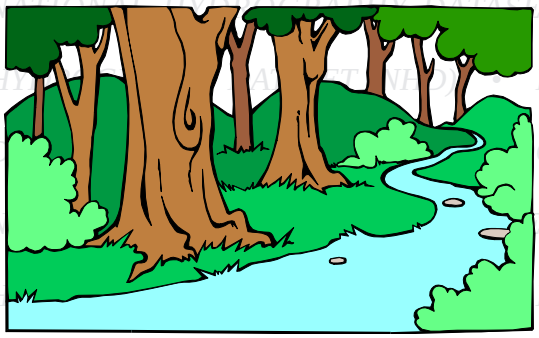
Center for Environmental Information and Statistics

US Environmental Protection Agency

Major Findings from the CEIS Review of the EPA and USGS

NATIONAL HYDROGRAPHY DATASET (NHD) DATABASE



July 13, 1999

Major Findings from the CEIS Review of EPA's NHD Database

1. INTRODUCTION 1

2. SUMMARY ANSWERS TO REVIEW QUESTIONS 2

3. DETAILED ANSWERS TO REVIEW QUESTIONS 4

3.1. What does the database cover? 4
3.2. Can the database be used for spatial analysis? 6
3.3. Can the database be used for temporal analysis? 6
3.4. How consistent are the variables over space and time? 7
3.5. Can data from NHD be linked with information from other databases? 7
3.6. How accurate are the data in NHD? 7
3.7. What are the limitations of NHD? 7
3.8. How can I get information on NHD? 8
3.9. Is there documentation on NHD? 8

1. INTRODUCTION

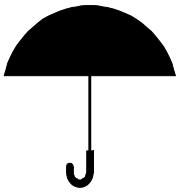
Key components of watershed assessment require the effective organization and management of a range of watershed information. The National Hydrography Dataset (NHD), a joint effort of the U.S. Environmental Protection Agency (EPA) and the United States Geological Survey (USGS), is the latest in a series of hydrological databases pertaining to the surface waters of the continental United States and Hawaii that provides a consistent national framework for managing watershed data. The NHD provides comprehensive coverage of hydrographic data for the United States. It uniquely identifies and interconnects the stream segments or "reaches" that comprise the nation's surface water drainage system. It also contains information for other common surface water bodies (e.g., lakes, reservoirs, estuaries, and coastlines). The NHD provides a digital referencing system for citing the location of specific points on streams and coasts of the United States, allowing users to move along the reaches and encounter water-related entities in hydrological sequence as they move up or downstream. Some anticipated end-user applications of the NHD are hydrological modeling and other water-quality analyses.

The NHD is the culmination of recent cooperative efforts of EPA and the USGS. In 1994, the USGS and EPA formalized a commitment to merge their hydrography files through a Memorandum of Understanding. The resultant National Hydrography Dataset combines the best of these Agencies' hydrography files including hydrological ordering, hydrological navigation for modeling applications, the unique identifiers for surface water features (i.e., reach codes) from EPA's Reach Files, and the spatial accuracy and comprehensiveness of USGS's Digital Line Graph (DLG) hydrography files. By incorporating these systems and data, the National Hydrography Dataset is more comprehensive, powerful, and useful than any of its components, and is designed to allow continued expansion and improvement by user-contributors.

This merger of complementary systems coincides with the Federal Geographic Data Committee (FGDC)-led efforts to develop the National Spatial Data Infrastructure (NSDI). The NSDI encompasses policies, standards, and procedures for all levels of government, the private and nonprofit sectors, and the academic community to promote production and sharing of geographic data. As part of the dynamic nature of the NHD, it will be made available via the internet.

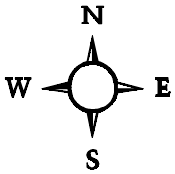
Because the NHD is designed to be more flexible and comprehensive than RF3, once the NHD is complete, the EPA will no longer support RF3. The balance of this review will focus on issues pertaining to the NHD.

2. SUMMARY ANSWERS TO REVIEW QUESTIONS



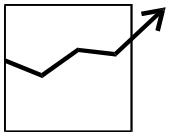
2.1. What does the database cover?

The National Hydrography Dataset (NHD) provides comprehensive coverage of hydrographic data and is a digital referencing system for citing the location of specific points on streams in the United States. The initial version of the NHD will contain more than 3,000,000 reaches. A reach is a segment of a river or stream between two significant tributaries and is made up of surface water features along the national surface water drainage network. The NHD will allows users to navigate upstream and downstream and encounter water-related entities in hydrological sequence along the network.



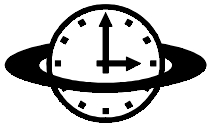
2.2. Can the database be used for spatial analysis?

NHD is useful for spatial analyses due to the provision of consistent latitude/longitude coordinates and unique identifiers for each reach. Once linked to the NHD by their reach address, the upstream/downstream relationships of water-related entities and any associated information about them can be analyzed using software tools ranging from spreadsheets to Geographic Information Systems (GIS).



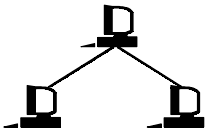
2.3. Can the database be used for temporal analysis?

The NHD provides spatial and other descriptive data for U.S. surface waters. For the most part, this information is not time-variant. However, NHD data can be used to provide an accurate and consistent basis for temporal analyses of water-related data from other systems.



2.4. How consistent are the variables over space and time?

While there is spatial consistency from reach to reach and from one area within the United States to another, there are inconsistencies in the density of the stream networks in different regions that were inherited from the USGS Digital Line Graph (DLG) data. The NHD will, however, be expanded over time to include a greater number of reaches. As mentioned above, NHD data are, for the most part, not time variant.



2.5. Can data from NHD be linked with information from other databases?

A primary objective for developing reach files like the NHD and its predecessors was to provide a common framework for interrelating data contained in many environmental databases. Integrating data from many surface-water related databases is possible by using reach codes and the surface water drainage network contained in NHD.

2.6. How accurate are the data in NHD?

The degree of locational accuracy of the coordinates representing the surface waters contained in the NHD is 40 meters. This accuracy is inherited from the USGS DLG data.

2.7. What are the limitations of NHD?

The NHD is limited primarily by the scale of the maps used in its development, which has resulted in inconsistencies in the density of the stream networks from one region to another within the U.S. However, plans for continuing updates and improvements to the NHD will eliminate most of these inconsistencies over time.

2.8. How can I get information on NHD?

Once available, the NHD data will be accessible via the internet.

2.9. Is there documentation on NHD?

Documentation on NHD will be available on a National Hydrography web site that is currently under development. For the time being, there is a web site created primarily as background and guidance for those producing the database. It contains limited documentation and project status information (http://nhd.fgdc.gov/).

3. DETAILED ANSWERS TO REVIEW QUESTIONS

3.1. What does the database cover?

The first full implementation of the Reach File concept was Reach File 1 (RF1), which was completed in 1982. RF1 was created from scanned 1:500,000-scale National Oceanic and Atmospheric Administration (NOAA) maps and consisted of approximately 68,000 reach segments comprising 650,000 miles of streams. RF1 still supports broad-based national applications. During 1987 and 1988 the Feature File of the U.S. Geological Survey's (USGS) Geographic Names Information System (GNIS), which contains information about physical and cultural geographic features in the United States, was used to add one new level of streams to RF1. The resulting file, known as RF2, contained 170,000 reach segments. Development of the latest Reach File, RF3, began in 1989 and was completed in 1992. The RF3 production process involved the overlay of the RF2 file, the GNIS II Feature File, and the USGS 1988 1:100,000-scale Digital Line Graph Version 3 (DLG3) hydrography data. RF3 contains 3,200,000 reaches.

The successor to RF3, the National Hydrography Data Set (NHD), is a joint effort between EPA and the USGS. This reach file is also based initially on the 1:100,000-scale DLG, but it has been designed to accommodate and encourage the development of the higher resolution data that many users require. The USGS DLG files contribute a national coverage of millions of features, including water bodies such as lakes and ponds, linear water features such as streams and rivers, and point features such as springs and wells. These files provide standardized feature types, delineation, and spatial accuracy. EPA's RF3 contributes hydrographic sequencing, upstream and downstream navigation for modeling applications, and unique identifiers for surface water features, or reach codes. Because the NHD is based on the same scale as RF3 and will be more comprehensive, EPA will no longer support RF3 once the NHD is complete.

A reach is defined as a segment of a river or stream of relatively constant size (flow) and composition between two significant tributaries. (A stream or river changes in size at the point of inflow of each significant tributary.) The segment of a stream below its point of origin and its first significant tributary is also defined to be a reach for the purposes of the NHD. Surface water features along the national surface water drainage network are combined into reaches that have unique codes. The reach codes then provide an addressing system for linking water-related information to the network. For example, reach codes are used to specify points that are on or near a reach, such as water quality monitoring sites; linear entities, such as stream segments, for which fish consumption advisories have been reported; and areal entities, such as sub-watersheds, that contain entire sub-networks of reaches.

NHD will facilitate integration of data from users at the national, state, and local levels. Permit writers, emergency management personnel, and other environmental managers will be able to identify water-related entities in hydrological sequence along the network when assessing the causes or implications of actual or potential pollution events.

The goal of the NHD design is to combine DLG accuracy and cartographic qualities and Reach File application orientation with enhancements to facilitate use in Geographic Information Systems (GIS) applications. Unlike Reach File 3, the NHD is designed to promote shared long-term maintenance and enhancement by users of the database.

Who Must Report?

There is no required reporting specifically to the NHD. The Reach Files were developed by EPA's Office of Water to provide a national database that provides common nomenclature for Federal and State reporting of surface water conditions. Additionally, it was to provide a framework for consistent national, Regional, and State analyses of water quality, wasteload allocations, and critical waterbody assessments such as those called for in Section 305 of the Clean Water Act.

How Are Data Reported?

The NHD has been designed to enable updating and maintenance by user-contributors. New reaches may be added to the database, as well as revisions and/or additions to data associated with reaches that are already in the database (e.g., basic features that create reaches). The updating process will rely on the contributions of Federal, State, and local users. Common identifiers for features are the basis for tracking and sharing deletions, additions, and modifications of features during maintenance, and are used to communicate and share corrections among organizations. NHD update guidance and capabilities will be provided incrementally over time, starting with simple attribute updates, such as correcting or adding reach names, and progressing to more complex spatial edits. As part of this process, higher resolution hydrography data (e.g., all water features on the USGS 1:24,000-scale maps) will also be integrated into the NHD.

Data Elements

NHD contains a number of features and attributes that describe the data. The basic identification variables include the reach code and reach name. Assigned to each reach is also a number of attribute variables, or features, that create the reach or can be found when navigating upstream or downstream along the reach. There are forty-three types of hydrographic features for 1:100,000-scale data in the NHD, and ten additional feature types will be added for 1:24,000-scale data. Additional feature types may be added in the future to accommodate data of higher resolution.

The reach code is the unique identifier assigned to each surface water reach. The reach code contains a cataloging unit (a unique numerical code assigned by the USGS to a watershed) and a segment (a unique four digit number assigned by the USGS to each reach within a given cataloging unit).

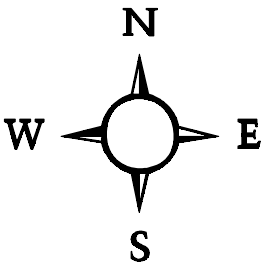
The reach codes uniquely identify individual components of the nation's rivers and lakes, providing a common nomenclature for Federal and State reporting of surface water conditions. This provides a common link for assigning reach addresses to water-related entities, such as industrial dischargers, drinking water supplies, fish habitat areas, and scenic rivers. Reach addresses establish the

locations of these entities relative to one another within the NHD surface water drainage network in a manner similar to street addresses. Therefore, the NHD can be combined with any other database containing reach codes, thus establishing the upstream and downstream relationships of these data.

Additional variables for each reach include coordinate entry attributes, navigation attributes, and names and physical descriptions of water bodies.

Coordinate attributes comprise latitude and longitude values that define the start, end, and internal shape points for each reach. Together, these values define a given coordinate entry to four decimal places for latitude and longitude and assign the coordinate entry a unique numerical identifier that ties back to the original source of information. Coordinates also possess a relational attribute that indicates point ordering from downstream to upstream.

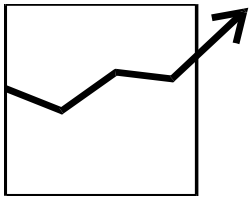
Navigational attributes define the connected stream network and provide connectivity despite the presence or absence of topologic continuity of the network's spatial representation. These attributes enable the NHD to provide hydrological ordering (the upstream or downstream sequence of events or conditions encountered in the stream network), as well as a graphical indication of the stream network, allowing users to track events either upstream or downstream. Traverse through open waters (e.g., wide rivers, lakes, and reservoirs) is also supported.



3.2. Can the database be used for spatial analysis?

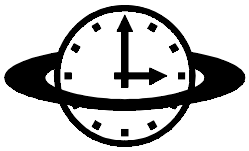
The inclusion of at least two sets of latitude and longitude values for each reach makes the National Hydrographic Dataset well-suited for use in spatial analyses. In addition, once linked to the NHD by their reach addresses, the upstream/downstream relationships of water-related entities and any associated information about them can be analyzed using software tools ranging from spreadsheets to GIS. GIS can also be used to combine NHD-based network analysis with other data layers, such as soils, land use, and population. Much of the value of the NHD lies in several unique characteristics of the data:

- Since hydrological connectivity is carried in its attributes, NHD maintains connectivity where topological discontinuity exists in the original data sources;
- Coordinates have been consistently ordered from downstream to upstream; and
- The system contains the names for major rivers, streams, and lakes.



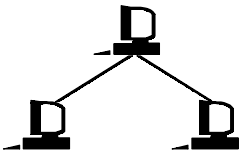
3.3. Can the database be used for temporal analysis?

The NHD provides spatial/locational data for U.S. water bodies. Most of this information is not time-variant. The data can be used to provide an accurate and consistent framework for temporal analyses of various kinds (using water quality and/or water flow data from other sources).



3.4. How consistent are the variables over space and time?

There is a high degree of spatial consistency from reach to reach and from one area within the United States to another, except with regard to the density of the stream network, which varies considerably. While RF3 has remained invariant since it was created, the NHD is designed to allow for expansion (to include higher levels of hydrography and additional reaches) and to allow for corrections and additions to the databases for reaches already contained in it. Ongoing improvements will gradually reduce the degree of inconsistency in the density of the stream network from one area to another.



3.5. Can data from NHD be linked with information from other databases?

One primary objective for developing the Reach Files and the NHD was to provide a common framework for interrelating data contained in the many EPA environmental databases. By linking data using reach codes and other data contained the NHD, it is possible to analyze data obtained from other data sources, and to enrich that analysis by associating upstream information with that from downstream. For example, pollutant discharges, described in the EPA's Permit Compliance System (PCS), may be related to downstream domestic water supply intakes stored in the EPA's drinking water system, and these data may be related to upstream and downstream water quality monitoring data from the EPA's STORET water quality database. Prior to the existence of EPA's Reach Files, it was extraordinarily difficult to develop such an analysis because the information was collected by separate organizations and had no common integrating framework.

The comparability of the NHD with other databases and the ability to make use of it to integrate data from other databases are determinations that must be made on a case-by-case basis. This is done by comparing the characteristics of the NHD with the corresponding characteristics of the other databases.

3.6. How accurate are the data in NHD?

The degree of locational accuracy of the coordinates representing the surface waters contained in the NHD is 40 meters. This is the accuracy inherited from the USGS DLG data which conforms to national map accuracy standards.

3.7. What are the limitations of NHD?

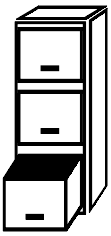
RF3 has been limited primarily by the scale of the maps used in its development, which may not provide the level of detail some users may want. In addition, the density of stream networks varies considerably from one region to another within the U.S.

While the initial version of the NHD will be drawn from the latest 1:100,000-scale DLG files, the NHD will gradually be expanded using the additional data

available on 1:24,000-scale maps. Additional improvements to the NHD, through the continuous updating and revision process by users, will incorporate higher levels of hydrography and additional reaches and should eventually eliminate most of these inconsistencies.

3.8. How can I get information on NHD?

Data from the NHD are not yet available, but will be available soon via the National Hydrography Dataset web site that is currently under development.



3.9. Is there documentation on NHD?

Documentation on the National Hydrography Dataset will be available on a National Hydrography web site that is currently under development. Once complete, this site will include: an informational document on NHD concepts and content; instructional documents on implementing the NHD in ARC/INFO; NHD software; NHD fact sheets; a tutorial series on how to use NHD; the data maintenance and update guidance; technical support; and an opportunity for user feedback.

There is also a web site that was created primarily as background and guidance for those producing the database. The site contains limited documentation and project status information. This information can be found at http://nhd.fgdc.gov/