United States Department of the Interior

GEOLOGICAL SURVEY
RESTON, VA 22092

NOV 21 1991

In Reply Refer To:
Mail Stop 437

Mr. Donald L. Chery, Jr.
Nuclear Regulatory Commission
Washington, D.C. 20555

Dear Mr. Chery:

The U.S. Geological Survey, Water Resources Division, is modernizing its hydrologic information system by designing and developing a new data base management system to replace the current Water Data Storage and Retrieval System (WATSTORE) and other related systems. The current system is used by many governmental and private organizations as a primary source of data for research and water resources planning and decisionmaking. The new National Water Information System (NWIS-II) will be a single, integrated system for processing and storing hydrologic data on a nationwide network of 32-bit micro computers. The design and development effort, which is being coordinated with the STORET program at the U.S. Environmental Protection Agency, began in 1989, and the first release of software is scheduled for October 1992.

This letter is to invite your review of the requirements of NWIS-II by reviewing and providing written comments on the enclosed paper, "National Water Information System II." We request your assistance in assuring that the review of the paper is coordinated throughout the Nuclear Regulatory Commission as needed to reflect possible impacts on your mission and systems. This review process is the mechanism through which your agency's formal comments on the planned improvements will be obtained. Your response is requested by February 1, 1992.

The enclosed paper is a very condensed version of the "NWIS-II System Requirements Specifications," a 750-page technical document that describes the specific requirements of the system. The paper emphasizes the requirements for data retrieval, output formats, and indexing because these will most affect outside users. Please address these requirements and the following questions during your review:

1. Are the proposed retrieval requirements adequate to meet the needs of your agency?

9111260285 911121
PDR WASTE
WM-1
PDR

2. Do you have any requirements for outputting data in formats other than those specified? If so, please describe the format.

3. Will the proposed data indexing system be useful to your agency? Would you use the system and provide index data?

If you have any questions about our request or need additional information about the NWIS-II requirements, please contact Tom Yorke, Chief, National Water Information System Program (703) 648-5659 or FTS 959-5659.

Sincerely yours,

James F. Daniel

James F. Daniel
Chairman, Subcommittee on Water Data
and Information Exchange

Enclosure

NATIONAL WATER INFORMATION SYSTEM II

INTRODUCTION

The National Water Information System Program of the U.S. Geological Survey (USGS), Water Resources Division, is designing and developing a new computer software system for processing and storing hydrologic data. The goal of the National Water Information System II (NWIS-II) effort is to implement a highly flexible hydrologic data management and processing system; one that can be easily changed and expanded in a rapidly changing technological environment. NWIS-II will be a single, integrated system that will include the functions of the current national Water Data Storage and Retrieval System (WATSTORE), the National Water Data Exchange (NAWDEX), and the NWIS-I systems, plus expanded capability for processing and managing additional chemical constituent, sediment, biological, and ancillary spatial data useful for GIS applications. The data base and processing system will be distributed across the Nation on the Distribution Information System II (DIS-II), a network of 32-bit microcomputers. It will include a national index of all data on the system and automatic polling for retrieving data from individual or multiple nodes.

The purpose of this paper is to provide information about NWIS-II to other Federal agencies, State and local agencies, and other interested groups, and to solicit comments on access and retrieval requirements for outside users of USGS hydrologic data bases. This paper will include a brief review of current systems, an overview of the development effort, the general requirements proposed by USGS users, and specific requirements proposed for data retrieval, output, and indexing. Considerable details on retrieval, output, and indexing are presented to facilitate comments on those aspects that most affect outside users.

CURRENT SYSTEMS

The current hydrologic data management and processing capability of the USGS consists of three separate software systems maintained on both a mainframe computer and a network of minicomputers. WATSTORE and NAWDEX are housed on a mainframe computer and NWIS-I is housed on a network of 54 minicomputers distributed across the Nation. WATSTORE consists of four major files and other ancillary files for storing surface-water, ground-water, and water-quality data. The Daily Values File, which contains historical data measured daily or continuously, has information from more than 32,000 sites. Additional data are being added for more than 12,000 sites each year. The Peak Flow File contains historical annual discharge values for more than 23,000 streamflow stations; 10,000 sites currently are in operation. The Water Quality File contains analyses for more than 300,000 sites, including 9,900 that currently are in operation. The Ground-Water Site Inventory contains geologic and well inventory data for more than 1,000,000 sites.

The NAWDEX program was established in 1976 as a confederation of organizations working together to make water data more readily accessible. Currently, NAWDEX comprises about 350

water-oriented organizations and consists of two files: (1) a Water Data Sources Directory of organizations that collect water data, the types of data they collect, and locations within these organizations from which water data may be obtained, and (2) a Master Water Data Index of sites for which water data are available. The Water Data Sources Directory contains information on 1,800 organizations. The Master Water Data Index identifies 480,000 sites at which water data are available. Information available about each site includes the geographic location, the types of data, and the period for which data are available.

NWIS-I consists of five files containing most of the data stored in WATSTORE. The functionality of the two systems is similar with one major exception. NWIS-I is a distributed interactive system designed to give each District office quicker response time for the processing and storage of the data collected in the field, whereas WATSTORE is a centralized batch processing system remote from the District offices. The 12,000 sites for which data are added each year to the WATSTORE Daily Values File are first processed and stored by utilizing the NWIS-I Daily Values File. Similarly, all new information is first processed and stored utilizing the NWIS-I Ground-Water Site Inventory and Water Quality File prior to being transmitted to WATSTORE. The State Water-Use File in NWIS-I contains site-specific water-use information and does not have a comparable file in WATSTORE. In addition, the NWIS-I Unit Values File contains the data collected more frequently than daily (e.g., every 5 minutes, 15 minutes, 1 hour) at 12,000 sites yearly.

SYSTEM DEVELOPMENT

The design and development of NWIS-II has required the coordination and combined efforts of five groups in the Water Resources Division. These groups, all critical to the effort, include the Strategic Planning Group, the User Groups, the Quality Assurance and Configuration Management Unit, the Design and Development Team, and the Data Management and Operations Unit. The discussion of the NWIS-II design and development effort focuses on a typical software life cycle. Each phase or step in the life cycle from initiation through operations and maintenance is described, including the role of each of the five groups involved in the effort.

Development of NWIS-II was initiated after a series of internal studies identified the shortcomings of the 20-year-old WATSTORE and the need for an integrated system. NWIS-II started with the formation of the Strategic Planning Group (SPG) in 1988. The SPG consists of most of the senior managers of the Water Resources Division, including the five Assistant Chief Hydrologists and the four Regional Hydrologists. The purpose of the SPG is to address policy issues and to aid in the implementation of NWIS-II. They are responsible for determining the scope of the effort and approving or delegating the approval of each phase of the software life cycle. They have committed the USGS to a complete redesign and rewrite of software instead of converting and updating existing software. They also are committed to accomplishing the effort with existing staff instead of using consulting services. As a result of this commitment, the other four groups involved in the design and development effort were formed, including 50 people to serve on User Groups and 40 people to serve on the Quality Assurance and Configuration Management Unit, the Design and Development Team, and the Data Management and Operations Unit.

The User Groups, which were formed in January 1989, consist of a cross section of WRD scientists currently using the WATSTORE, NAWDEX, and NWIS-I systems for processing and storing hydrologic data, and others who will be users of the new system. Forty-seven individuals were selected to serve on the user groups; 35 are data managers or investigators from WRD District offices and the others are management, headquarters, or research personnel. Eight User Groups represent specific disciplines or functional areas, which include the traditional WRD disciplines of surface water, ground water, quality of water, sediment, and water use. Also represented are groups for biology and spatial data in recognition of their importance to any future water programs. Finally, there is a NAWDEX group to address the non-USGS users of our data bases.

The purpose of the User Groups is to describe the user needs, including specific capabilities of inputting, computing, storing, and retrieving all forms of hydrologic and ancillary data. Most of the effort of the User Groups occurred during the analysis phase of the software life cycle; however, they will continue to be involved in all phases. Each User Group submitted a report describing their functional and data base requirements. Reports included input forms, computational algorithms, a description of tables and graphics, data definitions, and related characteristics to assist in data base design. The groups or specific individuals within the groups also will be used as technical consultants during the design and implementation phases of the system's life cycle. They will review formats for screens and menus, input forms, and tables and other output media.

The Quality Assurance and Configuration Management Unit is responsible for establishing a system for reviewing and assessing the design and performance of NWIS-II during its entire life cycle. The Unit will work with the Design and Development Team to establish standards for the requirements analysis, design, and implementation or coding phases of the project. They also will be instrumental in designing plans for testing modules, programs, and the complete system. Finally, they will implement a mechanism for assessing and documenting changes.

A vital part of the entire NWIS-II effort is the Design and Development Team. The Team is responsible for following through on all phases of the software life cycle from initiation through system acceptance testing. During the initiation phase, the SPG decided to undertake the design and development effort with Water Resources Division personnel. Plans were made in July 1989 to assemble the Team in Reston, Virginia. Part of the Team consists of permanent staff assigned to headquarters; these are individuals with extensive training in information system development and experience implementing the current NWIS-I system. Most of the Team consists of long-term detailees from District offices across the Nation and detailees from the U.S. Environmental Protection Agency and other Divisions of USGS. The team members have broad-based knowledge of hydrology and experience in how we process data; and they have been provided training in structured analysis, data base design, structured programming, and testing methodologies.

The Team is divided into three groups to accomplish the requirements analysis, design, implementation, and testing phases of the effort. One group is responsible for the functional requirements of the system such as input and edit, computations, and output options. Another group is responsible for the design of the logical data model and the physical design of the data base. The third group is responsible for the transfer of data from the current systems to NWIS-II. This group must compare the logical models of both the current systems and NWIS-II to ensure

that all data elements are accounted for in the new system. They will develop software to populate the new data base with data residing in the current WATSTORE and NWIS-I.

The fifth component of the NWIS-II effort is the Data Management and Operations Unit. This unit currently is responsible for maintaining the WATSTORE and NWIS-I systems. Members of the Unit will have a similar responsibility for maintaining NWIS-II, but they also will be involved in its design and development. The data-base managers who are responsible for the individual files of the current systems will be used by the data transfer group to help compare the logical data models. Because of their familiarity with the file structure, the data-base managers probably will design parts of the transfer programs. Other members of the Unit will be assessing the possible reuse of some of the existing code. A Fortran language analyzer will be used to evaluate programs written for NWIS-I that may meet the users' requirements. The Unit also will be involved in developing plans for and testing the new software, particularly the data transfer programs. After the initial design and development effort is complete, the Unit will have primary responsibility for providing software fixes and developing additional enhancements recommended by users.

SYSTEM REQUIREMENTS

The requirements-analysis phase of the NWIS-II development effort for WRD users has been completed. The eight user groups submitted reports of the discipline-specific requirements of the system on September 30, 1990. The Design and Development Team subsequently integrated the requirements of the eight user groups into the "System Requirements Specification (SRS) for the National Water Information System II." This 750-page document was reviewed in January 1991 by a team of 14 specialists, including the chairpersons of the eight user groups. It is currently being revised to address the comments of the reviewers and expanded to include additional details for input forms, reference lists, and output graphics and tables.

The remainder of this paper consists of abstractions from the SRS. They include a summary of the objectives and features of NWIS-II and the requirements proposed by the eight user groups.

Objectives

NWIS-II is being designed and developed to meet the following objectives:

- Provide a comprehensive data-base management system to support hydrologic analyses for USGS studies including those of the Federal-State Cooperative Program, hydrologic investigations, Thrust Programs, National Water Summary, and National Research Program
- Provide a national water data base
- Serve as an archive for all data used in completed and published work products of the Water Resources Division (WRD).
- Provide data-handling and data-storage procedures that are uniform across all disciplines of the WRD
- Provide a flexible and expandable system that is easy to use and understand by both the WRD and outside users

that all data elements are accounted for in the new system. They will develop software to populate the new data base with data residing in the current WATSTORE and NWIS-I.

The fifth component of the NWIS-II effort is the Data Management and Operations Unit. This unit currently is responsible for maintaining the WATSTORE and NWIS-I systems. Members of the Unit will have a similar responsibility for maintaining NWIS-II, but they also will be involved in its design and development. The data-base managers who are responsible for the individual files of the current systems will be used by the data transfer group to help compare the logical data models. Because of their familiarity with the file structure, the data-base managers probably will design parts of the transfer programs. Other members of the Unit will be assessing the possible reuse of some of the existing code. A Fortran language analyzer will be used to evaluate programs written for NWIS-I that may meet the users' requirements. The Unit also will be involved in developing plans for and testing the new software, particularly the data transfer programs. After the initial design and development effort is complete, the Unit will have primary responsibility for providing software fixes and developing additional enhancements recommended by users.

SYSTEM REQUIREMENTS

The requirements-analysis phase of the NWIS-II development effort for WRD users has been completed. The eight user groups submitted reports of the discipline-specific requirements of the system on September 30, 1990. The Design and Development Team subsequently integrated the requirements of the eight user groups into the "System Requirements Specification (SRS) for the National Water Information System II." This 750-page document was reviewed in January 1991 by a team of 14 specialists, including the chairpersons of the eight user groups. It is currently being revised to address the comments of the reviewers and expanded to include additional details for input forms, reference lists, and output graphics and tables.

The remainder of this paper consists of abstractions from the SRS. They include a summary of the objectives and features of NWIS-II and the requirements proposed by the eight user groups.

Objectives

NWIS-II is being designed and developed to meet the following objectives:

- Provide a comprehensive data-base management system to support hydrologic analyses for USGS studies including those of the Federal-State Cooperative Program, hydrologic investigations, Thrust Programs, National Water Summary, and National Research Program
- Provide a national water data base
- Serve as an archive for all data used in completed and published work products of the Water Resources Division (WRD).
- Provide data-handling and data-storage procedures that are uniform across all disciplines of the WRD
- Provide a flexible and expandable system that is easy to use and understand by both the WRD and outside users

- Provide data security and protection to insure the integrity of the data within the data base
- Implement with ANSI, FIPS, and industry software standards to minimize dependencies on hardware platforms and single vendor suppliers
- Serve as a clearinghouse of sources of water-resources data, including a comprehensive index of water data collected by all Federal agencies.
- Provide a fully-integrated data capability for the National Water Quality Laboratory (NWQL) and other laboratories

Features

The following features have been proposed by user groups or implied by their requirements. Most of these will be included in the initial releases of NWIS-II. Others probably will be added as additional requirements are defined.

- Consistent "look and feel" across disciplines and other major WRD software systems
- Desktop environment utilizing multi-window and multi-tasking capabilities of UNIX and X-Windows standards
- Menus in user selectable colors and also appropriate shades for monochrome monitors
- Multilevels of help at the option of the user
- Entry of data from Data Collection Platforms, Analog Digital Recorders, strip chart recorders, Electronic Data Loggers, radio and telephone, external files and keyboard through standard and user-defined input forms
- Graphical input, editing and analysis
- Verification invoked manually or automatically during input, retrieval, and editing of data including system defined and used-defined checks
- Entry of data by the use of numerous reference lists so entry of codes is not required
- Cataloging of data independent of the U.S. Environmental Protection Agency's STORET codes
- Taxonomic reference list that is coordinated with other agencies
- Automatic access of data during input, editing, verification and analysis
- Interface to a Geographic Information System to assist in the entry, quality assurance, error-checking, spatial verification, editing, retrieval, and display of hydrologic data in a geographic context
- Standard navigation paths for ease of use and efficient processing of basic data
- User-defined scripts for frequently repeated processes
- Data aging through change on data status type
- Access violation tracking
- Output to a variety of media in standard and user-defined formats
- Efficient data entry and retrieval by external software through provisions such as subroutine libraries and data piping between processes

- Links to other DIS-II software
- Data base preservation operations of backup, recover, audit trails, and archiving
- Store data histories
- An index of water data of USGS and other agencies
- Ability to transfer data between NWIS-II and STORET data bases

Functional Description

The Design and Development Team used the techniques of functional decomposition to subdivide the integrated requirement of the users into manageable parts or components. Based on the organizational structure of USGS and data-processing needs, the system was decomposed into nine parts: user interface, data retrieval, data protection, input and edit, data verification, computations, data output, project management, and maintenance.

User Interface

The user interface function can be described as the "look and feel" of the NWIS-II, because it affects users: 1) perceptually (e.g.,via the computer screen), 2) physically (e.g.,via the mouse), and 3) conceptually (via the mental image of the model). One of the goals of the NWIS-II is to provide a consistent and relevant "look and feel" across disciplines by providing standardized functionality and displays. In order to attain this goal, NWIS-II will utilize the multi-user and multitasking capabilities of a UNIX platform. The NWIS-II will support the use of various computer peripherals for data input and output (e.g., digitizers, digital-tape readers, scanners, plotters and printers). The exact models of peripheral devices that will be supported are those procured under the DIS-II contracts and selected peripheral devices that existed prior to the DIS-II. Due to the use of graphical windows and menus in NWIS-II, only color-graphics display devices that support the X-Window System will have the ease-of-use and total functionality of NWIS-II. Gray-scale terminals that support the X-Window System will have the total functionality of NWIS-II but not the same ease-of-use as color-graphics display devices. Character-based terminals will have limited functionality.

Data Retrieval

Data retrieval is the process of users selecting data from a data base. Data retrieval will function in an integrated, physically distributed system. All WRD hydrologic data, including selected National Water Quality Laboratory data, will be stored in the NWIS-II as one data base, rather than in separate discipline-specific data bases. Hence, the data retrieval function will be the same for all disciplines. Data from other WRD data bases (e.g., the electronic mail system and the Administrative Information System) also will be accessible by applications implemented in the NWIS-II. Retrievals from these data bases will function as if the data were within NWIS-II. Online indexes of the availability of data within the NWIS-II and in other data bases will be accessible by the retrieval function. NWIS-II data will be physically distributed among a network of nodes located throughout the Nation. Retrieval will function the same at any node. Specific data retrieval requirements are:

1. Provide the ability to retrieve on data elements in other WRD data bases

The retrieval function shall allow access to data elements in other WRD data bases, such as the Administrative Information System and WRD's EDOC User-Identification Data Base.

2. *Provide polling in data retrievals

NWIS-II shall be able to provide user-controlled and automatic polling for retrievals that cross nodes. Retrieval of data from any node will depend on the user's access rights. In retrieval requests, users shall be able to specify the nodes to be polled and when the nodes will be polled. Users also shall be able to allow the NWIS-II to automatically implement the polling necessary to complete a retrieval.

3. *Provide the ability to pause and warn about potential problems

As a default, users shall receive status messages before the data associated with a retrieval request are extracted, if any of the following occur:

- completion of the retrieval will take a long time
- an excessive number of records will be produced
- the user's allocated disk space will be exceeded
- other nodes will be polled
- data with restricted access are being requested

4. Provide the ability to suppress pause and warn and summary messages

Users shall have the ability to suppress the default pause for warning and summary messages and direct the output of these messages to a file.

5. Provide a retrieval summary

A default summary of a retrieval request shall be provided before the request is submitted. This summary shall consist of the results of the retrieval specifications, which will include a list of the data elements requested, their frequency of occurrence, and a status flag indicating whether they are original record, in working status, review status, or final status. Users will be able to retrieve only those records for which they have viewing rights. Users shall be able to access status information about polling at nodes. This status information will inform users about which nodes are closed to polling and provide a schedule for the known closing and opening time periods.

6. Provide the ability to access a retrieval before it is complete

If there is a system failure or any other event that causes the extraction of data from the data base to take a long time, users shall be able to access data already extracted and either continue or cancel the retrieval execution.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations..

7. *Provide the ability to use boolean logic for retrieval specification

The following boolean operators shall be available for specifying retrieval requests.

Boolean operators

equal to	equal to null
not equal to	not equal to null
greater than	equal to missing
greater than or equal to	not equal to missing
less than	in range
less than or equal to	not in range
contain	and
not contain	or

8. *Provide the ability to retrieve by qualifier codes

Users shall be able to make retrieval specifications using the following qualifier codes.

Qualifier codes used or needed in existing system

Qualifier codes (associated with a real value)	Definition
<	less than value shown
>	greater than value shown
E	value estimated
ND	not detected in sample
NC	sampled but not counted in grid
L	organism count less than 0.5%
D	organism count dominant at greater than or equal to 15%
&	organism estimated as dominant
K	based on colony count outside acceptable range
TNTC	too numerous to count
NICC	not in colony count
ZC	zero counts
M	missing or not collected
UN	value unknown
NE	value available but not yet entered
NA	not applicable

9. Provide retrieval of quality-assurance data

A system-defined retrieval option to retrieve quality-assurance data shall be available to users. This option will allow the user to subsequently review or perform computations in the analyses of quality-assurance data.

10. * Provide automatic selection of preferred values for data elements

If a retrieval involves the selection of data elements with multiple values, the values that have been designated as preferred shall be selected unless all values are selected. If multiple values exist for different agency codes, the values with USGS will be flagged preferred by default.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

11. Provide the ability to save and/or modify retrieval scenarios for reuse

Users shall have the ability to save, modify, and then execute a set of retrieval specifications. If the set of specifications was created with a menu, the user shall be able to use NWIS-II menu simulation to scroll through the selections and make changes. If the set of specifications was created with English commands, the user will be able to use an editor to make changes. Before committing to a retrieval specification, the user shall be shown the specification selected and given a chance to make changes, if running interactively on the system. If running in batch mode, the retrieval is automatically executed with no opportunity to change the specification.

12. Provide the ability to define a script to be used in a retrieval specification

Users shall be able to write and store scripts that can be used to make retrieval specifications, thereby requesting data in units in which they may not be stored or using algebraic notation on selected data elements to create new output variables.

13. *Provide the ability to selectively extract a specified number of records

Users shall be able to selectively extract and view a specified number of records from a data set determined by a set of retrieval specifications. Hence, users could extract a number of records from the front, middle, or end of the data set. They could extract a single record, single value, range of values, or period of record by dates.

14. *Provide the ability to retrieve by user-defined time periods

Users shall be able to define time units for use in retrieval specifications. For example a user could define a week to begin Monday and end Sunday, or a year to begin June 1 and end May 31. Certain system default time periods will be predefined, such as water year, calendar year, climatic year, month, and week.

15. *Provide the ability to retrieve data using delineated areas or points

Users shall have the ability to use a delineated area to specify a data set for retrieving information about NWIS-II sites that fall within the delineated area. The delineated area can be:

(1) selected from a thematic map. For example, users could select a county from a political boundary thematic map by pointing on the screen using a mouse and the requested information about the sites that fall within this area would be in the retrieval data set.

(2) a user-defined area. Circles, rectangles, and polygons are the types of areas users can define. They can be defined on the screen using a mouse or defined by pointing to a topographic map registered on a digitizer.

(3) a user-defined GIS coverage. This coverage must be in the same projection and coordinate system of the supplied thematic maps.

16. * Provide the ability to retrieve data by selecting sites

Users shall have the ability to specify a retrieval data set by selecting individual sites with a mouse that are graphically displayed on the screen.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations

17. *Provide the ability to retrieve by latitude and longitude

Users shall have the ability to specify retrievals by latitude and longitude by:

(1) delineating a polygon with 3 to 1,000 pairs of latitude/longitude coordinates entered in counterclockwise order from a keyboard or user file

(2) delineating a circle by latitude/longitude pair and radius

(3) specifying points or sites by a latitude/longitude pairs

18. Provide the ability to retrieve data using a generated buffer

Users shall be able to define buffering distances around thematic map features and NWIS-II sites that define a retrieval data set of NWIS-II information within buffer zones.

19. Provide the ability to retrieve data by overlaying thematic maps

Users shall be able to retrieve information based on spatial commonalities of overlaid thematic maps. The common areas are used to define a delineated area and the requested information about sites that fall within the area would be in the retrieval data set.

20. *Provide the ability to retrieve data using networked features

Users shall be able to retrieve on networked features by location and flow direction within the networked feature. It shall be possible to select features by the following specifications:

- up or down gradient of a location
- within a distance up or down gradient of a location
- between two locations
- defined segments of networked features (e.g., stream reaches of a networked tributary system) by gradient and number of NWIS-II monitoring sites

21. Provide the capability to retrieve spatial data involving calculations

Users shall be able to specify a retrieval based on calculations of spatial data. Examples: 1) For a given set of wells, determine the distance of each from a specified line (stream), polygon (lake), or point (site); 2) Locate the nearest well that has a value of fluoride concentration sampled within the last 12 months; or 3) Locate the 10 wells nearest to a specific feature.

22. Provide the ability to combine GIS and non-GIS retrieval specifications

Users shall have the ability to limit retrieval data sets by using the GIS in combination with non-GIS retrieval specifications.

Data Protection

The data-protection function is implemented to protect data and information in NWIS-II from deliberate or accidental alteration. Data are protected by restricting user access to processes that alter data. The goal of the data protection function is to give data managers a consistent method

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

for controlling data access. The protection scheme is based on two concepts: data aging and project-associated user access. Data aging describes the flow of data as processed for approval. In NWIS-II, a "project" is considered a group of data with a definable purpose and limits on user access. A project may coincide with the project designation for the USGS Management Information System. Project data can be subdivided into smaller groups of data, which have the same protection features.

Input and Edit

Input functionality describes the means by which data will be entered into NWIS-II. All data will initially interact with the input process; therefore, input functionality must be both flexible and robust. It will communicate transparently with a variety of field data collection and storage instrumentation, here categorized by data transfer and transmission methods. Communication among files is an inherent part of a relational data base, as values can be imported from one file to another (or entered to a selected set of records). Keyboard, scanned-image, and digitizer entry will also be supported. To provide flexibility, the process of input must be linked to the functionality of editing. Data often require manipulation to remove anomalies, errors, or false recordings. Editing on input allows a user the versatility to correct data values before they are stored in the data base.

An interface to a geographic information system (GIS) will facilitate the input and editing of site locations and associated attribute data. NWIS-II will support both character and graphical input and editing of data values and geographic locations. The ability to edit several parameters at the same time of the same or different disciplines will be inherent.

Data Verification

The data-verification function is the group of processes that are invoked manually or automatically to help ensure the correctness of data in the NWIS-II data base. Data verification occurs as error checks and data review. Error checks are executed against data when initially input to the data base and after new data have been calculated and/or existing data have been manipulated in some manner within NWIS-II. Once data entry and data manipulation ends, another verification process occurs in the form of data review. The rules, definitions, precision, and descriptive text associated with each verification (e.g., significant figures and precision rounding) will be outlined in the NWIS-II data dictionary.

Verification checks fall into three categories: system-wide, user-defined, and hydrologic-event-notification. System-wide checks are standard to NWIS-II and executed on data throughout the network. User-defined verification checks are executed on specific data selected by users. These checks can be used to supplement system-wide verification checks or may be independent of other verification checks. Hydrologic-event-notification checks are used to check specific data for conditions that require immediate notification and/or some sort of action.

Computations

Computations are used to manipulate and analyze basic water-resources-related data. The computations provided for by NWIS-II can be executed either interactively or from a script and will include mathematical functions, statistical functions, graphical functions, and a transparent

interface to use a geographical information system and spreadsheet package. The system will provide specific capability for computing the traditional data collected and analyzed by USGS, including water discharge, sediment and chemical constituent loads, water use, and biological indicators.

Data Output

Data output is the production of information delivered to a user as paper copy or video display, or transferred to tape or disk. Data output results from a user's request or query for data in NWIS-II using the retrieval subfunctions described previously. Data output will be displayed in a standard format with predefined units of measurement and number of significant digits. Data also may be output using user-defined formats, units of measurement, and number of significant digits. Some standard output formats will duplicate input forms used to enter data into the NWIS-II data base. Information not stored in NWIS-II, but derived from NWIS-II data through calculations, will be output. Specific data output requirements are:

1. *Allow users to output data to a variety of media

NWIS-II shall support the output of data to graphics and character-based monitors or storage media, such as floppy disks, magnetic tapes, and paper.

2. *Allow users to specify units of measure for output data

Tables and figures shall be output with default units of measure to meet publication and report requirements. Users can specify units of measure for an output that differ from those stored in the data base. Users shall be able to specify the time system; for example, PST, CST, EST, CDST, MST and UTC. Users also will be able to supplement the list of default units of measure applicable for specified outputs with additional units.

3. Allow the output of special characters, data flags, and symbols

Special characters are handled in NWIS-II supplied outputs by embedded software that is transparent to users. When requested by users, NWIS-II shall show a complete or pertinent partial listing of data flags, special characters, and symbols with their associated meanings.

4. *Output data with different levels of significant digits or rounding

The precision associated with numeric data can vary due to methods of data collection, data analyses and output, and computer representation of the data. NWIS-II shall allow users to output numeric data by one of three methods:

(1) Unrounded numeric data: The actual numeric representation or precision of the data as input and stored in the data base.

(2) NWIS-II data dictionary rounding: the rounding or precision of numeric data is based upon the type of constituent, its method of collection and analysis, and use of final output.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

(3) User-defined rounding.

5. *Provide the ability to output textual descriptions

Users shall have the ability to output text to aid in creating manuscripts, reports, notes, memos, or tables.

6. Provide the ability to output scanned images

Users shall have the ability to output scanned images from NWIS-II. These images may include sketches of sites, reference points, and well-measuring points; diagrams of pumps; images of taxa (flora and fauna); and schematics of water-user systems.

7. *Provide the ability to output a standard and consistent plotted or printed map

Users will routinely require maps plotted or printed with selected NWIS-II features overlaid on the appropriate thematic layers. These maps shall be consistent, all will have scale bars, north arrows, explanations, titles, and meaningful symbols by default. Output shall be at standard scales such as 1:24,000, 1:100,000, 1:250,000, and 1:2,000,000. Users will be able to register the maps to USGS topographic quadrangle maps. Page size plots, 8.5 x 11 inches, shall be an available output. Users shall have a default set of thematic layers available for use as a base map.

8. *Provide the ability to output graphics plot, charts, and diagrams

Users shall have the ability to output graphics plots, charts and diagrams. Many of the graphics functions will be provided by the Office of the Assistant Chief Hydrologist for Program Coordination and Technical support. The NWIS-II design and development team shall develop functions for the following graphical outputs:

- x-y plots
- logx-logy plots
- probability plots
- time-series plots
- duration hydrographs
- channel cross-sections

9. Provide the ability to incrementally move overlaid time-series plots

Users shall have the ability to move time-series plots for comparison purposes. With this ability, a user could overlay a flood hydrograph of a downstream station on a hydrograph of the same flood at an upstream station and move the plots until the peaks matched.

10. Provide the ability to output user-customized figures

Users shall have the ability to annotate and enhance figures for customized output and to graphically display data using different scale notations. Examples of scale notations are listed below:

- arithmetic

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

- logarithmic (base 10 & natural)
- exponential
- probability

11. Provide the ability to output publication-quality tables

Users shall have the ability to output tables of data that conform to WRD publication standards. The table output should be camera-ready quality so that no additional editing is necessary before including in a manuscript.

12. *Provide the ability to output time-based data in tabular form

The system will be able to output tables of unit values, daily values, or data based on discrete time periods. These tables also will display extreme values either by day, month, year, or period of record.

13. Provide the ability to output tabular forms of ratings

Users shall have the ability to output lists of points that define rating curves and possible offsets. Lists of points may originate from graphical representations of surface-water stage-discharge or sediment-transport curves. The output form for the rating table will include gage-heights to the nearest 100th of a foot and the corresponding discharge.

14. *Provide the ability to output a summary of retrieval specifications

The system shall allow users to output a summary of retrieval specifications. This summary will contain the number of times each requested keyword was found in the data base using boolean logic.

15. Provide the ability to output data dictionary information

NWIS-II will be able to output data dictionary and meta data listings, as well as special characters and symbology information. This will include the definition of the data or meta data listed, size allowable and the range of values allowed for the data element. Associated constituent and method identities also are included as information that may be output.

16. *Provide the ability to output a data base dump

Users shall have the ability to output the complete contents of a data base. The data could be used for simple viewing, archiving, or repopulating a damaged file after making fixes.

17. *Provide the ability to output unit-values data

The standard output of unit-values data will be provided in two tabular formats to be combined with a plot of the computed daily values of any parameters selected. One format will provide adjusted recorded values, interpolated for each hour of the day, along with the mean, maximum, and minimum daily recorded and computed values and their corresponding corrections. This format shall be the default of standard output. Users shall have the option to select a second format, which will provide daily minimum, maximum, and mean values of the

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations

recorded and computed values, along with their corresponding corrections. Both forms of standard output also shall include any flags generated during processing and selected information about measurements made during a given day.

18. *Provide the ability to output daily values

Daily values shall be output in a standard format for publication in the annual data report; the exact format will depend on the parameter selected. These tables shall be combined with a plot of the selected daily values.

19. *Provide the ability to output standard water-use tables that are periodically modifiable

Output for water-use discipline tables for aggregation will be standard nationally but NWIS-II shall provide new table formats for the report. In addition, users will be able to output different sources of water and different years of data at one facility. The default unit of measure in output forms will be million gallons of water per day. Users also will be able to specify units of measure. For population numbers, the default will be in thousands of persons. Both elements will be reported to two decimal places to the right of the decimal point.

20. Provide the ability to output a table of non-target compounds

Non-target organic substances form a special class of analytical results, which are determined through a mass spectrometer. There are thousands of potential compounds listed in mass spectral libraries. These substances are tentatively identified and semi-quantitatively estimated. They are not on the U.S. Environmental Protection Agency list of organic compounds. The values do not meet USGS standards for acceptance, but they represent potentially useful information to users so they would be stored on line.

21. Provide the ability to output quality-assurance documentation

Users shall have the ability to output tables of quality-assurance (QA) information. Special QA tables may be produced from QA protocols that use reference, blind, or field blank samples. There also are two types of laboratory quality-assurance tables: brief and complete. The message in the brief table will contain information on a single file. The message in the complete table will contain information on all parameters that exceed the QA criteria.

22. Provide a tabling package that conforms to accepted publishing guidelines for taxonomic nomenclature

NWIS-II shall include a tabling package that conforms to USGS publishing standards and accepted publishing guidelines for taxonomic nomenclature. For taxonomic names, the genus and species should be underlined or italicized and the first letter of the genus name should be capitalized.

23. Provide for output of discharge measurement and shift analysis

For every measurement selected for output, the number of the rating curve or shift curve used to compare the measurement will be displayed.

---

* Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

24. Provide the ability to output summary of gage-height corrections

Users will be able to output a summary of all gage-height corrections at a specified station for the water year.

25. Provide the ability to output an end-of-year summary

Users will be able to output a tabular summary of maximums, minimums, and peaks above base for the entire water year.

26. Provide the ability to output a rating curve with shift curves included

Users will be able to output a rating curve with shift curves included.

27. *Provide the ability to output tables of selected meteorologic data

NWIS-II will provide the ability to output tables of meteorologic measurements, such as rainfall (intensity and cumulative amount), wind (speed, duration, and vector direction), and sunlight (intensity and duration).

28. *Provide the ability to output tables of sediment loads

The amount of suspended sediment and bed load detected at surface-water sites will be used to estimate the amount of material moving downstream over a period of time. Users will have the ability to output sediment loads in tons per day, as well as to make monthly and annual sediment load tables from daily load information. If direct computations are being performed, users will be able to output the computed instantaneous readings and the interpolated discharge; and if a regression is being used, then regression coefficients may be output. Users also shall have the ability to output concentrations, particle size, and bed materials.

29. Provide the ability to customize output forms

Users shall be able to customize existing output table formats or create new ones. The user-customized output format shall interact with the data base retrieval software so that a user's data are retrieved and displayed in the designated format. Customization of output table formats shall exist at the District, Project and User-defined levels. User-defined forms are created for specific users so others will not be able to modify these forms. Project- level forms are defined by project personnel for a group of users. District forms are designed for users in a District. The system also shall provide the ability for self-scaling of column widths. For example, if a field for well name is 40 characters wide, but the longest name retrieved is 25 characters, the field should optionally be rescaled to 25 characters.

30. Provide the ability to output GIS-compatible formats

Users shall have the ability to output NWIS-II information into two GIS compatible formats: (1) GIS coverage and (2) GIS export file.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

31. *Provide the ability to output in WATSTORE card format

Provide data output in the WATSTORE card formats for the following data: unit values, daily values, and water quality.

32. * Provide the ability to output in STORET file format

Provide data output as STORET-compatible ASCII text files.

33. *Provide the ability to export data in industry standard formats

NWIS-II shall have the ability to output information in the following formats:

- DIF format
- Postscript
- ASCII and ASCII-delimited
- CGM files and/or other graphics formats

34. Provide the ability to sort postprocessed data

Users will have sorting capabilities, which include:

- Sort output data set on any element by numeric values, to include maximum and minimum values
- Sort output data set on any element by alpha values
- Use boolean logic
- Compare differences and similarities.

35. Provide the ability to reformat data by row and column position

Reformatting of data shall include:

- Output columns of data that are aligned
- Output data separated by spaces
- Trim rows and columns to eliminate labels such as date
- Delete columns or rows of data

36. Provide the ability to merge data sets

The system should be able to merge whole or sections of output files.

37. Provide the ability to manipulate special characters

NWIS-II's use and output of special characters could present problems to other software packages unable to read or interpret these characters or symbols. Users should have the ability to visualize these special characters and use tools to substitute or remove these characters and symbols as necessary. The manipulation software shall allow users to redefine the special characters and symbols used in NWIS-II output data, so the data can be interpreted by external software packages.

---

*Requirements marked with an asterisk (*) are applicable to all users; those without an asterisk are specific to WRD users with DIS-II workstations.

Project Management

The functions for project management are designed to help the manager of a project to schedule and monitor hydrologic activities such as data collection, data analyses and review, and report preparation. The use of the word project in this section refers to a hydrologic study, a hydrologic data network, a special studies section, a hydrologic data section, or any other special group involved in the collection, analyses, or reporting of hydrologic data. A necessary subfunction included within project management is hydrologic event notification. This subfunction provides for the means to inform project personnel about occurrences of relative hydrologic significance that need validating and may require some alternate action.

Maintenance

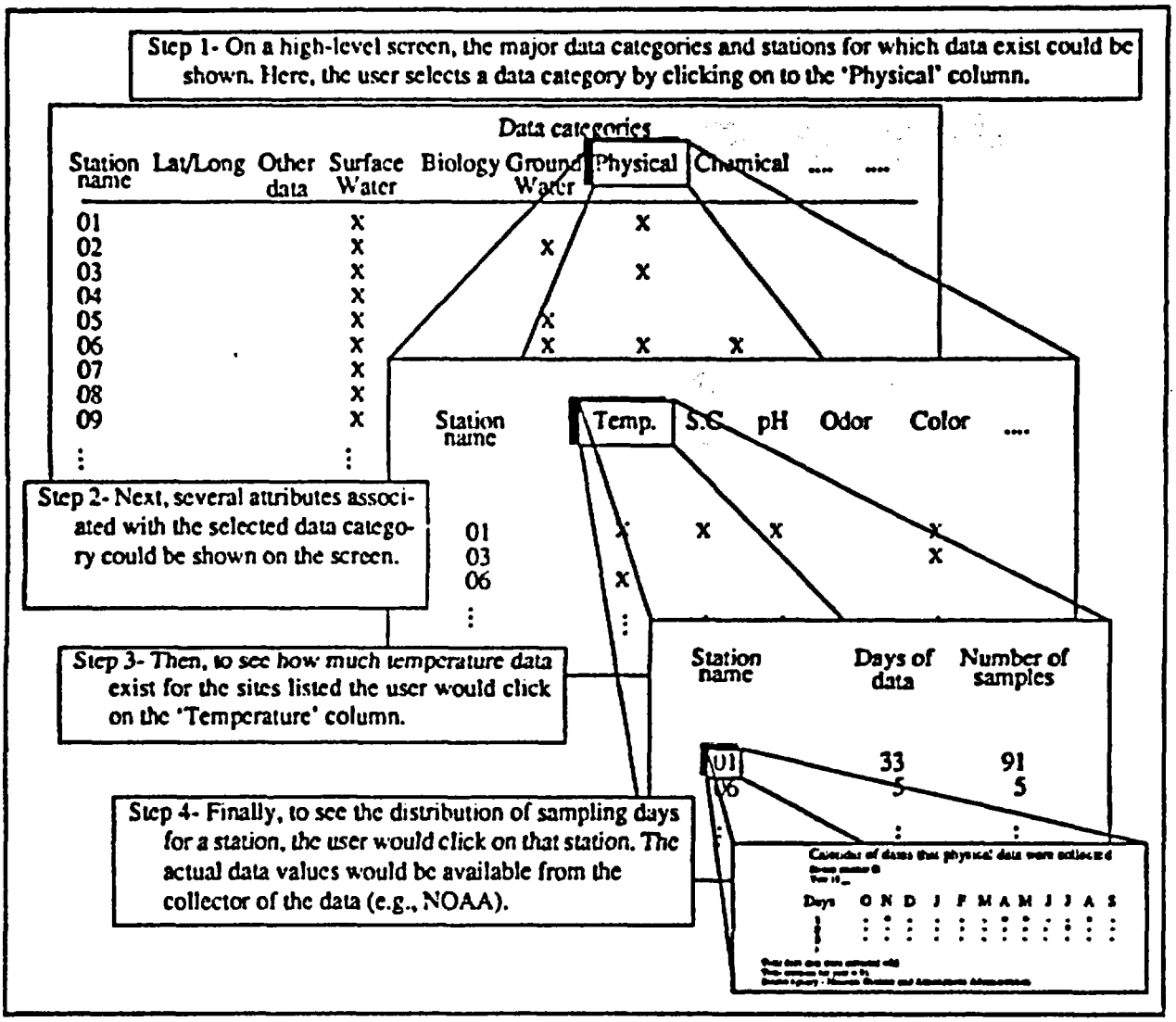
The maintenance functional requirements cover the "care and tending" of NWIS-II. This includes maintenance of the data stored in the data base, the software used to interact with the data base, the operating system that controls the interaction between the data base and input and output devices, and third-party software associated with the data base. Electronic mail software and USGS computer models will not be maintained as parts of the NWIS-II.

Maintenance requirements are organized in the following subsections: data distribution, data base maintenance, and system maintenance and administration. Data are distributed using functions that allow the transfer of data within the NWIS-II data base. Data base maintenance involves functions that protect and maintain data base files. System maintenance and administration involves functions that control changes in NWIS-II.

Data Indexing

The NAWDEX User Group has proposed a major change in how water data are indexed. The current Master Water Data Index identifies water data availability by groups of parameters and frequency of collection. For example, the index contains information that metals data are available and the frequency of collection for the last year of record. It does not have information on specific constituents, total number of analyses available, or specific dates when samples were collected or some other determination was made.

The index proposed as requirements for NWIS-II is a much more extensive and robust system than NAWDEX. The index would provide information on individual constituents or physical parameters and the actual dates when data are available. Instead of a broad category of physical parameters, each parameter would be indexed. For example, at a particular site, the user could determine the total number of temperature measurements and the days when measurements were made. For an interactive retrieval of index data, the user might query the system through a series of screens. The first screen would provide information that is broad in scope. Each successive screen would be narrower in scope, which would lead users to the most specific information in the index. The following figure is a conceptual diagram of how an on-screen, data-availability retrieval might work:


Step 1- On a high-level screen, the major data categories and stations for which data exist could be shown. Here, the user selects a data category by clicking on to the 'Physical' column.
Data categories
Station name
Lat/Long
Other data
Surface Water
Biology
Ground Water
Physical
Chemical
Step 2- Next, several attributes associated with the selected data category could be shown on the screen.
Station name
Temp.
S.C
pH
Odor
Color
Step 3- Then, to see how much temperature data exist for the sites listed the user would click on the 'Temperature' column.
Station name
Days of data
Number of samples
Step 4- Finally, to see the distribution of sampling days for a station, the user would click on that station. The actual data values would be available from the collector of the data (e.g., NOAA).

The utility of the proposed index in obvious; it will provide comprehensive information on available water data. However, there is a high cost associated with collecting and storing the index data. The information in the current NAWDEX is provided to USGS by most other Federal agencies and State and local agencies via forms that list site information, parameter group, and frequency code. Instead of information for 214 data components and parameters groups and 33 frequency codes, the proposed index will require information for thousands of constituents and an unlimited number of dates. This mass of information can only be exchanged through electronic media, and this will require software to generate the index data from the data base of each agency. Some agencies that have their data stored is a relational data base similar to NWIS-II, will be able to use the NWIS-II indexing software. However, most agencies probably will have to write software to generate the index data in the NWIS-II format.