research, evaluation, development, and testing
- o Software archiving and dissemination
- o Central indexing of the NWIS distributed-data network
- o Global access to the NWIS network
- o Processing and dissemination of water and water-related data requiring centralized storage

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

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

- o Nodes with NWIS data
- o Nodes without NWIS data

- Nodes with NWIS data that additionally support Direct Readout Ground Receive Stations (DRGS) for the receipt and processing of data relayed by satellite from remote locations and 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 have the responsibility for the processing, storage, indexing, management, and dissemination of 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 State or subdistrict node 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 system of their host node. These nodes will operate and maintain a partial configuration of the NWIS tailored to their needs which will allow them to receive and temporarily store data for local application and use.

Some District nodes will have the additional responsibility for receiving data transmitted by data collection platforms (DCP's) via satellite to Direct Readout Ground Receive Stations. Each receiving node will have the responsibility for disseminating these data to several designated District, State, and subdistrict nodes for processing and storage. Each receiving node will also have responsibility for providing backup support to at least one other receiving node.

All District nodes and many 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 Division. AHDAS is being developed to collect data at remote locations 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 which 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 Division's regional offices located in Reston, Virginia; Atlanta, Georgia; Denver (Lakewood), Colorado; and Menlo Park, California 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 research-related data and data collected by regional projects.

3.1.1.4 Central Laboratory Node

The Central Laboratory located in Denver (Arvada), Colorado processes water samples collected by Division offices nationwide and will serve as a major

source of data for the NWIS. The laboratory will maintain and operate a partial configuration of the NWIS sufficient to interface with its internal processing systems and temporarily store data prior to its transfer to other nodes for subsequent processing, validation, and storage.

3.1.1.5 Hydrologic Instrumentation Facility Node

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

3.1.1.6 Special Application Nodes

Several computer nodes are expected to exist which will be configured and dedicated to special applications. Examples of such nodes are those used for operating and maintaining geographic information system applications and those that may be dedicated to local or regional research projects and modelling 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 not compatible 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, rather, will allow access to the NWIS by incompatible external systems or access by the NWIS to incompatible external systems.

3.1.1.8 External Users and Systems

The Division currently (1986) provides access to its NAWDEX and WATSTORE data systems to over 170 organizations at over 260 remote locations nationwide. This service will continue to be provided by the NWIS. These organizations will not operate or maintain the NWIS system. Rather, they will access the NWIS for data-retrieval purposes. Most users will continue to be serviced by dial-up 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 network.

3.1.2 Local Area Networks

Use of microcomputers, word processors, graphics devices, laser printers,

and other hardware and software systems is rapidly increasing within the WRD. These systems can be most effectively used in a shared environment. Local area networks (LAN's) are expected to be used extensively at the NWIS nodes for the interoffice and intraoffice sharing of data, hardware, and software systems.

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 on the network. A LAN allows a large number of intelligent devices to share resources such 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 that will be discussed in subsequent sections of this report.

As shown in figure 1, each LAN will allow access to GEONET and any workstation on the LAN to communicate with computers and other devices throughout the NWIS network to share and exchange data and software.

3.2 Hardware Configurations

The original conceptual design for the NWIS developed in 1983 was directed at the operation of the NWIS on the existing network of minicomputers of the Distributed Information System (DIS). The NWIS has a projected development life cycle of five to seven years with nearly full implementation scheduled for 1990. The existing DIS minicomputer hardware has a projected life cycle of eight 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 and not the current DIS hardware. For this reason, it was decided to begin development of the NWIS 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 the 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.
- The prototyped NWIS will become progressively larger over time, allowing for accurate definition of 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 cycle. Selections of inadequate hardware will be avoided and hardware can be selected to fit the specific requirements of the NWIS software and data-base systems. This will prevent costly retro-

fitting and modifications to the developed system after the hardware procurement.

The first prototype of the NWIS is currently (1986) being developed using microcomputer workstations interfaced with the existing minicomputers. Subsequent prototypes will be developed using only microcomputers. By the end of the development 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 clearly defined.

As shown in figure 2, a variety of computer and peripheral hardware may be used in the operation of the NWIS. This hardware will be linked and shared over local area networks which will be a part of the DIS network. The various types of hardware that may be used are discussed in the subsections that follow.

3.2.1 Microcomputer Central Processors

One or more microcomputers will be the central processor(s) at each node of the NWIS computing network. 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 purposes. The following are desirable features for any single microcomputer to be tested and evaluated for this purpose:

Microcomputer:

- Full 32-bit word memory transfer
- Expandabe memory from 4 to 16 MBytes
- Connectable to selected LAN's
- Connectable to the X.25 GEONET network
- Support of up to 32 concurrent users
- Multi-tasking capability
- Distributed data base capability
- Shared data base capability

Peripheral Devices:

- Fixed disk storage of 500 MBytes to 2 GBytes
- Floppy disk storage
- Streamer and 9-track tape storage

Software Support:

- Relational data base management system
- UNIX or UNIX-based operating system
- High-level language compilers:
  - -- FORTRAN 77
  - -- C
  - -- COBOL
  - -- PASCAL
  - -- BASIC
- SQL language

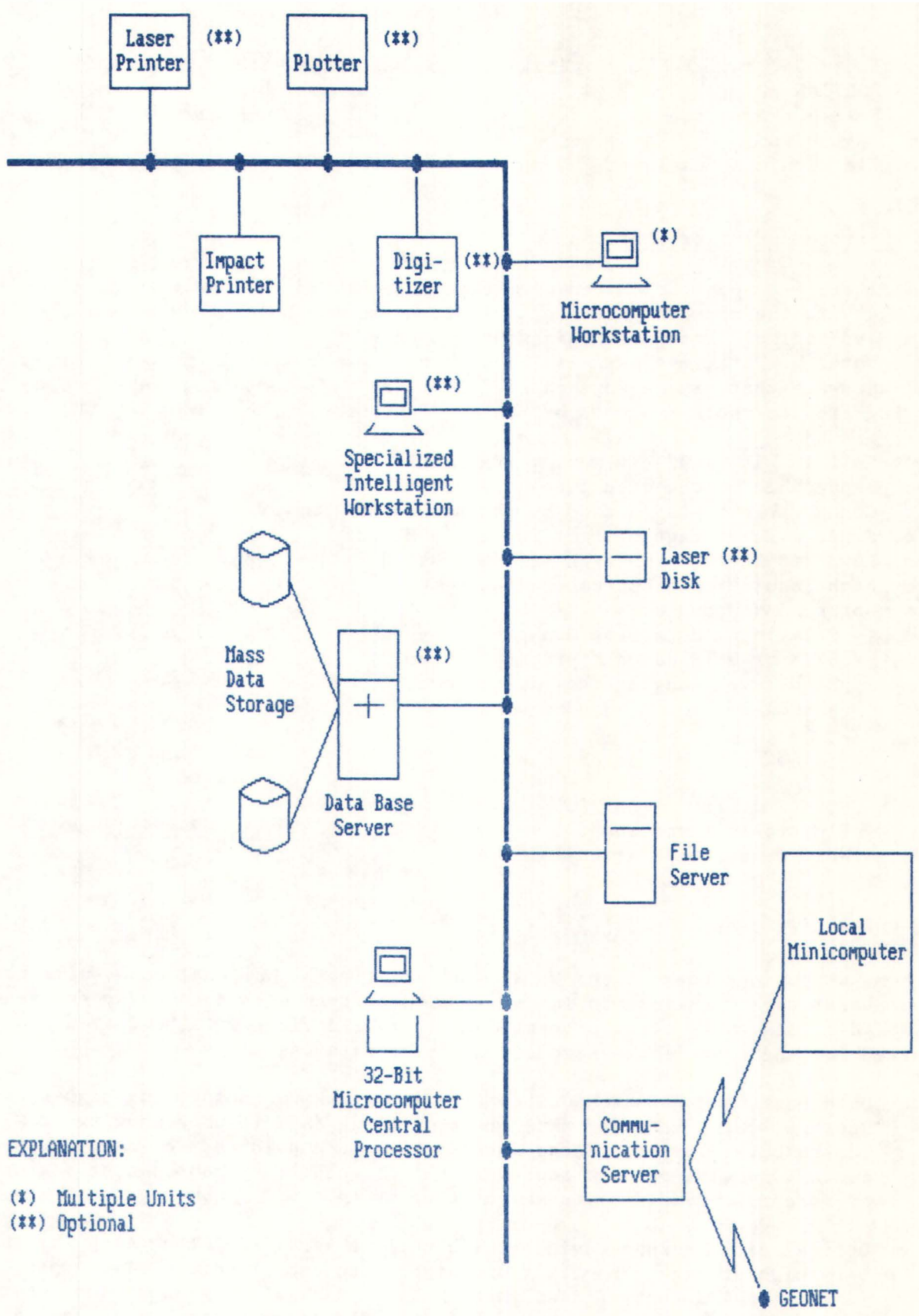


Figure 2.--Potential computer and peripherial hardware for use in operating the National Water Information System.

- o Index Sequential File Management (ISAM) utility with interface to FORTRAN 77
- o Full-screen editor
- o Text editor
- o Electronic spreadsheet

3.2.2 Microcomputer Workstations

Smaller, less-powerful microcomputers will be used as workstations interconnected 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 computing nodes. The following are desirable features of the microcomputer workstations:

- o Full 16-bit or 32-bit memory transfer
- o Expandable memory from 1 to 4 MBytes
- o Connectable to LAN's used by the Geological Survey
- o Fixed disk storage of from 20 to 100 MBytes
- o Low (360 KBytes) and high (1.2 MBytes) density floppy disk storage
- o High-resolution graphics monitor
- o Software support of:
  - -- Relational data base management system
  - -- UNIX or UNIX-based operating system
  - -- High-level language compilers for:
    - FORTRAN 77
    - C
    - PASCAL
    - BASIC
  - -- SQL Language
  - -- Electronic spreadsheet
  - -- Word processor/text editor

3.2.3 Mass Data Storage Devices

Some computing nodes of the NWIS, specifically the National node, will have large data-storage requirements in excess of one GByte with a need for frequent, high-speed access. Two types of mass-data storage devices will be tested and evaluated for the efficient management of large data bases:

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

- o Optical disk storage devices which utilize laser-disk technology having very high storage capacities and high-speed access rates. Two types exist: (1) Read-only drives which have no disk-writing capability where master disks are created and duplicated for distribution using specialized equipment; and (2) Write-once and 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 which require no, or infrequent, updating. Examples of these types of data are: support data such as State codes, county codes, 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 the NWIS network.

3.2.4 Peripheral Hardware

A variety of peripheral hardware will be required in support of the computer and software systems of the NWIS. This will include the following devices:

- Impact and laser printers with graphics capability
- Line (vector) and raster plotters
- Digitizers
- File servers which 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

Over six billion bytes (GBytes) of water data and related information are stored in the existing data systems of the Division. These data must be converted and transferred for distributed storage in the NWIS data base. Nearly an equal amount of additional water and water-related data have been identified in project and other special data systems throughout the Division which are desirable for conversion and storage in the NWIS data base. Major objectives in the design of a data base to support these data were to:

- Provide ease and quickness in accessing data
- Integrate the data, to the highest degree possible, into a single data base
- Minimize the redundancy of data storage
- Provide for the integrated use of all types of data stored in the data base

The designs and procedures developed to achieve these objectives are discussed in the subsections that follow.

3.3.1 Data Integration

The existing water-data and information systems of the Division consist of 13 major files and data bases managed by the NAWDEX, WATSTORE, NWUIP, and WRSIC programs. Each of these files or data bases are managed separately by the individual programs and have their own associated data management, processing, and

application software systems. As shown in figure 3, the design of the NWIS data base integrates the 13 major files and data bases into six discrete data-base components. 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 be managed separately but will be used in an integrated manner with common software systems. The Selected Water Resources Abstracts component will be managed under a separate data base management system to provide better to provide more efficient use of its text-type data.

As shown in figure 3, the following integrations of data have been accomplished in the NWIS data base design:

- 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 inintegrated into a single Site Index component.

- The data of 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 as a relational data model (Edler, A., The MITRE Corporation, written commun., 1986). As shown in figure 4, the data model consists of eleven major components, each of which contains data of different characteristics and usage requirements. The following are descriptions of these components:

- **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

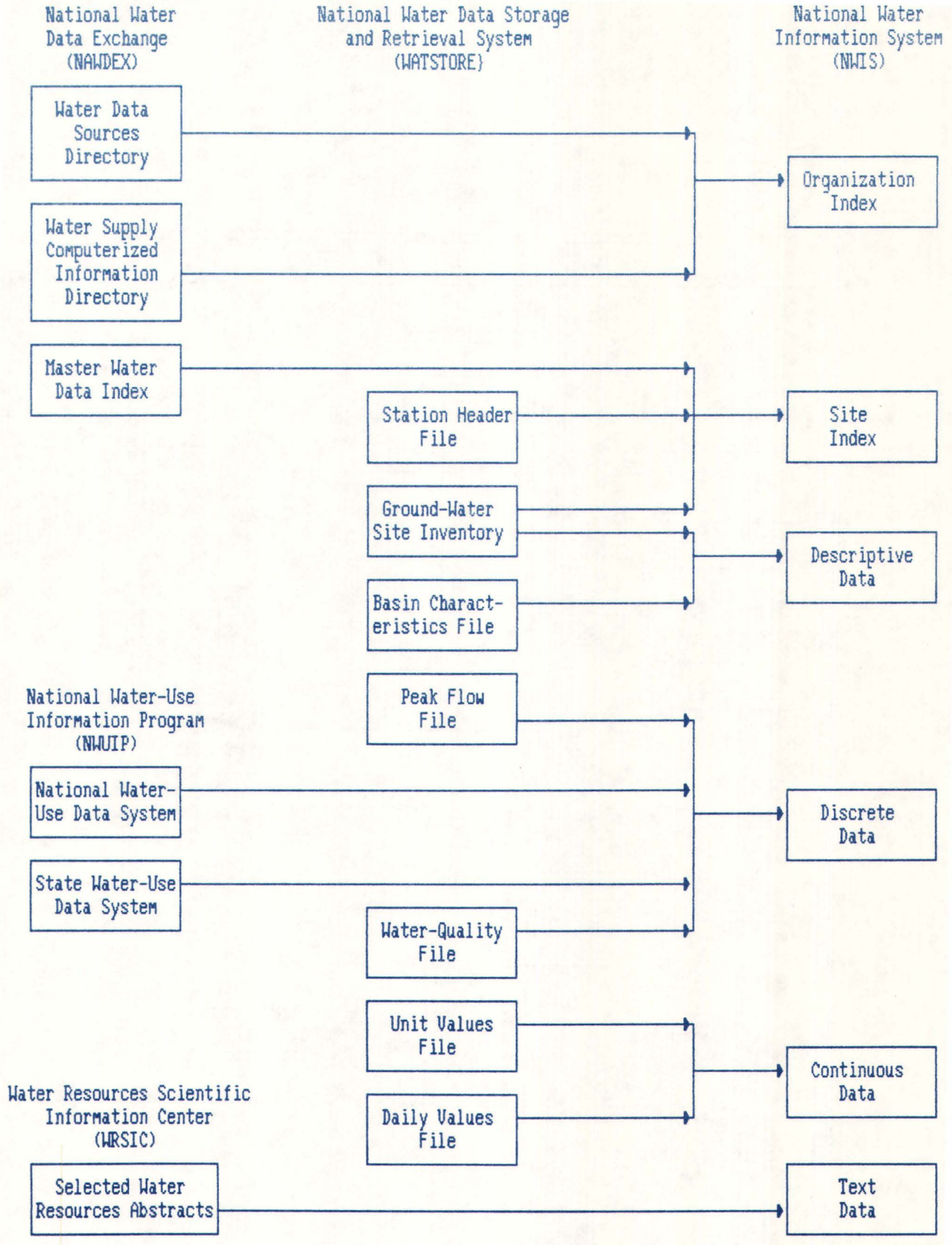


Figure 3.--Conversion of existing data systems to the National Water Information System.

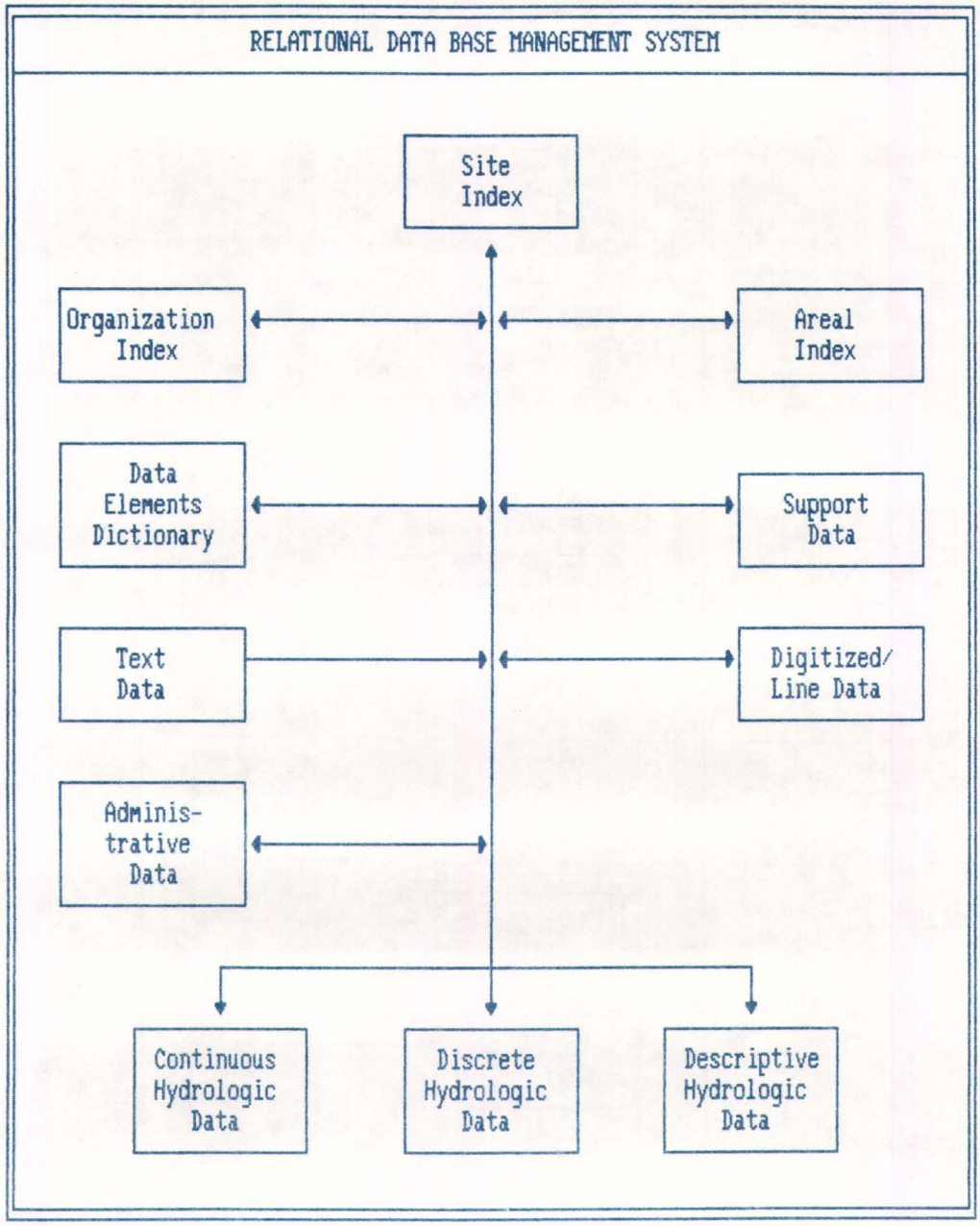


Figure 4.--Simplified structure of the National Water Information System data base.

through the NAWDEX program. This will include information about each site's geographic location, the types of data available, the frequency of measurement of the data available, the time periods for which data are available, the history of data collected at the site, and the data base or computer-node location of the data.

- o 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.

- o 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.

- o Data Elements Dictionary: The data base and software systems of the NWIS will be supported by a Data Elements Dictionary (DED) component of the data base. 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 data values associated with each data element.

- o Support Data: A variety of support data will be stored in the NWIS in conjunction with the DED. These data ininclude the following:

  Rating Data: These data include rating curves, calibration curves, and other relationships required to convert 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.

  Validation Codes: A number of 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 include the following:

-- State and county codes and names
-- City and place codes and names
-- Country codes and names
-- Standard Metropolitan Statistical Area 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: A variety of text-type data will be stored in the NWIS. These data will include online systems documentation and user instructions, descriptive headings for hydrologic sites, and abstracts and bibliographic citations of water-related literature.

- Digitized/Line Data: A variety of 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 derivations (State, county and hydrologic unit codes), data retrievals, data mapping, and a variety of other applications. The data will include the following:

  -- 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 a variety of mapping and other purposes

- Administrative Data: A variety of administrative data will be required to support the operation and management of the NWIS. This will include the following:

  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 use of data stored in the NWIS.

Addresses: Mailing addresses for data dissemination, user billing, and other activities related to the 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 quality of water parameters.

- Discrete Hydrologic Data: This component will store data collected or measured at irregular or fixed time intervals less frequent than daily (weekly, quarterly, monthly, annually, or other periodic intervals). These could be single-parameter data such as ground-water levels or multi-parameter 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 quality of water alert values.

- 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 ground-water 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. As shown in figure 5, data will be stored in the NWIS relational data base in the form of two-dimensional tables which are technically referred to as relations. (NOTE: The column headings shown in figure 5 are provided for clarification only and do not exist in the tables comprising the data base). Data stored within each table are organized in 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

Attributes

Tuples

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

Figure 5.--Example of a relational data table.

general, existing data base management systems will allow up to 100 tables per data base, 250 attributes per table, and up to 150,000,000 tuples per table. This mode of storage allows data to be grouped and managed in separate tables and allows tables within the data base to be related to each other. This will 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.

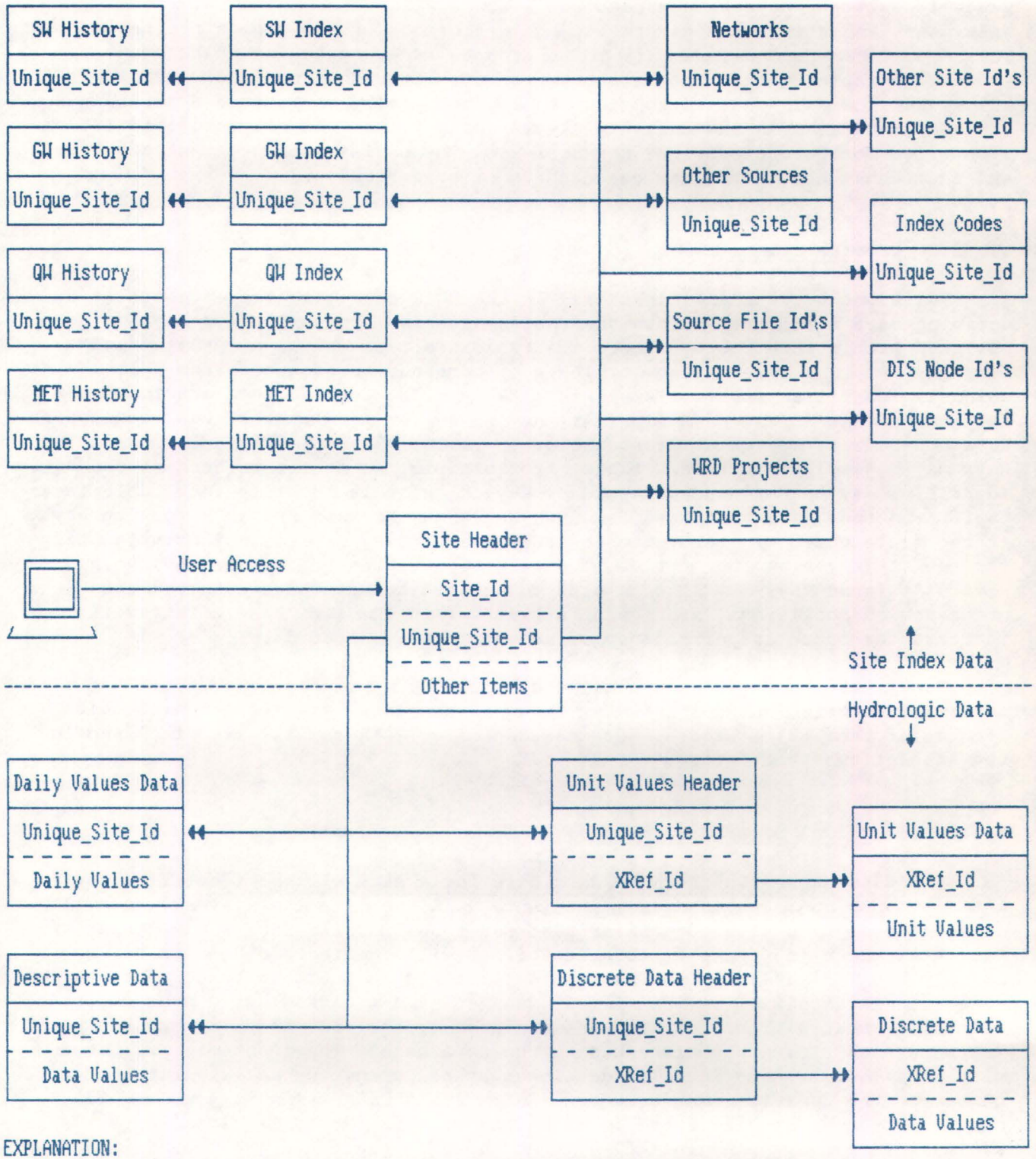
3.3.2.2 Data Relations

Relations between data in tables in the NWIS data base are established by defining data attributes within the tables as keys. Up to 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. The foreign keys are those attributes that appear as the primary key in one or more tables, but also appear as a non-primary 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 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 the two keys. Assigned keys may also be used to create indexes which identify the existence and positional storage of attributes in the data base. This will 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 relationships between tables in the data base: (1) A unique site identifier, and (2) A cross-reference 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 identified by a site identifier that may range in size from 8 to 15 characters. As shown in figure 6, this identifier will continue to be used as the primary key for access by the user to data stored in the data base. Internally, however, all data will be identified, managed, and accessed 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. The identifier will consist of a 9-digit numeric value developed by concatenating a 7-digit sequence number to the 2-digit numeric Federal Information Processing Standard (FIPS) code that represents the State in which the site is physically located (U.S. National Bureau of Standards, 1970). As shown in figure 6, all tables in the data base, with the exception of those requiring irregular time relations, will be related with this identifier. Figure 6 also shows that there can be a one-to-one (only one occurrence of the table can exist in the data base) or a one-to-many (multiple occurrences of the table can exist in the data base) relation between tables.



EXPLANATION:

→ One-to-One Relation

→→ One-to-Many Relation

The use of the 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.
- Since other agencies, for which data are indexed, frequently use non-numeric site identifiers, they 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 4 bytes of storage for each occurrence in the data base.

3.3.2.2.2 Cross-Reference Identifier

Much data stored in the data base must be related to irregular time intervals. Two specific examples of these data are discrete data, such as the chemical analyses of water samples and unit-values continuous data. In each case, data may be stored for irregular date intervals and irregular time intervals within a date. For the storage of these data, two types of tables have been established in the data model: (1) A Header Table, and (2) A Data Table. As shown in figure 6, the Header Table is related to other data for the site via the unique site identifier. 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. As also shown in figure 6, an attribute designated as a cross-reference identifier (XREF) is used to relate data in the two tables to each other. 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 relationship between the Header Table and the Site Header Table and a one-to-many relationship between the Header Table and the Data Table. Again, significant storage savings result in the use of the XREF since date values (year, month, and day) and time intervals (begin time and end time) need only be stored once for each set of discrete or continuous data associated with a date and time interval. The XREF is stored as the 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 Division 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 the District or to State and subdistrict nodes in order 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 which 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 which are not amenable to geographic distribution. Each node of the NWIS network will store data and maintain and operate only those software components of the NWIS required to support its data processing and application needs. A summary of the anticipated distribution of data in the NWIS is presented in figure 7.

3.3.4 Data Security

The maintenance and operation of the NWIS data in a distributed environment requires an adequate program of security to assure the integrity of the distributed data bases. 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 using a variety of techniques, including the following:

- 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 of the data base
- 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 over 1400 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 being redesigned. However, the new software system must retain nearly all of the functionality 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 that would minimize redundant source code and reduce the volume of source code

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	▲	▲	○	○	○		▲	▲

EXPLANATION:

● All Network Data
○ Geographic Area of Responsibility
▲ As Needed
■ Network Summary
□ Current Data Only

HIF Hydrologic Instrumentation Facility
SMSA Standard Metropolitan Statistical Area

Figure 7.--Distribution of data in the National Water Information System

- Develop software packages that would be executed 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 Division served as the basis for development of the conceptual design of the NWIS software architecture.

An overview of the structure of the NWIS software system is presented in figure 8. As shown in the figure, the software system will be developed and operated in a modular manner. Many modules will be callable subroutines supporting common functions existing in modules 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 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 include the following:

- Interactive and batch processing capabilities
- Integrated "HELP" facilities for users
- Online user documentation
- Integration of user-supplied software

The major software modules of the software system shown in figure 8 are discussed in the subsections that follow.

3.4.1 Interactive Controller

As shown in figure 8, access to the NWIS software system will be through an interactive controller. The controller is a family of user and system invoked menus, subroutines, and procedures used to perform a variety of control, administrative, and housekeeping functions within the NWIS.

The 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. The menus will provide information about data and

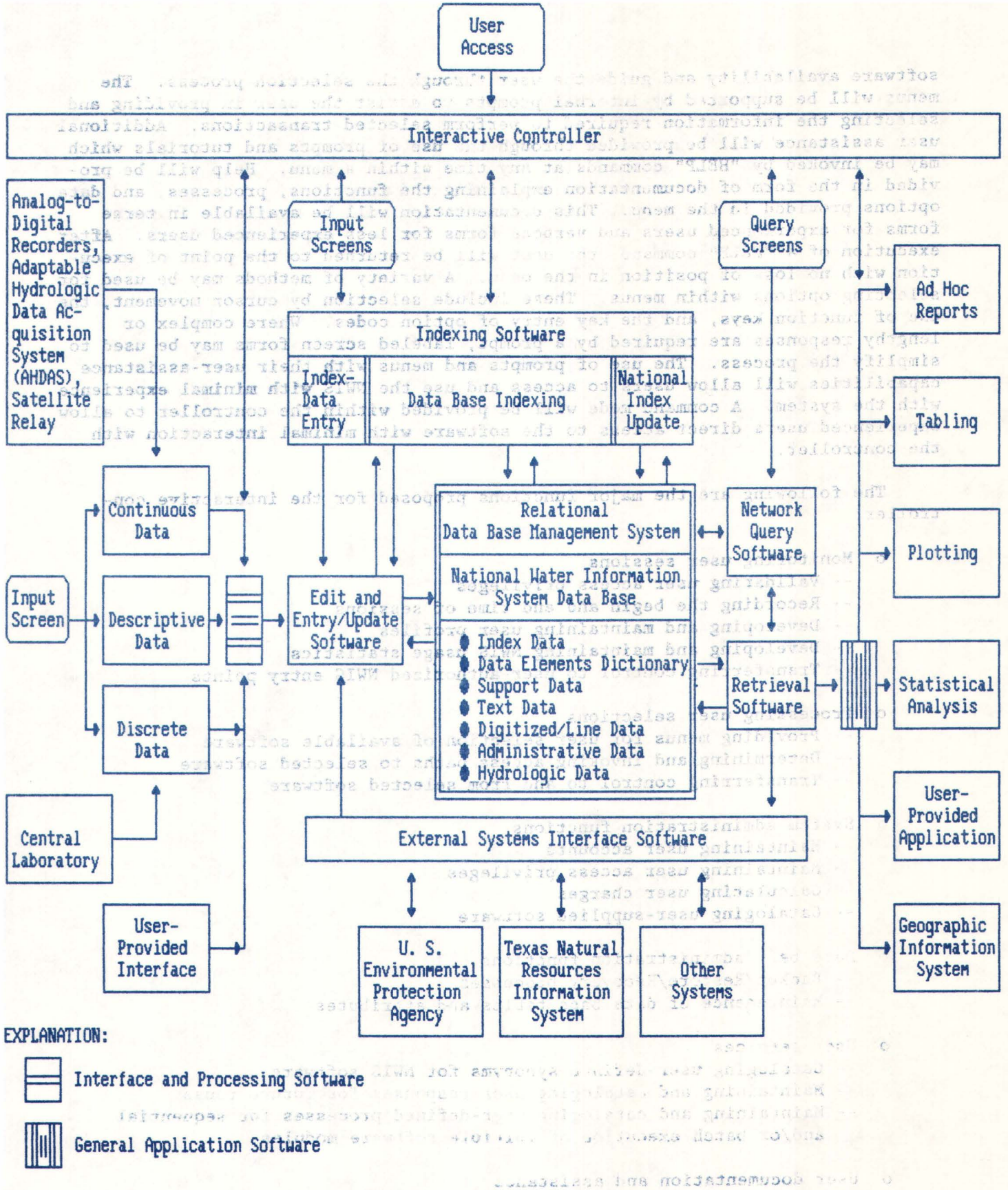


Figure 8.--Overview of the structure of the National Water Information System software system.

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 which may be invoked by "HELP" commands at any time within a menu. Help will be provided in the form of documentation explaining the functions, processes, and data options provided in the menu. This documentation will be available in terse forms for experienced users and verbose forms for less experienced 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. A variety of methods may be used for selecting options within menus. These include selection by cursor movement, the use of function keys, and the 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 to access and use the NWIS with minimal experience with the system. A command mode will be provided within the controller to allow experienced users direct access to the software with minimal interaction with the controller.

The following are the major functions proposed for the interactive controller:

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

- o 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

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

- o Data base administration functions
  - -- Backup/Restore/Recovery processes
  - -- Maintenance of data base tables and attributes

- o 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

- o User documentation and assistance
  - -- Maintaining an NWIS information desk
    - Descriptions of NWIS subsystems

- Descriptions of NWIS applications and procedures
- Descriptions of NWIS software modules

o 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

As previously shown in figure 8, a relational data base management system (RDBMS) will serve as the central kernel of the NWIS software system. All data in the NWIS will be stored and managed by the RDBMS. The RDBMS will also provide software support to the NWIS. A variety of commercially available systems are currently (1986) being tested and evaluated for use in the NWIS. The RDBMS selected should have most, or all, of the following features:

- o Full data base management and maintenance capability including load, backup, restore, recover, copy, and audit trail capabilities.
- o Extensive data base administration tools including data base design and development aids, a data definition language, and a data manipulation language.
- o Data base security at the table, tuple, and attribute levels.
- o Full screen generation capability and screen management facilities.
- o Menu selection of the RDBMS facilities and capabilities.
- o Callable interfaces with the FORTRAN and C programming languages.
- o Data Dictionary capabilities.
- o Data edit facilities.
- o Interactive query language with both update and retrieval capabilities.
- o Report generator.
- o Electronic spreadsheet capabilities.
- o Indexing capabilities at the table, tuple, and attribute levels.
- o Plotting and graphics capabilities.
- o Minimal word-processing and text-editing capability.
- o Internal programming capability with either the available query language or a high-level programming language.
- o Code generation or application generation capability in the available query language or a high-level programming language.
- o Command storage and cataloging capability.

Most of the above features are currently available in relational data base management systems and will contribute to a significant reduction in the amount of source code that must be developed inhouse.

3.4.3 Interface and Processing Software

The NWIS software system will contain a variety of software modules that will interface with the various types of data being entered for storage in the

data base. These will include processing systems for data entered by the WRD Central Laboratory; data collected by the Adaptable Hydrologic Data Acquisition System (AHDAS), analog-to-digital recorders (ADR's), and other automated instrumentation; and data relayed by satellite to Direct Readout Ground Receive Stations (DRGS's) for processing and entry to the data base. Additional processing systems will be provided 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, as shown in figure 8. This module will interact directly with the relational data base management system and perform the following functions:

- Validation of data values using edit criteria stored in the NWIS Data Elements Dictionary (DED) and support data tables or using special edit criteria and procedures included in the software
- Derivation and/or computation of data values required prior to storage 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. As shown in figure 8, indexing software for the NWIS will consist of three major modules: (1) Index-Data Entry, (2) Data Base Indexing, and (3) National Index Update. The functions of these three modules are discussed in the subsections that follow.

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 prior to 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. It will also 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 relational data base management system. 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 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 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 the organization index of the data base. 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 which will consist of a concatenation of all local indexes maintained throughout the NWIS network as well as information about water data that are available from other organizations throughout the country. This national index will serve as the central locator for determining the existence and computer node location of data stored throughout the network. Since 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, this software will be operated at the national node in a reverse function. Copies of tuples added 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, as shown in figure 8. This software will interact directly with the relational data base management system and will perform the following functions:

- Retrieval of specified tuples and attributes for tables in the data base based upon criteria supplied by the user.
- Derivation or computation of data values, as requested by the user, using derivation and computational algorithms stored in the Data Elements Dictionary.
- Integration of data, as specified by the user, from specified tables in the 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 to users in machine-readable form 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 requests to all applicable computer nodes in the NWIS distributed network to obtain the requested data.
- Route 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

A variety of application software modules will be provided in the NWIS architecture to support general, commonly used applications. Software will be provided to perform the following general applications:

- Tabling
- Spreadsheets
- Plotting
- Mapping
- General computations
- Statistical analyses
- Numerical analyses

More specific application software is expected to be provided by the user community and is discussed in subsection 3.4.10.

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. Examples of this are the 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) of the State of Texas, and many others. For those systems that cannot use data generated in the standardized output formats of the NWIS retrieval software, a large number of software modules will be integrated into the NWIS software system to perform the data and format transformations needed to achieve compatibility with external data and software 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. This will include 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 system. In this manner, software 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 system.

3.4.11 Programming and Query Languages

FORTRAN 77 has been selected as the primary programming language for the NWIS software. Other programming languages such as COBOL, C, PASCAL, BASIC, and others may, however, be used to interface with proprietary software packages, to take advantage of 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 possibility of using 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. The Structure Query Language (SQL) has been tentatively (1986) selected for use as the standard query language of the NWIS. SQL is an English-like, high-level language that is designed for use by non-programmer personnel. Experience, to date, has shown the language to be easy to learn and use. Versions of SQL may vary among vendors of data base management systems selected for use in the NWIS. However, it is the query language most commonly used by these systems at the current time and is available on several of the 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 Division. To assure adequate exchange of information pertinent to the development of 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. The dictionary will contain the following types of

information:

- o Data element definitions (type, name, format)
- o Data base table and external file structures
- o Record formats
- o Input/output structures
- o Screen formats
- o Naming conventions (software, files, data elements)
- o Data and development standards
- o Identification of existing programs and subroutines

The design data dictionary will assure consistency in systems 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 reduce redundancy.

3.4.13 Software Documentation

All software subsystems and modules within the NWIS will be documented and published as official U.S. Geological Survey publications. All subsystems will be documented in functional requirements and conceptual design documents. Each software module will be documented by the following:

- o Program specifications used for development
- o Test plan
- o Test analysis report
- o User manual
- o Operations manual
- o Program maintenance manual

Federal Information Processing Standards Publication 38, "Guidelines for Documentation of Computer Programs and Automated Data Systems" issued by the National Bureau of Standards on February 15, 1976, has been adopted by the U.S. Geological Survey as the standard for preparing 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 user and operations manual will be distributed to users of the NWIS. Also, the user and operations manuals will be abstracted and stored online as part of the NWIS software system in the forms of an NWIS information desk, prompts, and tutorials to be executed by "HELP" commands (See subsection 3.4.1). These forms of online documentation will significantly reduce the need for users to work with printed manuals.

3.4.14 Software Distribution

As previously discussed in subsection 3.1.1, it will not be necessary for each computer node of the NWIS network to maintain a full configuration of the NWIS data base and software systems. Each node will acquire only the software necessary for its local operations. The components of the NWIS software system, previously shown in figure 8, will be distributed in the following manner:

- 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
  - -- Relational Data Base Management System
  - -- 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 system. Once certified for release, modules will be integrated in 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 in 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 Corporation, 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.