Policy Recommendations for Management and Retention of Hydrologic Data of the USGS

Compiled by E. F. Hubbard

ABSTRACT

A data-policy committee was formed in the late spring of 1991 at the request of the Strategic Planning Group, which is composed of senior managers of the Water Resources Division (WRD) of the U.S. Geological Survey (USGS) engaged in planning the National Water Information System (NWIS II). NWIS II is a distributed-information system that will provide electronic-data storage, management, manipulation, accessibility, and security for most of the hydrologic and related information collected or complied by WRD. Once data in NWIS II are reviewed and approved, they will become available for public viewing and copying, be maintained indefinitely, and be designated as collected by WRD or by some other organization. The committee was commissioned to consider five issues important to the design of the system. The five issues, part of which also apply to records on paper or in other form, and key recommendations of the committee are:

1. What are original data? The committee must define the term.

The concept of original data has changed somewhat from the days when a pen trace or hydrographer's notes represented original data. For present day usage, the committee defined original data in this manner:

Original data--from automated data-collection sites, laboratories, outside sources, and non-automated field observations--are those data that shall be preserved unmodified as collected or received, once in conventional units (engineering units, generally with a decimal).

2. What original data collected by the WRD (such as 15-minute stage data) and other data stored in WRD data bases should be placed in data archives? The committee should draft a policy.

All original data that are published or support published scientific analyses shall be placed in archives. The WRD should establish central coordination of archiving original hydrologic data to ensure that valuable, irreplaceable data are safely preserved in accordance with the USGS Files Management Program provisions and the WRD Mission-Specific Records Disposition Schedule.

3. What minimum data set should be maintained on line or available for immediate access at all times?

Relatively new data that are being compiled, modified, reviewed and frequently referred to require high levels of accessibility. As these data become older, lower levels of accessibility are typically needed. Data in the data base should generally move from areas with high accessibility and low security to areas with lower accessibility and higher security as need for access diminishes and it becomes finalized (moved from working and review files to approved files). As this occurs, data may be moved from on line to near on line or off line, as appropriate. It is anticipated, however, that much of the data will remain on line. These data, once approved, will be considered as being in archives, regardless of whether on line or off line.

4. What data should be stored in NWIS II in addition to data collected by WRD, such as: (a) water data collected by other agencies or (b) ancillary data, including spatial data coverages, such as a geographic information system, and digital model simulations?

Current (1992) policy requires that all water data collected as part of the routine data collection of the WRD be stored in computer files of the National Water Information System. This policy is expanded by the addition of: Data collected by others--such as cooperators, universities, or consultants--that is used to support published USGS documents and not published or archived elsewhere, shall be placed in NWIS II. Other outside data may be placed in the data base at the appropriate manager's option. It is not anticipated that spatial data coverages and model simulations will be part of the data base residing in approved files, except for the streamflow-simulation results of the USGS BRANCH model.

5. Who should have access to data stored in NWIS II? The committee should draft a policy defining who should have access to the NWIS II data base, including provisional data, with particular emphasis on non-USGS users.

A protection scheme, granting access rights to WRD personnel and other users, will be part of NWIS II. The general public will have access to all approved, non-protected files through the National Water Data Exchange system For approved (archived) files this access will be limited to viewing or copying, except for a very few WRD employees with duties that include data-base changes. Cooperators and others may be granted more direct access to local working files at the discretion of an appropriate manager.

INTRODUCTION

The U.S. Geological Survey (USGS) is presently acquiring (January 1992) a network of computer work stations that will link nearly all Water Resources Division (WRD) offices across the country. This network is a new Distributed Information System (DIS II) and will take the place of DIS I, which was a network of mini-computers. DIS II will provide nearly all of the computer support required by the technical staff of the WRD, as well as providing administrative and data-base functions.

The storage and management of hydrologic data is a function of the DIS II that is very important to WRD. Thus, the Division is preparing software, containing a data base--a new National Water Information System (NWIS II)--that will replace the present system, NWIS I.

The Strategic Planning Group (SPG), composed of members of the WRD Senior Staff, is directing the acquisition of hardware and the preparation of software, of which NWIS II is a key and major part. The Committee on Hydrologic Data Policy (the committee) was commissioned in response to a request from the SPG to consider five fundamental issues important to the design of the NWIS II data base. Some of these issues are also quite important to the retention and management of non-electronically recorded data, in paper or other form. The five issues are as follows:

1. What are original data? The committee must define the term.
2. What original data collected by the WRD (such as 15-minute stage data) and other data stored in WRD data bases should be placed in data archives? The committee should draft a policy.
3. What minimum data set should be maintained on line or available for immediate access at all times?
4. What data should be stored in NWIS II in addition to data collected by WRD, such as: (a) water data collected by other agencies or (b) ancillary data, including spatial data coverages and digital model results?
5. Who should have access to data stored in NWIS II? The committee should draft a policy defining who should have access to the NWIS II data base, including provisional data, with particular emphasis on non-USGS users.

Members of the committee, who were selected to provide a broad range of technical and organizational perspective, are:

John Briggs	Alan W. Burns
Alberto Condes de la Torre	Robert E. Faye
Ernest F. Hubbard (chair)	Jayne E. May
Nick B. Melcher	Joe A. Moreland
Kathy D. Peter	Robert R. Pierce
Keith R. Prince	Peter F. Rogerson
Vernon B. Sauer	William G. Shope, Jr.
Wayne B. Solley	Wayne E. Webb

Additionally, this group was assisted by Jeffrey D. Christman, James L. Kiesler, John S. McLean, and Timothy C. Stamey--who provided technical consultation and advice, both in their areas of expertise and in general, as the issues were considered. Early versions of the committee's recommendations were furnished to NWIS II staff members involved with designing the system, who raised questions and gave information that were helpful in finalizing the recommendations. Chairmen of the user requirements analysis groups, which were responsible for providing advice concerning the design of NWIS II in specific technical areas, were also asked to comment on the recommendations. Those who responded provided useful suggestions, helping shape this document.

The committee, which met three times and exchanged much correspondence during the period from May through July 1991, prepared a draft of recommendations dealing with the five issues. This draft was presented to the SPG in August. At the suggestion of the chairman of SPG, the draft recommendations were distributed for comment to all district, area, and regional offices, and to the assistant chief hydrologists. All responses were positive: some were outright endorsements; others were helpful comments and questions.

The committee met again in October and January to consider the comments and questions and modify the recommendations as necessary. Except for a few comments or questions that dealt with matters beyond the scope of the committee's charge, virtually every response resulted in clarification or modification of the recommendations.

In January 1992, the revised recommendations were presented to the SPG, who asked that the recommendations be published as an open-file report and, subsequently, transmitted under a WRD numbered memorandum as policy of the USGS. Specific recommendations that address the five fundamental issues on the design of NWIS II and on the management and retention of non-electronically recorded data are discussed below.

ISSUE 1--DEFINITION OF ORIGINAL DATA

Issue 1 concerns the definition of what constitutes original data. The committee originally restricted debate on this subject to data in electronic form, then expanded it to other types of original data at the suggestion of the team that is preparing NWIS II. Original data as defined by the committee are those data that should be preserved unmodified as collected or received.

As this definition may imply, the term "original data" has lost some of the meaning it had in the past. Discussions of the committee included the thought that the term was useful when one could point to tangible evidence such as a pen trace or a pencil mark and testify that this was original data. At one time the original record would have some additional credibility because in most cases it was difficult or at least cumbersome to fabricate. With electronic data collection there is little tangible evidence of its authenticity, and because the data can easily be fabricated or manipulated, the term "original data" has lost relevance in this sense.

An earlier WRD work group, concerned with the placing of hydrologic data in electronic form in archives, was chaired by Vernon B. Sauer. In a recommendation arising from a 1988 meeting in Atlanta, the work group defined original data in terms of surface-water stages. The following definitions and descriptions expand and paraphrase their thoughts to include all kinds of hydrologic data collected in the field, determined in the laboratory, or received from other sources:

Recommendations on Issue 1

- Original data for automated data-collection sites are defined as unaltered (no changes in magnitude) data acquired from the primary sensor (and back-up sensor, as needed) and converted to conventional (engineering) units in a form and format suitable for entry into NWIS. For stage data this means that the uncorrected unit values once converted to electronic form in feet and hundredths of feet, with a decimal included, are the original data. This record is uncorrected for datum and time and will be permanently preserved, or archived, as original data. [It may be noted that this definition differs from the usage in the Automated Data Processing System (ADAPS), which is the surface-water data processing system currently used by WRD. The SATIN utility of ADAPS, which processes electronic data from data-collection platforms and data loggers, has archive files that contain raw data that may not converted to conventional units (depending on the source). This discrepancy should be corrected in the surface-water data processing system (and other systems) to be incorporated in NWIS II such that original data for archiving will be unaltered data converted to engineering units.]
- Once the stage data are corrected to reflect as accurately as possible the actual events recorded, they also should be preserved indefinitely, but should not be considered original data. The work group referred to these data as "edited."
- Original data from other sensors--dissolved oxygen, pH, temperature, precipitation, specific conductance, ground-water levels, and so forth--will be treated similarly:

The uncorrected data, when first converted to the conventional units in which it will be reported, are considered original. Corrected data would be considered edited.

- Records of the algorithms used to convert the data to conventional units are to be preserved with the data, if not documented elsewhere. Some of these algorithms change slightly with time. When substantive changes do occur, the superseded algorithm shall be preserved with the dates for which it was in effect. Records of the corrections applied, giving field observations and other supporting data, shall be preserved with the edited data, if in electronic form or if not documented elsewhere.
- For data from the National Water Quality Laboratory (NWQL) and other WRD laboratories, typically in electronic form, original data are those received from the laboratory for entry into the district NWIS II data base prior to any modification. (Some of the data that results from laboratory operations, prior to the transmittal of information to the user, will also be stored and preserved. Management and retention of laboratory internal data is beyond the scope of this report, however.)
- Data that are obtained from outside sources (cooperating agencies, contractors, consultants, outside laboratories, other agencies, universities, and similar sources) are considered original as received. If essential for WRD analyses (that is, used to support published conclusions), WRD shall preserve these data if not published or indefinitely preserved elsewhere. These records shall be preserved in the form in which they are received. For example, paper records may be manually transcribed to electronic form, but the resulting file(s) would not be considered original data. Thus, the paper records should be preserved. If data are received in electronic form, then they are considered original for WRD purposes and should be entered into the NWIS II data base as such.
- Non-automated data collected by WRD (such as field determinations of water-quality parameters, discharge-measurement notes, crest-gage inspections, measurements of water levels in wells, differential level notes, well-inventory data, observer's cards, and similar data) shall be preserved as original data in the form in which they were recorded in the field. Actual copies (not transcriptions) of these

data, such as scanned images or photocopies, may also be preserved as original data.

The committee has determined, on the basis of published legal opinions, consultation with experts, prior opinions of the Department of the Interior solicitors, and the Federal legal code, that exact copies of the original data are considered original for legal purposes in Federal court. (There are some minor differences in individual State codes.) This determination pertains to paper records that have been copied by a copying or imaging machine, punched paper tapes that have been converted into electronic records, and data that have been copied from one electronic medium to another (magnetic tape to optical disk, for example). Records that have been transcribed (copied by a process that involves potential human error) are not considered original in a court of law. For example, field notes that are manually transcribed from paper to electronic form would not be considered original in the electronic form. Data of this type might be accepted into evidence but would not have the impeccability of the original.

ISSUE 2--DATA TO BE PLACED IN ARCHIVES

The second issue addresses what data should be placed in archives. To answer this question, the committee defined two terms for use in the discussion and the resulting recommendations. The first term, "electronic archiving," referring to data in an electronic data base, involves moving data to a lower level of accessibility. (For example, taking data off line.) The second term, "archiving," is broader, referring to long-term preservation of data in both paper and electronic form:

Electronic archiving.--The systematic process of removing from the active, on-line data base original data and data that require minimum access, retaining it indefinitely and providing the capability of reading it or returning it to the active data base. (This definition closely follows the data-management industry definition for archiving.)

Archiving.--The systematic process of placing valuable data into archives to preserve and protect it from inadvertent change or loss by providing the necessary security measures and procedures. "Archiving" also implies that the data in archives will be indexed and can be retrieved. (Note that archiving is synonymous with the term "preservation" as used in the preceding recommendation for Issue 1, and elsewhere in this document.)

Archiving, when used with data in electronic form, does not solely mean a computer back-up procedure for protection in case of disk failure, or whatever. Rather, archiving is a set of procedures such that data (both in electronic and paper form) once approved, can be maintained without change in a secure and accessible environment. Revisions are allowed, but a record of the transaction must be maintained. The old, unrevised data are also retained.

"Placing data in archives," as used in this document, is the systematic and permanent storage of hydrologic records of historical value such that they reside in a system with suitable access, indexing, and uniformly high security. This process is contrasted with that in which records are saved. For example, saved records might be data on paper or electronically recorded that are stored in file cabinets in a data or project office, stacked in the corner of an office, or placed in boxes under a project leader's desk. Lack of security and accessibility can prevent these methods of storage from being considered archived, but a central reason is that this storage is non-systematic.

Some of the desirable characteristics of an archival system are:

- The data are on media that can be permanently maintained.
- Systematic archival procedures are established and maintained.
- Archiving is for an indefinite period.
- Data are readily accessible. First priority will be given to archiving data on a media that can be accessed directly by computers. When this is not feasible, data may be stored on other media.
- The data are preserved in a non-volatile state or one of extremely low volatility. (Volatility, in this sense, refers to the tendency of data, or the media on which it is stored, to deteriorate, rendering the data unreadable.)
- Data are verified to be accurate and complete before archiving.
- Changes can be made to the data, but a record of the transaction is archived.
- There is central coordination of the archival system, such that uniform policies, procedures, and guidelines are established for use by those archiving data, and adherence to the policies is monitored and required.
- Data are indexed before archiving.

Archiving of data is not the same as the saving of data, which would have some of the following characteristics:

- Storage is short-term or non-permanent.
- Media are difficult to maintain indefinitely and may readily deteriorate.
- Data are volatile.
- Organizational units use their own discretion in following broad or vague guidelines.
- Security of data is low.
- Accessibility is inadequate or uneven among data of the same kind.
- Changes may be made to the data without record of the transaction.
- Verification, to determine if the data are accurate and complete, is incomplete or nonexistent.

The Survey Manual, SM 432.1 and SM 431.9 (U.S. Geological Survey, 1991a and 1988, respectively), supplemented by Survey Manual Handbook 432-1-H (U.S. Geological Survey, 1990a), prescribes how records shall be managed. Disposition of hydrologic and related data is described in the document, "Water Resources Division Mission-Specific Records Disposition Schedule, October 1990," (U.S. Geological Survey, 1990b) which shows that electronically recorded records in the national Water Data and Retrieval System (WATSTORE), and presumably NWIS II, are permanent and will be preserved indefinitely. Other records are scheduled for destruction. It specifies that some records, however, are to be microfilmed and stored by USGS prior to destruction of the paper copies. These provisions cover most of the types of data identified in the recommendation below. Thus, the hydrologic data that the committee has identified for being placed in archives are, for the most part, consistent with the data designated as permanent or scheduled for indefinite retention by the USGS.

Recommendations on Issue 2

The archive for hydrologic data in electronic form will be NWIS II. Data files in NWIS II, other than the working and review files, will be preserved. Thus, all approved files are archived. This concept agrees with the document, "System requirements specification for the National Water Information System II," Open-File Report 91-525 (U.S. Geological Survey, 1991b), which addresses data protection and data aging. (Subsequent recommendations for Issue 3 show, in tabular form, examples of the types of data in electronic form that will be archived.)

In general, all data published in WRD reports, or used to support scientific analyses leading to conclusions in these reports, are archived. This report specifies what data--both in electronic and paper form--are to be archived and establishes when and why these data are to be placed in archives.

The committee strongly recommends that a cataloging or indexing system become an integral part of NWIS II so that data can be easily located. The system should enable the user to determine the type of data and the amount or period of record.

To prepare an overall policy on data archiving, the committee reviewed and modified a proposed policy statement that resulted from the activities of an earlier committee on data archiving in 1990. This policy will ensure a consistent national approach by WRD to the process of preserving original and other essential data by placing them in archives with appropriate security and accessibility.

Overall Policy on Data Archiving

What to archive.--It has long been WRD policy that all original field notes, measurements, and observations shall be preserved indefinitely. Table 1 contains lists, by type and discipline, of examples of those data, notes, and original calculations that are included.

Notes:

- Although these lists are fairly exhaustive, they are not intended to contain every possible type of hydrologic data that should be archived. Similar data collected in these and other hydrologic disciplines should also be preserved in archives.
- Some hydrologic data kinds listed below are currently on paper but might be collected or received in electronic form in the future. Also, data presently on paper may at some time be scanned into electronic form.

When to archive.--To specify when records should be archived, it is convenient to consider paper and electronic material separately. Paper documents may be saved (retained in suitable field facilities), for 6 months to 3 years after finalization. ("Finalization" is generally the publication of data or interpretive analyses.) After that, the paper documents must be placed in the archival system. Electronic records will be archived in NWIS II according to the schedules in tables 2-6, which appear subsequently in this document. Until NWIS II is available, paper copies of the records will be archived within 3 years after publication. Also, an electronic copy of the data shall be saved using the "archival" facility of ADAPS (or other records processing system) until NWIS II is available for archiving.

Archiving is an ongoing process. It is recommended that data be placed in archives as soon as is reasonably feasible. For example, records should be placed in the archival system soon after the resulting report is approved or after computation and review. Original hydrologic data in electronic form should be archived immediately after they are submitted to NWIS and prior to any editing or modifications.

Why Archive?--The reasons why records should be archived are mostly self-evident. They include:

- Legal requirements or potential needs in legal matters
- Data for future research and investigations
- Support of published reports
- Support of working and on-line computer data bases
- Security and accessibility of original data

Recent challenges to published hydrologic information have pointed to the need to retain field data to support conclusions, calculated values, and statistical determinations. Challenges to the veracity of published hydrologic information may be made tens or even hundreds of years after the data are collected.

How Data Shall Be Archived.--This document does not establish how data will be archived--neither in regard to the ultimate form and repository for paper records nor to the design of NWIS II. Various entities and work groups within WRD have and will deal with these issues. Nor is this document intended to serve as a manual or guide book for data management and retention. It is only intended to establish policy. Details of how the policy is to be implemented remain the responsibility of the user--at least until further guidelines are issued.

Archival Coordination.--Data archiving in WRD has been carried out in a distributed manner. With the advent of electronic data and the unlikely prospect of records centers continuing to serve as a permanent place for archiving some types of hydrologic data, it is now necessary to provide more central leadership in this area. The Assistant Chief Hydrologist for Program Coordination and Technical Support (ACH, PC&TS) will assume the responsibilities for national coordination of data archiving. The ACH, PC&TS will implement the policies contained herein, coordinate the data-archiving activities of the field and project offices, and monitor to ensure that all appropriate records are archived in the designated secure manner.

Archival redundancy.--Copies of the archived data should reside in at least two suitably safe locations. Copies will be remade and verified at appropriate intervals to ensure against loss of records resulting from media deterioration. Paper records and those copied on microfilm will remain in the records centers or in some other centrally located repositories, although standards may be developed such that some or all paper records may be archived in field offices. The activities will be administered by the ACH, PC&TS, who will offer continuing leadership and monitoring in data archiving.

Table 1. --Hydrologic data to be placed in archive.

[Data will likely be collected or received in electronic form (E), on paper (P), or both (P,E).]

Automated-Station Data
(Continuous record, or equivalent)

*Primary sensor data:
- Recorder charts (P)
- Digital-punch tapes (if not archived electronically) (P)
- Electronically recorded records and processed digital-tape records (E)
  1. Original data (E)
  2. Edited data (E)
- Observer's notes and readings
- Beginning and ending information for recorded data (P,E)

* These data include, but are not limited to: ground- and surface-water levels; stream velocities; water-quality characteristics, such as pH, conductance, temperature, dissolved oxygen, and suspended sediment concentration; and meteorological measurements, such as precipitation, wind speed and direction, temperature, humidity, solar radiation, and snow depth.

Secondary, or back-up, records are not generally archived unless subsequently used to replace or reconstruct primary records.

Supporting information:
- Station descriptions (P,E)
- Station analyses (P,E)
- Rating curves and tables (P,E)
- Transport curves and tables (P,E)
- Concentration as a function of time curves (P,E)
- "Box" coefficients data (P)
- Key computations (P,E)
- Shift records (P,E)
- Level notes (P)
- Field notes and observations (P)

Calculated data:
- Unit values (E)
- Daily, or longer period, values (E)

[Data will likely be collected or received in electronic form (E), on paper (P), or both (P,E).]
Surface Water

Current-meter discharge measurements (P)
Spin-test results (P)
Methods-development data (Not in NWIS II)
Laboratory data (P,E)
Crest-gage stage records (P)
Various field data (P,E)
Key calculations (P)
Analyses, photographs, and plots (P)
Time-of-travel data (P)
Dye-dilution data (P)
Miscellaneous field notes (P)
Verification data and computations:
    Channel-roughness coefficients (P)
    Bridge, culvert, and weir coefficients (P)
    Photographs, and stereoscopic slides (P)
Indirect Measurements:
    Field-survey notes (P)
    Computations (P)
    Photographs, and stereo slides (P)
Interpretive studies:
    Data and plots (P)
    Final statistical analyses (P)
    Final model parameters and variables (P)
    Final model version (P,E)
    Project description (P)
    Laboratory data (P)
    7Original field observations, notes, and measurements (P)
Cableway inspection forms (P)

Ground Water
Aquifer-test data (P) Drill-core descriptions (P) Drill-cuttings descriptions (P) Field notes, geologic and hydrologic (P) Geophysical logs (P,E) Pumping records (P,E) Soil-moisture data (E) Soil-temperature data (E) Surveying records (P) Spring schedules (P,E) Geologic maps (P) Water-level data sheets (P) Well-field maps (P) Well logs (P) Well schedules (P) Surface geophysical data (P,E)
Water Quality
Analyst notes for qualitative biological determinations (P) Analyst work sheet (district lab analyses) (P,E) Calibration readings for field instruments (P) Laboratory analyses (P,E) Sampling protocols (P,E) Field notes, including field determinations (P) Quality-assurance data (P,E)

[Data will likely be collected or received in electronic form (E), on paper (P), or both (P,E).]
Water Use
Field notes (P)
Supporting data (P,E)
Copies of withdrawal or discharge (P)
Questionnaires by cooperators (P,E)
Location maps (P)
Aggregate data (E)
Data-Documentation file (E)
Other
Project description forms (all years) (P,E)
Selected photographs (P,E)
Station descriptions (P,E)
Station operation records (P,E)
Geographic position system results (E)
Quality-assurance data (P,E)
Scanned original field notes and other data (E)

ISSUE 3--DATA TO BE PLACED ON LINE

Issue 3 addresses what minimal data set should be on line or near on line in NWIS II. (The term "near on line" as used herein refers to storage devices that are not on line for immediate retrieval but may be accessed automatically for retrieval that is only a little slower.) Although this question appears simple, three aspects must be considered: the need for a data base to be in NWIS II at all, the length of time a data base must be immediately accessible (and perhaps the length of time it might be accessed quickly but not immediately), and the needs for local and national accessibility.

To deal with these aspects, the committee included Tables 2-6 below. The tables, which deal almost exclusively with data in electronic form, show the relative need for accessibility for all principal types of data that will be included in NWIS II. The committee did not, however, completely dissect the details of the user group recommendations. Thus, some supporting data bases may not appear in the tables, but recommendations for related data will be indicative of their status.

The tables should not be construed as recommending duplicate data bases. The committee recommends against this approach and presumes that the data set will be in a singular system that may be maintained by changes, modifications, or revisions to that system alone. Data that require a high degree of security, however, such as that in archives, may be maintained separately and updated less frequently than working files.

Tables 2-6 show the number of years from date of collection, acquisition, or generation in which different types of electronic data should remain at various levels of accessibility. Both national (Nat.) and local (Loc.) needs are defined. National need refers to aggregated data bases accessible by anyone in the network; local need refers to regional data bases primarily accessible by those in the originating office or offices.

The columns in the tables are arranged from left to right in decreasing accessibility and increasing security. The column heading "Many References (Refer.), Many Changes," indicates the need for immediate access and that most of the data contained therein has the status of a "working file" as defined in the system requirements specification document (U.S. Geological Survey, 1991). Once data are approved, they will be moved periodically to a level of less accessibility. The heading, "Many Refer., Few Changes," indicates immediate access and an "approved file" with more security. These two categories, "Many Refer., Many Changes," and "Many Refer., Few Changes," have a recommended minimum level of accessibility characterized as "on line"; their being "near on line" or "off line" is not acceptable. "Few Refer., Few Changes," indicates that slower access is acceptable--a few minutes to an hour or so. This category has a recommended minimum level of accessibility characterized as "near on line". "Elec. Archive" indicates that a high level of security is needed, and that longer access time might be tolerated--a few hours to overnight, perhaps. This category has a recommended minimum level of accessibility characterized as "off line".

Needless to say, it would be ideal if the system can be designed to economically provide immediate access to all approved data.

Once data have moved from the level of accessibility entitled "Many Refer., Many Changes" to the other levels of accessibility and security, they are in a state of preservation. That is, those data have been placed in archives--even if they are on line or near on line instead of being off line, which is the minimum requirement for the level of accessibility entitled "Elec. Archive." The determination that data in these latter three levels of accessibility have been archived negates the need to archive paper copies, superseding previous requirements.

The committee did not attempt to provide design recommendations such as where the data should physically reside. Thus, the characterizations of the data bases as "on line," "near on line," or "off line," are intended to be explanatory of the accessibility requirements and not indicative of system design.

The specifications of time that data bases should remain in a particular accessibility category are judgmental and are given as a guide. Subsequent deliberation may call for some modification of these times.

It would be a highly desirable feature of the NWIS system if data could be maintained in a higher accessibility level than designated in tables 2-6, when requested by a project or district. It would not, however, be allowed for data to be placed in a lower level of accessibility where this would adversely affect national retrievals. For data that are only accessible locally, the data base administrator should have the capability to move it from one level to another.

The time designated as "0" in the tables is when project personnel actually have access to the data. It could be when the data are collected, received from a laboratory, or obtained from an outside source--such as some water-use or well-log information. The frequent use of 1.5 years as a time when data changes accessibility level reflects the requirements to publish the annual data report 6 months after the end of the water year. The designation of 1.5 years means the data should migrate to a different level of accessibility about the time the report is published.

Recommendations on Issue 3

Surface water.--

Needs for access to surface-water data are closely related to the goal of publishing data within 6 months of the end of the water year in which it was collected. Thus, the 1.5-year times, which frequently appear in table 2, correspond to the publication of the data and when the file can be moved to a data base with, perhaps, a higher level of security and/or a lower level of accessibility.

Table 2. --Recommended accessibility for surface-water data bases, in years.

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , an indefinite number of years; Yrs., Years; Orig., Original; GH, Gage Height; res., reservoir; Edit., Edited; Correct., Corrections; Descrip., Descriptions; UVM, Ultrasonic Velocity Meter; rev., time of revision; Q, Discharge; Char., Characteristics; Do., ditto.]

Data type	Nat'l retrieval	Local retrieval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Daily discharge	x			1.5-∞			Data are on line and archived after 1.5 years.
Daily discharge		x	0-1.5	1.5-∞			
Original unit GH data		x				0-∞	Includes stage-only and reservoir-level data
Edited unit GH data		x	0-1.5			1.5-∞	Includes stage-only and reservoir-level data
GH and datum corrections		x	0-1.5			1.5-∞	

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , an indefinite number of years; Yrs., Years; Orig., Original; GH, Gage Height; res., reservoir; Edit., Edited; Correct., Corrections; Descrip., Descriptions; UVM, Ultrasonic Velocity Meter; rev., time of revision; Q, Discharge; Char., Characteristics; Do., ditto.]

Data type	Nat'l retri-eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Unit discharges		x	0-1.5			1.5-∞	
Ratings, analyses, shifts, and supporting information		x	0-1.5			1.5-∞	
Station description		x		0-rev.		rev.-∞	Current description on line; past in archive
Velocity data--UVM, vane, and similar		x	0-1.5			1.5-∞	
Lists of Q measurements		x	0-1.5	1.5-∞			
Level notes or summaries							Not generally electronic
Original Q measurements							Not electronic
Precipitation and temperature		x	0-?				Needed on line during project
Site data		x	0-1.5				
Site data	x			1.5-∞			
Peak data		x	0-1.5				
Peak data	x			1.5-∞			
Basin characteristic	x	x			0-∞		

Ground water.--

The committee identified the minimum ground-water data set for on-line or rapid access, as indicated in table 3. The structure is based on the assumption that the need exists, both locally and nationally, to quickly identify site locations where data are available. Most data retrievals at the national level are assumed to not require immediate access and may best be stored in a lower level of accessibility, with periodic (annual) updates, in contrast to the frequent and rapid data access and update capabilities that are mandatory at the local level for the data base to be useful to district operation.

The Ground Water Site Inventory (GWSI) data base has the potential to store over 300 different types of data, all of which fall into eight major categories. Data types identified in table 3 are based on these eight major data categories, as currently (1992) exist in the GWSI data base. In identifying these data categories and access levels, the committee did not refer to the systems requirements specifications document to determine if all data types specified in the NWIS-II data base structure were addressed in table 3. It was assumed that the ground-water data bases will be parallel enough in structure and content that the suggestions made by the committee can be easily interpreted and, where necessary, extended to data groups not currently in GWSI but to be considered by the NWIS-II design team.

As indicated in table 3, ground-water data will be placed in the appropriate storage and access categories for the life of the data base; data will not be periodically down loaded to less accessible storage media as time progresses. This design requirement is dictated by the basic differences in which ground-water data are collected and used in WRD studies. Ground-water data are spatial and temporal in nature and require the ability to add information to and update the data base as needed. For example, little more than location information might be available for a given site, but when an investigation is begun in the area, it might be possible to add information on water-levels, well construction, and geophysical logs to the data base. This concept is in clear contrast to surface-water data where, once records are computed and published, the data are rarely changed and can be placed in electronic archives. All data, however, that have been moved to any of the three lower levels of accessibility, as indicated in the table below, are considered to have been placed in archives.

Table 3. --Recommended accessibility for ground-water data bases, in years

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , indicates an indefinite number of years; Constr., Construction; Orig., Original; WL, Water Level; Edit., Edited; GW, Ground-Water; Geohydr., Geohydrologic; Obser., Observation; Do., ditto.]

Data type	Nat'l retri--eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Site data	x			0-			
Site data		x	0-				
Well construction	x			0-∞			
Well construction		x		0-∞			

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , indicates an indefinite number of years; Constr., Construction; Orig., Original; WL, Water Level; Edit., Edited; GW, Ground-Water; Geohydr., Geohydrologic; Obser., Observation; Do., ditto.]							
Data type	Nat'l retri--eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Original unit water level data		x				0-∞	
Edited unit water level data		x	0-1.5			1.5-∞	
Periodic water level data	x			0-∞			
Periodic water level data		x		0-∞			
Ground-water discharge	x				0-∞		
Ground-water discharge		x	0-∞				
Geohydrologic logs	x				0-∞		
Geohydrologic logs		x		0-∞			
Miscellaneous	x				0-∞		
Miscellaneous		x		0-∞			
Observation well information	x				0-∞		
Observation well information		x		0-∞			
Hydraulic data	x				0-∞		
Hydraulic data		x		0-∞			

Water quality.--

The entries in table 4 are meant to be indicative of the accessibility needs for all classes and attributes of water-quality data currently listed in the system requirements specifications (U.S. Geological Survey, 1991) for NWIS II. These requirements should also apply to similar biological data and to some types of sediment data. The following describes the row labels in table 4.

- Header information refers to station identification, date, time, who collected, how collected, and similar information regarding the sampling event.
- Sample management system (sms) information refers to the log-in data sent to and from the laboratory about the sample such as date sent, date received, received condition, requested analyses, and similar information.
- Sample types identifies whether this is an environmental sample, quality-assurance sample (blank, spike, duplicate, or other) or similar sample-type information.
- Constituents refers to the information on concentrations of a constituent, accompanied by method used, significant figures, and other identifying information.
- Qualifiers refers to information that might apply to a constituent, groups of constituents, events, or groups of events. This information provides an explanation of circumstances that might have affected the analyses.
- Sequential analyses refers to multiple analyses, resulting from either laboratory- or user-requested reruns. Typically, one of these values is chosen or derived to represent the most correct value and is so designated in "constituents."
- Laboratory data, NWQL, and other, refer to the original analysis data, as received from the laboratory, before any modifications or deletions are made. These are placed in archives as a record of the original laboratory determinations.

Table 4. --Recommended accessibility for water-quality (including biologic) data bases, in years

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , indicates an indefinite number of years; QW, Quality of Water; sms, sample management system; Anal., Analyses; Orig., Original; Auto., Automated; Edit., Edited; Descrip., Description; NWQL, National Water Quality Laboratory; Do., ditto.]

Data type	Nat'l retri-eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Quality water data header	x	x	0-1.5	1.5-∞			
Log in (sample management system)		x	0-1.5	1.5-3		3-∞	
Sample types	x	x	0-1.5	1.5-∞			
Constituents	x	x	0-1.5	1.5-∞			
Qualifiers	x	x	0-1.5	1.5-∞			
Sequential analyses		x	0-1.5	1.5-∞			

[Abbreviations: Nat., National; Loc., Local; Refer., References; Elec., Electronic; x, indicates whether accessibility is local or national; , indicates an indefinite number of years; QW, Quality of Water; sms, sample management system; Anal., Analyses; Orig., Original; Auto., Automated; Edit., Edited; Descrip., Description; NWQL, National Water Quality Laboratory; Do., ditto.]

Data type	Nat'l retri--eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Gilogical data	x		0-1.5	1.5-∞			
Original unit values		x				0-∞	Automated monitor data
Edited unit values		x	0-1.5			1.5-∞	Automated monitor data
Daily values	x			1.5-∞			
Daily values		x	0-1.5	1.5-∞			
Station description		x		0-rev.		rev.-∞	Current description on line; past in archive
Field note sheets							Not electronic
Laboratory data							Transmitted to Districts
National Water Quality Laboratory		x				0-∞	
Other		x				0-∞	

Sequential analyses, which will only be kept locally, should be accessible for use at the same level as the constituents and sample type. Qualifiers must be available nationally at the same level of accessibility as constituents and sample type, because this information explains anomalies in the data or unusual circumstances in the collection of the sample that would be of interest to the user. Non-environmental (quality-assurance) sample results--such as spikes, blanks, and splits--also should be available nationally as well as locally, so WRD will have the capability to make much better use of these data than in the past.

Sediment.--

The general needs for access to data bases containing data used for computation of daily sediment records are defined in table 5. Like those for surface-water-discharge data, the needs reflect the annual cycle of data collection, computation, and publication.

Table 5. --Recommended minimum accessibility for sediment data bases, in years.

{Abbreviations: x, indicates whether accessibility is local or national; , an indefinite number of years.}

Data type	Nat'l retri- -eval	Local retri- eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Suspended Sediment:							
Daily Data	x			1.5-∞			
Daily Data		x	0-1.5	1.5-∞			
Field notes		x					Not electronic
Lab analyses	x	x	0-1.5	1.5-∞			
Cross-section coeffi- cients		x	0-1.5			1.5-∞	
Transport curves			0-1.5			1.5-∞	
Concentration curves			0-1.5			1.5-∞	
Computations			0-1.5			1.5-∞	
Bedload:							
Daily Data	x			1.5-∞			
Daily Data		x	0-1.5	1.5-∞		rev.-∞	
Field notes		x					Not electronic
Lab analyses	x	x	0-1.5	1.5-∞			
Computations		x	0-1.5	1.5-∞		0-∞	If electronic

Water use.--

Water-use data are collected and compiled annually at the field level. Every 5 years there is a compilation at headquarters to prepare a national report. The needs for data accessibility associated with these activities are defined in table 6.

Much of the water-use data that support WRD publications are compiled or collected by others. If not archived or published elsewhere, this information shall be preserved indefinitely. If these data are received in electronic form, they shall be preserved (archived) in NWIS II. If received in paper form, they shall be placed in archives, according to WRD policies.

Table 6. --Recommended accessibility for water-use data bases, in years.

{Abbreviations: x, indicates whether accessibility is local or national; , an indefinite number of years.}

Data type	Nat'l retri--eval	Local retri-eval	Decreasing accessibility ⇒				Remarks
			on line		near on line	off line	
			Man refer. many changes	Man refer., few changes	Few refer., few changes	Elec. archive	
Aggregate data	x			1-5	5-∞		Compiled every 5 years
Aggregate data		x	0-3	3-∞			
Site-specific data		x	0-2	2-∞			

General Concepts

Although tables 2-6 address accessibility by WRD users, they do not address access by others. For example, as will be discussed subsequently in relation to issue 4, non-WRD data can reside in the NWIS II data bases that are not proprietary but, nevertheless, should not be released, such as that in non-approved working files to the public. Procedures are needed to permit selective public access to the data bases.

One major item to consider is data ownership, as defined by collector or measurer of the data. It is important to define the WRD data for which WRD has an obligation to store and distribute electronically. If hydrologically related data from outside sources are used in WRD publications and are not published elsewhere, we have an obligation to be able to provide these data to the interested public and to be able to reproduce these data to support our published conclusions.

We recommend defining WRD data as that collected for or by WRD programs. NWIS II files that are publicly accessible should contain these WRD data. The data should be retrievable by specified WRD program or by areal aggregation.

All WRD data stored for eventual distribution must be associated with enough ancillary information to allow proper interpretation. Identifying the program or the project, for which data are collected, provides a general understanding of the nature of the data and differentiates WRD data from that of other agencies.

The information contained in tables 2-6 is intended as a guide for design implementation and not to establish procedures for data handling. Related to the design of NWIS II, however, are strong implications for data-handling procedures.

ISSUE 4--STORAGE OF NON-U.S. GEOLOGICAL SURVEY DATA

Issue 4 deals with what data shall be stored in NWIS II in addition to data collected by WRD, such as that from other agencies and ancillary data. The committee was advised that SPG had agreed that the two examples of ancillary data--spatial data coverages and hypothetical digital-model results--shall not be placed in the NWIS II data base. Evidently, this type of information will be placed in the other data bases designed to interface with the NWIS system. Therefore, the committee only considered the issues of whether data from other agencies and results of computational procedures used to quantify actual hydrologic events shall be placed in NWIS II.

The only model results that are presently (1992) acceptable in NWIS II are surface-water discharges computed with the USGS BRANCH model. Results from calibrated ground-water or a water-quality model are, for purposes of this recommendation, considered hypothetical data and, thus, not allowable in NWIS II.

Recommendations on Issue 4

The current policy in WRD is that all water data collected as part of the routine data collection of the WRD (both basic and project data) must be stored in the computer files of the National Water Information System. One purpose of this policy is to enable all WRD work to be verifiable and repeatable to the greatest extent possible at any time in the future.

The committee recommends that this policy be expanded for inclusion of: Any hydrologically related data collected by another agency--either unchanged or modified--that is published in or used for scientific interpretations or analyses published in a WRD report shall be placed in a NWIS II data base, if appropriate data-base capabilities exist, unless the data are otherwise preserved by publication or placed in archives elsewhere. The system should have the capability to store:

- Designation of the source of the data.
- Designation of modification, acceptance of correctness, or verification (if any) by person, date, and method.
- Entry of the source, modification, acceptance, and verification information for data collected by others should be mandatory.
- When non-USGS data are retrieved, they should be clearly identified as being collected by others.

This policy will allow project personnel to modify data to meet project needs or otherwise improve its utility and to make these changes apparent to subsequent users, who would be aware that the data were derived from a non-USGS source but are changed in some way. The system will date the time the data were considered correct by WRD. Thus, even though the originating agency makes further changes to the data in their data base, it is clear that the data in NWIS II is that which was considered correct as of the designated date.

Other data, such as that furnished by cooperators or other agencies but not appearing in a WRD report, could be placed in the NWIS II data base at the discretion of the local WRD office. The quality of some of these data may meet WRD standards and could be released to the public as part of WRD's data set, some may not. NWIS II should be a system that will accept all sources of data, leaving the question of quality and usability to the WRD hydrologist using the data. Therefore, entry of the data into the system should not automatically make it part of the publicly accessible WRD data set that may be released to requestors without regard to the source or quality. A designation should exist for other agencies' data that are not to become part of the publicly accessible WRD data set, such as non-approved data in working files. (Actually such data might be ultimately available to the public, if specifically requested, under the provisions of the Freedom of Information Act.)

Data that have not been approved by WRD should be available only through an explicit request for that type data made by parties with authority to retrieve it.

Historically, the storage of data in WATSTORE data bases has constituted "publication," unless the data are declared proprietary. Because of this, WRD generally has stored data from cooperators and other agencies only by mutual consent. NWIS II, however, should permit the entry of data, regardless of the source or status of publication, but if the data are private or unpublished, it should not be placed in a publicly accessible state without prior consent from the organization that collected it.

Hydrologic data, resulting from a USGS-approved model or computational procedure, used to generate values for publication that represent actual hydrologic conditions may be entered into NWIS II. Results generated to simulate hypothetical conditions will not be entered into NWIS II. Presently, the only permissible data would be the results of the USGS BRANCH model used to compute flow in a stream reach. Predictive models that answer "what if..." questions are considered hypothetical, and their results shall not be allowed in NWIS II.

Ancillary data used by WRD that does not fit into NWIS II will be stored separately and may be indexed in the Master Water Data Index in NWIS II. Minimal description for the data should be included, just as would be required for its use--such as source, type, location, date of collection, value, and units of measurement.

ISSUE 5--POLICY ON ACCESS TO DATA BASE

Issue 5 deals with the need for a policy on who should have access to data stored in NWIS II, including provisional data, with particular emphasis on outside users. Policies were established in the past that addressed this issue in regard to NWIS I, the data bases residing on the Prime[1] computers, and other computers of the Distributed Information System. Generally, the committee followed these earlier policies by adapting them to NWIS II.

[1]The use of brand names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

Recommendations on Issue 5

Direct access to the NWIS II data base.--

- The protection scheme and access rights granted in the system requirements specification document (U.S. Geological Survey, 1991) should be adopted. These provisions include both WRD and non-USGS users.
- Cooperators and other users, seeking local data, may be granted access to specific files by districts or projects at the discretion of the appropriate manager. These users can be afforded access rights similar to WRD users, as deemed appropriate by the local office. As a matter of policy, this access is for local data only.
- Other users (and the users above) seeking national data should obtain access through the National Water Data Exchange (NAWDEX). These users would include registered NAWDEX member organizations and the general public. The NAWDEX and DIS II staffs should take the lead in determining the resources

needed to support this access and to establish reimbursement procedures and user priorities.

SUMMARY

A data-policy committee was formed in the late spring of 1991 at the request of the Strategic Planning Group (SPG) to consider five issues important to the design of a new National Water Information System (NWIS II), which will provide electronic-data storage, management, access, and protection for most of the hydrologic and related information collected or compiled by the U.S. Geological Survey (USGS), Water Resources Division (WRD). The five issues, part of which also apply to records on paper or in other form, and key recommendations of the committee are:

1. What are original data? The committee must define the term.

Original data--from automated data-collection sites, laboratories, outside sources, and non-automated field observations--are those data that should be preserved unmodified as collected or received. Once in conventional units, these data should be preserved, unmodified and unedited.

Based on legal opinions and review of the Federal legal code, the committee has determined that photocopies, or copies produced by other processes, where an exact image or copy is produced without the chance of human error occurring, constitutes an "original" copy for legal purposes. This would include copies of electronically recorded data on magnetic tape or similar device, data transferred to microfilm, and scanned images copied to an optical disk.

2. What original data collected by the WRD (such as 15-minute stage data) and other data stored in WRD data bases should be placed in data archives? The committee should draft a policy.

All original data that are published or used to support published scientific analyses shall be placed in archives. Specific examples of the data that shall be placed in archives are listed in the body of the report.

The WRD should establish central coordination of archiving original hydrologic data to ensure that valuable, irreplaceable data are safely preserved in accordance with the USGS Files Management Program provisions and the WRD Mission-Specific Records Disposition Schedule.

3. What minimum data set shall be maintained on line or available for immediate access at all times?

The body of the report lists, by general hydrologic discipline, examples of the data that will be in NWIS II and show the required levels of accessibility as the data become more final and are referred to less often. Data in the data base should generally move from levels with high accessibility and low security to levels with lower accessibility and higher security as they become finalized and need for access diminishes. As this occurs, data could be placed near on line or off line, as appropriate.

4. What data should be stored in NWIS II in addition to data collected by WRD, such as: (a) water data collected by other agencies or (b) ancillary data, including spatial data coverages and digital model results?

All data collected by others shall be placed in the data base, if used to support published USGS documents and not published or archived elsewhere. Other outside data may be placed in the data base at the appropriate manager's option. It is not anticipated that spatial data coverages and model results will be part of the data base, except for the results of the BRANCH model to simulate streamflow.

For data collected by others, the committee recommends that the capability exists in NWIS II to designate the source of the data and its modification, acceptance of correctness, or verification (if any) by person, date, and method. This information should be mandatory, and when the data are retrieved, it should be clearly marked as being collected by others.

5. Who should have access to data stored in NWIS II? The committee should draft a policy defining who should have access to the NWIS II data base, including provisional data, with particular emphasis on outside users.

A protection scheme, granting access rights to WRD personnel and other users, will be part of NWIS II. The general public will have access to all approved, non-protected files through the National Water Data Exchange system. Cooperators and others may be granted more direct access to local files at the discretion of appropriate manager. These provisions generally follow the established WRD policies in regard to access to hydrologic data stored in the National Water Information System.

REFERENCES

Schaffranek, R.W., Baltzer, R.A., and Goldberg, D.E., 1981, A model for simulation of flow in singular and interconnected channels: Techniques of Water-Resources Investigations of the U.S. Geological Survey, Book 7, Chapter C3.

U.S. Geological Survey, 1988, Information storage and retrieval technology--micrographics: U.S. Geological Survey Manual 431.9.

____1990a, Handbook for managing USGS records (432-1-H) Administrative Division report, U.S. Geological Survey.

____1990b, Water Resources Division mission-specific records disposition schedule, October 1990, supplements Handbook for managing USGS records: U.S. Geological Survey unnumbered report.
____1991a, Files Management Program: U.S. Geological Survey Manual 432.1.
____1991b, System requirements specification for the National Water Information System II: Open-File Report 91-525, 626 p.