NUTRIENT STATIONS GEOREFERENCED TO NHD: VALUE TO NUTRIENT CRITERIA DEVELOPMENT

Ifeyinwa Davis, U.S. Environmental Protection Agency
William Cooter, RTI International
James Rineer, RTI International
James Sinnott, RTI International
Randy Dodd, RTI International
Bill Arnold, Great Lakes Environmental Center

Abstract

The U.S. Environmental Protection Agency (USEPA) Nutrient Criteria Program developed a prototype Web-based database (NUTDB) and ArcIMS mapping application to support states and tribes in developing nutrient criteria recommendations for surface waters. Monitoring station nutrient data were compiled from multiple sources (EPA Legacy STORET, USGS, and others through EPA Regional Offices). Approximately 65,000 monitoring stations were batch georeferenced to the National Hydrography Dataset (NHD). The methods and results of the georeferencing are discussed in this paper. Linking nutrient data to the NHD allows geospatial interoperability with national environmental databases and facilitates query and display of water-related nutrient data and analyses, such as a station's upstream/downstream relation to discharge points. The ArcIMS application uses multiple layers, including the USGS National Elevation Data (NED), National Land Cover Data (NLCD), and orthophotography. The paper discusses how the current prototype utilizes comprehensive nutrient data and map layers for the Maryland, Virginia, and the District of Columbia pilot study area.

Introduction

Nitrogen and phosphorus have consistently ranked among the top three causes of use impairment in U.S. waters for more than a decade. Excess nutrients lead to significant water quality problems, including harmful algal blooms, hypoxia, and declines in wildlife and wildlife habitat, and have also been linked to increases in human pathogens. In recognition of this issue, the U.S. Environmental Protection Agency (USEPA), in cooperation with many other agencies and stakeholders, has developed a strategy to reduce impairment from nutrients (USEPA a,b, 2000). This strategy includes guidance for the development of nutrient criteria for both causal (i.e., phosphorus and nitrogen) and response (algal biomass/chlorophyll *a*, and measures of water clarity such as Secchi depth) variables. To support this effort, EPA has developed the National Nutrient Database (NUTDB),[i] a nutrient monitoring database that stores and analyzes nutrient water quality data and serves as an information resource for states, tribes, and others in establishing scientifically defensible numeric nutrient criteria. The database contains ambient data from EPA's Legacy STOrage and RETrieval (STORET) data system, the U.S. Geological Survey's (USGS's) National Stream Quality Accounting Network (NASQAN) data and National Water Quality Assessment (NAWQA) data, and other relevant sources such as universities and states/tribes. The ultimate aim of the database is to derive ecoregionally representative, as well as waterbody-specific numeric nutrient criteria that can be applied to all waterbodies of the United States. The NUTDB is a Web-enabled Oracle® database that includes locational (latitude and longitude coordinates of monitoring stations) as well as station information and analytically measured values for the nutrient parameters of interest. This database is linked to EPA's web-enabled EnviroMapper application. Georeferencing the monitoring locations in the database is a necessary prerequisite to harnessing the information contained in the database and supporting the programmatic and environmental goals to reduce the impairment of waterbodies by nutrients.

Georeferencing NUTDB Stations to the National Hydrography Dataset (NHD)

Overview of the NHD Georeferencing Process

Georeferencing is the process of locating an entity in "real world" coordinates. For example, you would georeference your house by determining its latitude and longitude coordinates or by reference to an existing geographic information system (GIS) data layer showing roads and street addresses. For information located on surface water features, the National Hydrography Dataset (NHD) provides the equivalent of a system of street addresses, and is the national standard for surface water data. The NHD is a digital network that represents surface water features, such as rivers and lakes. EPA maintains a surface water data system (Watershed Assessment, Tracking, and Environmental Results System [WATERS]) that includes data on impaired and assessed waters, designated uses, monitoring stations, and wastewater outfalls and no discharge zones, beach closures, fish consumption advisories, and nonpoint source projects. For this system, the geographic representation of water features in the NHD serves as a framework for organizing and integrating program attribute information. When represented in this manner, the program features can then be analyzed and displayed by computer-based or Web-based tools. By connecting program data to the NHD, water resources information can be managed and applied based on its geographic location. Because the NHD provides a unique and standard identifier for each segment of water across the country, it can be thought of as the reference framework for surface water analysis and display between EPA Office of Water programs, states, and other users

The USEPA and the USGS have developed standard methods to define features called "drains," which provide the foundation for the NHD hydrological network. For entities such as stream segments, linear events combining portions of one or more drains are used in the NHD georeferencing process. For entities such as point source discharges or monitoring sites, a point event within a single NHD drain is the preferred way to set up the georeferencing instructions (Figure 1).

Figure 1 illustrates how linear events, for information such as Section 303(d) impaired waters, and point events, for the locations of Permit Compliance System (PCS) discharges, can be applied using desktop or Web-enabled tools to create GIS mapping displays.[ii]

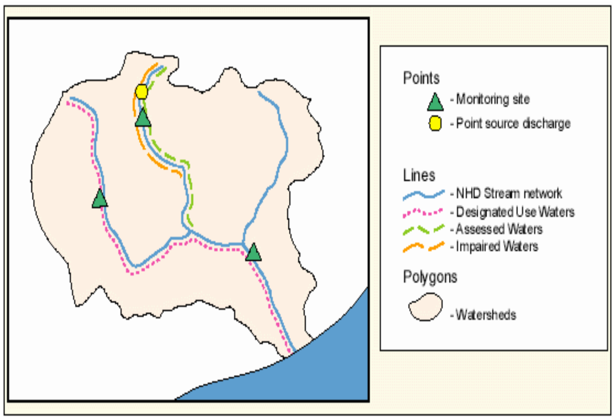

Figure 1. Point and line data georeferenced to the NHD.

The Oracle Spatial Batch Georeferencing Tool for Point Latitude/Longitude Coordinates

EPA has supported development of a Web-enabled locational improvement tool called the Web-based Reach Indexing Tool (WebRIT)[iii] (Rineer et al., 2003). The WebRIT can be used to take existing sets of latitude/longitude points, manually select positions on NHD drains, and then initiate a set of software procedures that convert this information into GIS point events georeferenced to the NHD. (EPA programs such as the Clean Water Act Section 319(h) Grant Tracking System [GRTS] and the Office of Science and Technology's [OST's] Beaches Environmental Assessment, Closure, and Health (BEACHES) program have encouraged use of the WebRIT in cases where a state agency is responsible for certifying the locational accuracy of a limited number of entities.) However, for georeferencing large data sets such as the NUTDB, an initial automated batch process followed by manual correction is far more efficient than a strictly manual approach. The software system that performs the NHD georeferencing in the WebRIT is based on an Oracle Spatial tool. This tool can be applied to run in an automated batch mode that takes an input data set of "off-NHD" latitude/longitude coordinates (as in the NUTDB) and produces output GIS materials that provide "on-NHD" point events georeferenced to NHD drains.

Because there is no human operator to select a preferred NHD drain to use in the automated on-NHD georeferencing process, the Oracle Spatial batch tool as currently implemented searches for up to five NHD drains in the vicinity of a latitude/longitude point and then performs the NHD georeferencing for the drain that has the shortest "snap distance" to the original off-NHD point. The major steps in this process are as follows:

1. Find the five closest drains using unprojected station and NHD locations (i.e., geographic data).
2. Project the five closest drains and the off-NHD station latitude/longitude to an Albers spheroid system to get more accurate distances.
3. Find the closest point to the projected station on the closest drain of the five candidate projected drains.
4. Determine the measure associated with the closest point on the closest drain (to create the needed GIS point event instructions).
5. Create a new point unprojected station location based on this and store it in the GIS output files.

These steps are then repeated for all of the latitude/longitude coordinates in the input non-NHD data set.[iv] Preliminary checks to identify off-NHD latitude/longitude points that fall in foreign countries (e.g., Canada) or more than 3 miles beyond the NHD coastal shorelines are conducted. These kinds of off-NHD points cannot be precisely placed on NHD drains, and segregating these coordinates is part of the quality assurance (QA) process.

Locational Precision and the Oracle Spatial Batch Georeferencing Process

The Oracle Spatial batch georeferencing approach has been applied to several large point data sets, including PCS pipe discharge points, surface intake points from the federal version of the Safe Drinking Water Information System (SDWIS/FED), USGS gaging station sites, and now the EPA's Office of Science and Technology, Health and Ecological Criteria Division's (OST/HECD's) NUTDB stations. Batch georeferencing such large data sets to the 1:100K NHD results in varying degrees of locational precision. For example, USGS took the batch georeferencing results for the gaging stations and performed a comprehensive manual evaluation of all the stations. USGS found a need for locational refinements for about 10 percent of these automatically georeferenced stations. For most of the stations for which corrections were deemed appropriate, USGS discovered problems with the original latitude/longitude coordinates. In general, therefore, if the original input latitude/longitude values are accurate, the Oracle Spatial batch georeferencing results will generally have acceptable levels of precision as a percentage of the total georeferenced data set. Batch georeferencing, then, can produce certain results where locational precision may still be an issue:

Adjacent Branches: Figure 2 shows two cases (A and B) where an off-NHD point has been snapped to an on-NHD location. For the points labeled "A," the original latitude/longitude point and the on-NHD point are in virtually the same place, so the georeferencing results seem quite acceptable. For the points labeled “B,” however, deciding whether the georeferencing is “correct” can require the actual data owner’s knowledge of the station locations. The upper end of the hypothetical stream network shown in Figure 2 divides into several branches. Unless the latitude/longitude values are extremely precise, it is possible that the shortest snap distance could move the on-NHD point onto the wrong branch.

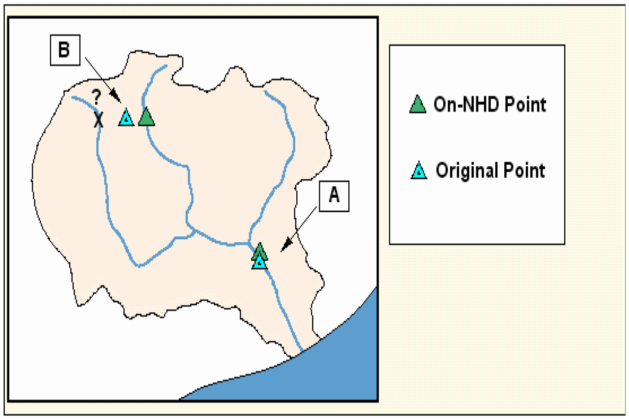

Figure 2. Some locational precision issues for stream networks.

Map resolution: There can also be problems with the automated georeferencing results when the original latitude/longitude points were developed for water features that appear only on higher resolution (1:24K) maps. Once again, detailed knowledge of the original station data will usually be required to be fully aware of such issues. Figure 3 shows NUTDB stations that actually lie on small water features to the south of Watts Branch, a stream that empties into the lower tidal Anacostia River in the District of Columbia. Currently, georeferencing tools are available only for the 1:100K resolution version of the NHD. Where these sorts of resolution issues can be documented, the preferred alternative would be to try to move the point to a confluence point of a stream drain that appears in the 1:100K NHD. If this is not an acceptable option, then the original off-NHD location could be used. However, such off-NHD points will lack upstream/downstream navigation capabilities and the ability to interact with other NHD-georeferenced/indexed databases. This type of situation is commonly called the “missing stream” (or “missing lake”) problem.

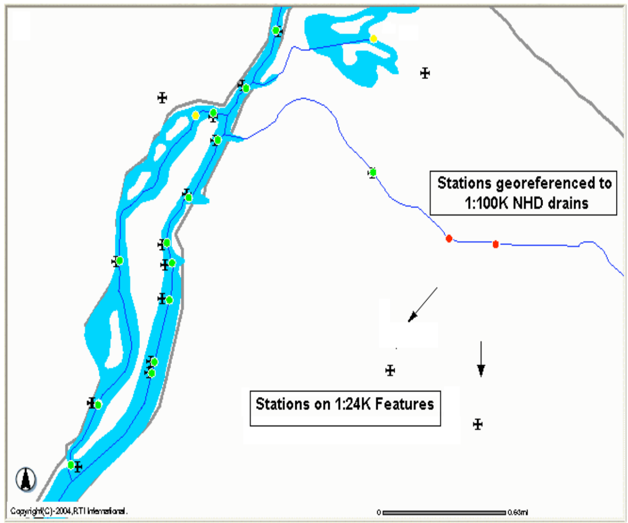

Figure 3. Precision issues for 1:24K mapping materials.

NHD Production Consistency: Another situation that can lead to locational precision issues is when portions of the NHD have production problems, notably line density variations. It is fairly common in all parts of the United States for adjacent "tiles" of parent digital line graph traces (used to generate the NHD) to exhibit major differences in trace density.

Figure 4 shows an example from New Hampshire, where the rectilinear tile edges related to these trace density shifts are readily apparent. Where the drain densities in a certain area are very low or transitional, batch georeferencing results can present locational precision problems. Other production issues exist for a small percentage of the drains in NHD that make it impossible to set up event tables that include these drains. There is currently no way to implement automated batch on-NHD georeferencing where these holes in the NHD dynamic segmentation system occur.

NHD Spatial Displacement: A final issue worth noting is that for certain situations, on-NHD georeferencing will entail some degree of displacement from the original off-NHD latitude/longitude locations. For large rivers (wide rivers), large natural lakes and reservoirs, and other open water areas such as estuaries, georeferencing onto NHD drains places the on-NHD points along centerlines that may not match the original latitude/longitude placements. The original points may have aimed to indicate precisely whether a station was on the left or right bank of a wide river or in the literal zone of a large lake (or reservoir).

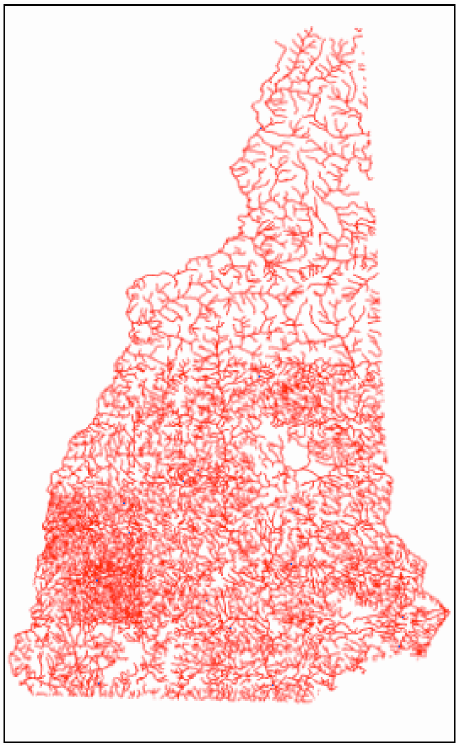

Figure 4. NHD drain density changes in New Hampshire.

Figure 5 illustrates the results of on-NHD georeferencing for a hypothetical open water feature. Although the NHD contains two-dimensional polygon features, dynamic segmentation cannot be implemented using polygons. The preferred approach is to snap a point onto a drain related to the transport reaches within the waterbody polygons. In such situations, the snap distances can be much larger than for simple single-lined streams.

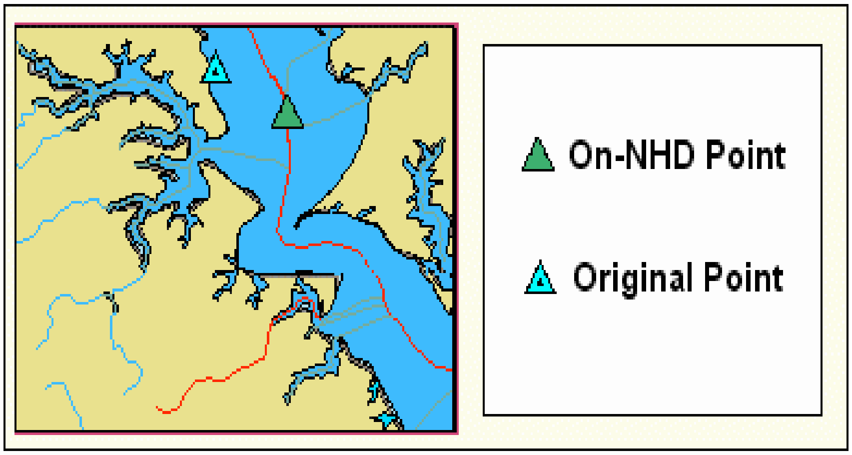

Figure 5. Illustration of the results of on-NHD georeferencing for wide rivers and other large open water features.

For automated batch georeferencing then, inherent limitations of the underlying NHD reference map and certain snapping process results can potentially require significant further investigation to identify, evaluate, and possibly correct locational accuracy.

Applying the Oracle Spatial Batch Georeferencing Tool for NUTDB Stations

USEPA's NUTDB database was accessed in July 2004 to obtain copies of tables with station location information for all 64,381 stations that had water quality parameter information. Original latitude/longitude coordinates were available for 63,315 stations, and the Oracle Spatial batch georeferencing process was applied to 63,250 of those stations, excluding those in areas where the NHD is not available or where the station coordinates would place them outside U.S. or coastal boundaries.

Original latitude/longitude coordinate values were not available for 1,066 stations with monitoring information. Many (70%) of these stations were located in Maine and involved small lakes or headwater streams. For these stations, RTI applied a set of tools that matched NUTDB information on the name of the waterbody and other relevant locational information (e.g., state and county in which the monitoring station was located) against data from the USGS Geographical Names Information System (GNIS). These latitude/longitude coordinates were then processed using the Oracle Spatial georeferencing tool.

Results of the Oracle Spatial Georeferencing Process and Initial QA Approaches

Of the stations where latitude/longitude coordinates were already provided in the NUTDB, georeferencing could be performed for 63,250 stations. Automated georeferencing was also performed on an additional 713 stations (of the 1,066) that lacked latitude/longitude coordinates through the automated GNIS-based matching routine that provided a set of suitable coordinates. Consequently, 63,963 stations have georeferencing results, which represents 99 percent of the 64,381 NUTDB stations with monitoring information.

In assessing the accuracy of georeferencing results from the Oracle Spatial procedure, a general rule of thumb is that the results are more likely to be accurate when the snap distance is short. From past experience in applying the batch georeferencing tool for other EPA or USGS data sets, we have found that snap distances greater than 1 kilometer (1,000 m) are likely to have locational precision issues warranting follow-up attention. As an initial QA evaluation, we manually inspected 10 of the 50 stations with the largest snap distances, selecting stations from different states. For 7 of these 10 points, the large snap distance was a function of the sparseness of NHD hydrography data. The other 3 points were associated with large open water bodies with no nearby NHD feature.

To facilitate the review of the georeferencing results, a simple Web-based ArcIMS® map viewer was deployed. An example of the content available in this simple review and evaluation tool for the District of Columbia is provided in Figure 6. The review site provides a way to zoom in to an area the size of a typical county, where a series of useful map features become visible. These map features include the following:

- On-NHD georeferenced points (and labels with the station IDs)
- Original off-NHD points based on latitude/longitude coordinates
- An image layer showing the linear and waterbody features in NHD
- Layers with state and county boundaries and major roads
- A "wallpaper" image layers toggle that displays either USGS topographical maps (7.5 minute Digital Raster Graphics [DRG]) or USGS Digital Ortho-Quad (DOQ) aerial photographs.

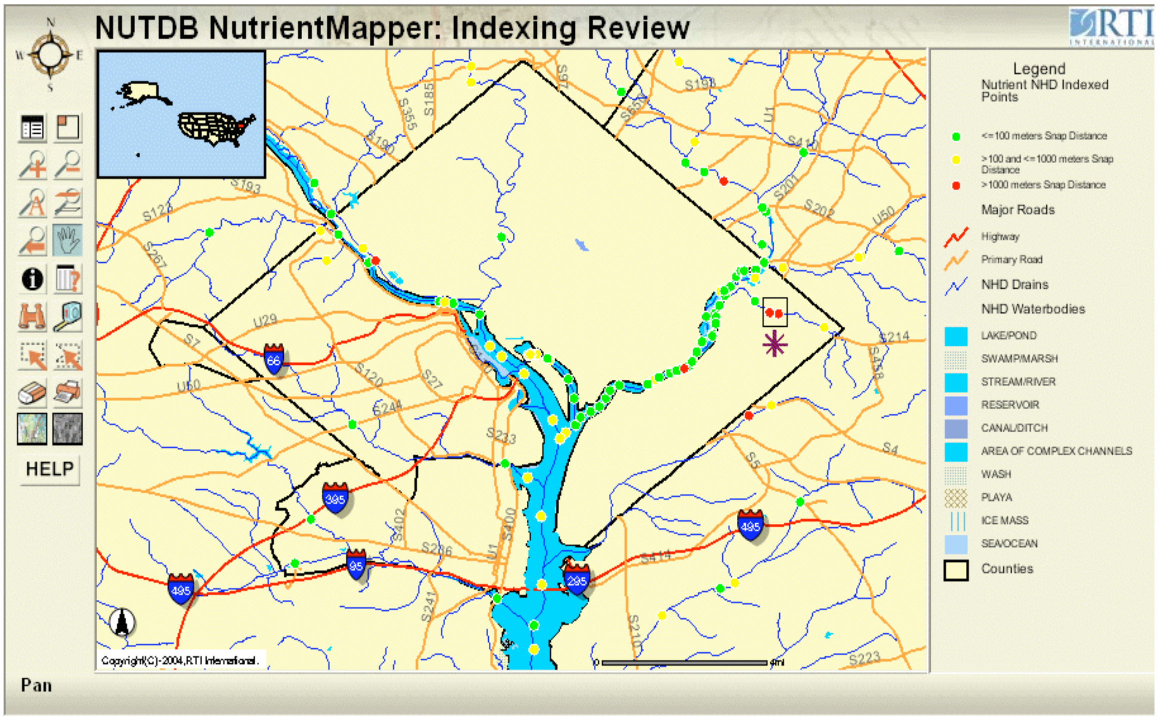

Figure 6. Georeferencing review tool showing the Washington, DC, area

The on-NHD georeferenced points are classified into three categories: points in green have georeferencing snap distances of less than 100 meters; yellow, 100 to 1,000 meters; and red, greater than 1,000 meters.

For the 63,963 georeferenced stations, the distribution according to snap distances is

Category	Distance	Count and % of Total
Small	100 m	39,049 (61%)
Medium	100 to 1,000 m	19,275 (30%)
Large	> 1,000 m	5,639 (9%)

For the stations with the largest snap distances (greater than 1,000 m), the predominant waterbody types were lakes or ponds (4,078) and reservoirs (42), accounting for about 73 percent of the stations in this category. There are also 788 river or stream stations (14%) and 735 estuary or coastal marine stations (13%) with large snap distances. As noted previously, georeferencing lakes or estuary waterbodies to NHD drains can sometimes result in large snap distances if the original station locations are in littoral or shoreline areas.

The number of stations with medium and large snap distances also varies by state. This distribution of snap variance by states provides an important first step in the aforementioned process of identification, evaluation, and correction of automated georeferencing results. Appendix A shows station counts organized by EPA Region and state for stations with snap distances in the 100 to 1,000 meter range and greater than 1,000 meters.

3. Development of a Prototype Web-Based NUTDB Display Tool

Main Objectives for the Prototype Display Tool

To assess the georeferencing work, a prototype display tool was created. The prototype display tool shows all of the on-NHD georeferenced NUTDB stations. Additional layers such as nutrient data and dams are included for Maryland, Virginia, and the District of Columbia. These data cover nutrient stations in several ecoregions that include all major waterbody types: rivers, lakes, and estuaries. The display tool adapts proven features of established Office of Water tools, such as the WebRIT, to provide a wide range of GIS and database query functionality using readily available HTML browsers, to provide the interface controls for end-user access.

The prototype display tool currently includes a query function that produces useful summary statistics for monitoring station data. The availability of data in the three states for total phosphorus (TP) and total nitrogen (TN) is illustrated in Figure 7.

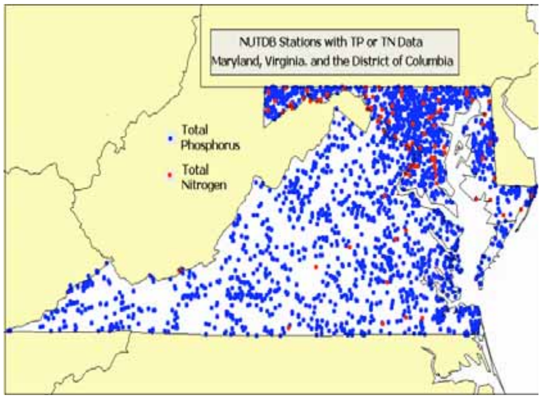

Figure 7. Availability of TP and TN data for Maryland, Virginia, and DC.

Figure 8 shows the layout for the query form window.

This portion of the prototype system is designed to use features in Oracle Application Server.[v] The outputs from a query can be used to generate tabular displays of water quality statistics for single stations or for groups of stations for a specified waterbody type and a specified geographical extent (e.g., for a specific aggregated EPA nutrient ecoregion within a state). The mapping capabilities in the prototype display tools are based on ESRI's ArcIMS software technology. ArcIMS provides a better platform for implementing Enviromapper-style GIS viewers in Web-based applications than the current MapObjects-based mapper system The prototype display tool makes use of ArcGIS-based techniques also used by Office of Water's WebRIT, the Office of Ground Water and Drinking Water's (OGWDW's) Drinking Water Mapping Application (DWMA), the American Indian Environmental Office's Federal Integrated Tribal Information Management System (TIMS), and the Office of Environmental Information's (OEI's) Window to My Environment (WME) system.

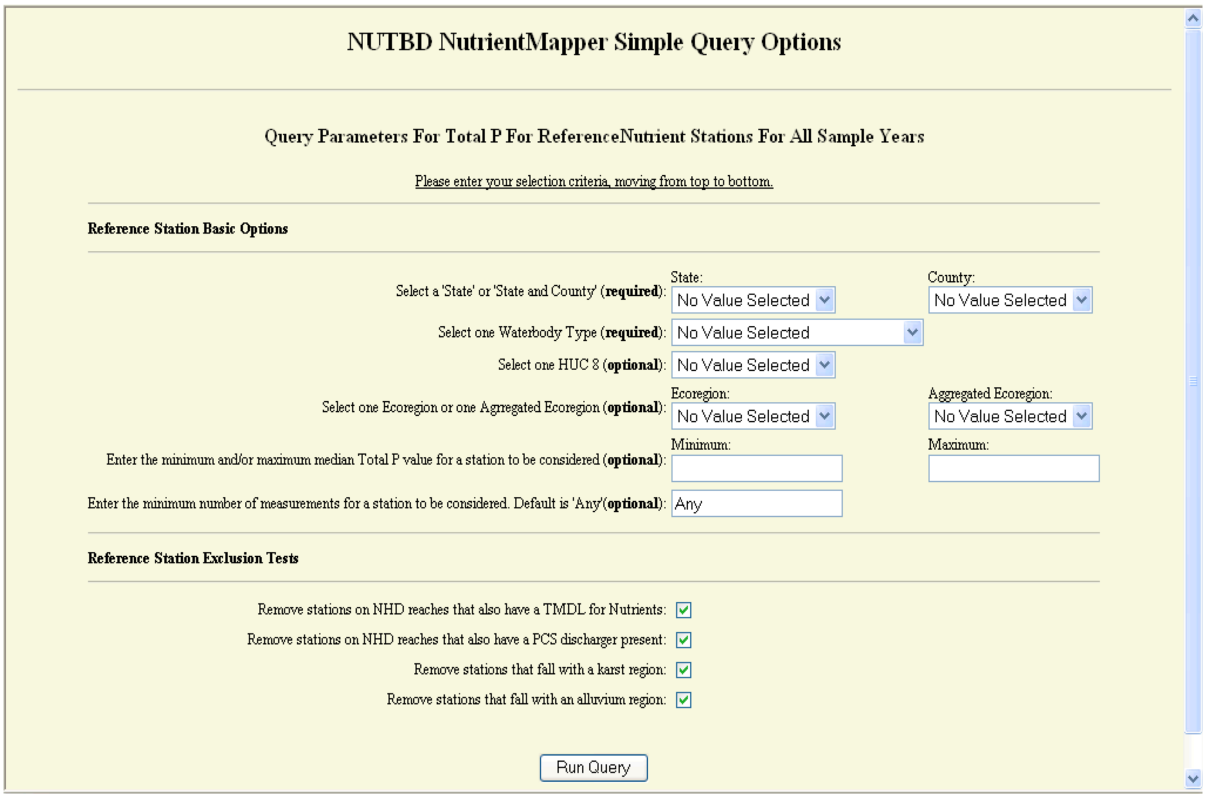

Figure 8. Layout for a prototype display tool query window.

As with any GIS application, the functionality of the prototype display tool, both in terms of the ability to generate reports or statistical summaries from database queries and the utility of the map viewer displays, is dependent on the underlying data content and map layers. In the following discussion, screen shots from the map viewer itself are used to illustrate how the tool can help study data content. Figure 9 shows map views at a regional (low resolution) small spatial scale, where only a few mapping features can be displayed. Even at this low resolution, several image layers from a USGS Web service can be applied to a map view. In this case, a National Elevation Data (NED) terrain relief layer is activated. These USGS image layers are provided through the National Map Seamless Data Distribution System (SDDS).

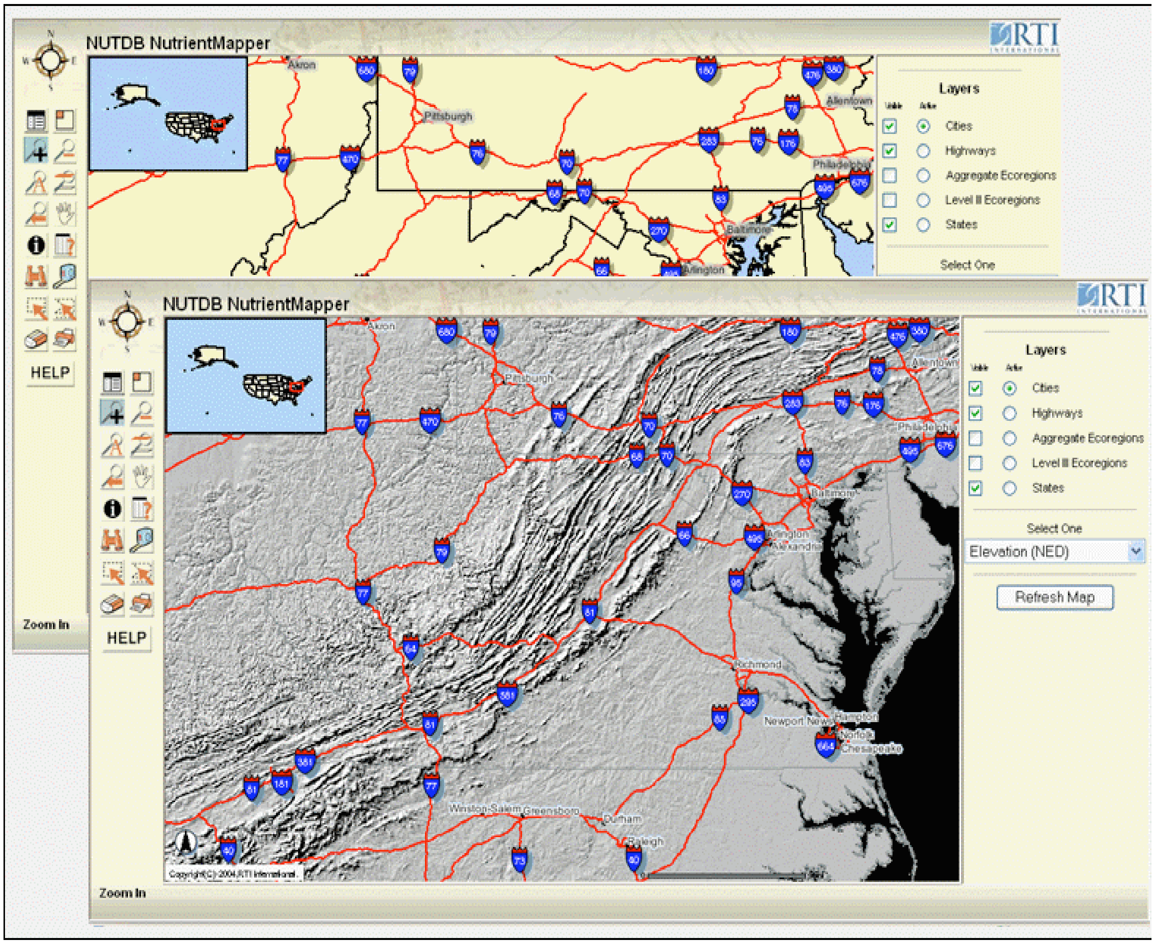

Figure 9. Map view at a regional scale showing USGS Web-service NED image layer.

Figure 10 shows a larger scale (higher resolution) view of the District of Columbia area. For areas in the size range of the District and typical county administrative units, the complete range of display tool GIS layers and image layers become available. The display in Figure 10 shows the National Land Cover Data (NLCD) image layer activated. All of the data sets, GIS layers, and Web-service image layers in the prototype tool are derived from available resources through USEPA systems, USGS Web services, or private Web services such as Microsoft TerraServer.[vi]

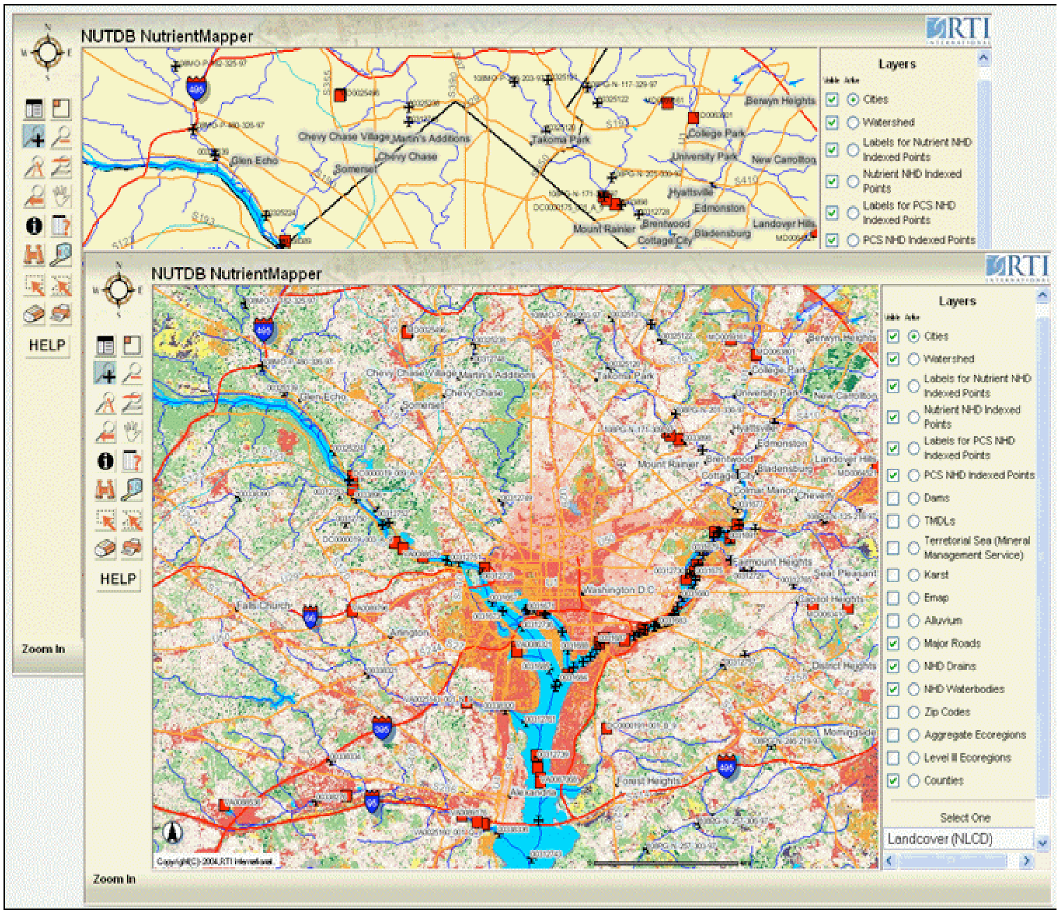

Figure 10. High-resolution map view of the DC area showing the full set of mapping features and the NLCD image layer selected as a background layer.

The map-viewing features and image layers for the prototype display tool are listed in Appendix B, along with a short summary description. This metadata information is included as part of an online information/help facility within the display tool.

Overview of Opportunities for Enhancements to the NUTDB Web-Based Tools

Several factors are involved in making recommendations for enhancements to the currently deployed NUTDB tools. Changes in the current tools will be required over the next year or so simply to accommodate upgrades to the current versions of USEPA enterprise network software services such as upgrades to Oracle and SAS versions. As with many of the data systems that share in the current WATERS system design, there will be a shift from map viewers based on the ESRI MapObjects image server to ESRI's ArcIMS. These pending changes in the entire collection of database and GIS tools that support the NUTDB provide opportunities to consider ways to factor in a number of other system enhancements.

Progress in georeferencing the NUTDB stations to the NHD is another major factor encouraging system enhancements. Modifications to the current NUTDB query and mapping procedures are needed to take advantage of the new georeferencing information. A whole range of new geospatial analysis approaches (e.g., easily determining if there are any Section 303(d) listed waters in upstream or downstream proximity to a nutrient station) are now possible. The content and features are now available to add a range of new functionality to the NUTDB tools, but upgrades in these Web-based applications will be needed to add these new query and mapping capabilities.

In the process of improving the NUTDB application and tools, there are also opportunities to reorganize the NUTDB tools to better serve the needs of specific stakeholders and clients/users. For example, the underlying database and GIS layer content used in the NUTDB system will likely become more widely available. A useful design direction would be to set up different tiers of user access to satisfy the needs of the general public on one level and internal agency/governmental data owners on another level.

Summary, Conclusions, and Recommendations

Under this project, USEPA's OST/HECD has completed the initial georeferencing of 99 percent of the 64,381 stations with water quality data in the NUTDB to the USGS 1:100K NHD. This georeferencing process has assigned "on-NHD" latitude and longitude coordinates to the stations over the NUTDB stations original latitude/longitude coordinates. The resulting map layer can then be used as a GIS map layer in concert with other map- and NHD-based data for display, analysis, and map outputs. The georeferencing was based on the version of the USGS (NHD) used in the USEPA Office of Water's Reach Address Database (RAD Version 2). Georeferencing point locations to the drain identifiers contained in the NHD satisfies USEPA's business rules for documenting how locational information is expressed for use in geospatial analyses. This paper summarizes the steps taken to complete this initial NHD georeferencing process and discusses the GIS data layers and Web-based viewing tools for review and ongoing management of the NUTDB database.

NHD georeferencing of the NUTDB opens a number of new dimensions in how Web-based query and map analyses can be performed. The information in the NUTDB can be very effectively integrated with a growing number of other data systems using NHD georeferencing. Under this project, a special Web-based georeferencing review tool was deployed. This simple ArcIMS-based viewer tool was designed to facilitate evaluation of the initial georeferencing results and help formulate strategies to improve georeferencing results over time. In addition to the review tool, a prototype display tool with more features was created and deployed to provide working examples of useful enhancements and upgrades to the NUTDB mapping analysis tools currently deployed on USEPA's computer system.

Recommendations for Ongoing Improvements in Spatial Accuracy and Precision of Georeferencing

As steps to continue to improve on locational data in the NUTDB, we recommend several general principles:

1. Continuing to frame/define reliability of batch results through statistical analyses and testing
2. Developing manual georeferencing routines suitable for NUTDB on NHD GIS data
3. Accessing data owners and knowledge at the regional, state, and local level
4. Utilizing the georeferencing review tool and the prototype display tool to support improved georeferencing.

Over the short term, we recommend that two mapping tools be maintained. The basic georeferencing review tool should be maintained online to display the original NUTDB station locations along with the new NHD-georeferenced stations. The prototype display tool, on the other hand, should screen data for certain quality parameters before stations could be input into the system. A major expectation would be that the snap distances should not exceed appropriate thresholds unless additional evaluation reviews determined that stations with larger snap distances should be included in the display tool. For rivers, an initial recommendation is that stations be included in the display tool only if the snap distances are less than or equal to 100 meters. Lake and estuary original latitude/longitude placements are often intended to capture sites on littoral or shoreline areas rather than NHD centerline drains; therefore, for nonriver waterbody types, the snap distance expectations might be relaxed to around 500 meters.

The basic georeferencing review tool should be maintained with only those stations where snap distances are over acceptable thresholds or where station georeferencing results have not been validated as acceptable by qualified experts. As documented in Appendix A, in most states there are a fairly small number of stations where the snap distances suggest the need for adjustments in the initial georeferencing results. Regional/state/local experts could provide a review for USEPA Headquarters with suggestions for corrected latitude/longitude coordinate values. After such updates, stations with validated georeferencing results could be deployed into the production display tool.

Acknowledgments

This project was fully funded by the U.S. Environmental Protection Agency through the Office of Science and Technology, Health and Ecological Criteria Division contract TECHNICAL AND REGULATORY SUPPORT FOR THE DEVELOPMENT OF CRITERIA FOR WATER MEDIA (ECOLOGICAL EMPHASIS) – 68-C-04-006 with Great Lakes Environmental Center (GLEC) Inc. RTI International Inc. is a sub-contractor to GLEC. We would like to thank Steve Beaulieu at RTI

International for his contractual oversight and Anne Marie Miller also at RTI International for her help with the ArcIMS viewers used in this work.

Appendix A

Table A-1. Summary of Stations with Large Snap Distances by USEPA Region and State

USEPA Region	State	Number of Stations With Snap Distance	
		100–1,000 m	> 1,000 m
01	Connecticut	19	2
01	Massachusetts	186	77
01	Maine	236	32
01	New Hampshire	316	28
01	Rhode Island	41	15
01	Vermont	277	114
02	New Jersey	175	87
02	New York	250	46
02	Puerto Rico	99	0
03	District of Columbia	36	8
03	Delaware	166	19
03	Maryland	298	20
03	Pennsylvania	258	3
03	Virginia	169	10
03	West Virginia	839	29
04	Alabama	211	16
04	Florida	6,115	3,286
04	Georgia	151	5
04	Kentucky	99	14
04	Mississippi	57	4
04	North Carolina	390	31
04	South Carolina	430	53
04	Tennessee	209	8
05	Illinois	750	128
05	Indiana	44	11
05	Michigan	247	114

(continued)

Table A-1. (continued)

USEPA Region	State	Number of Stations With Snap Distance	
		100–1,000 m	> 1,000 m
05	Minnesota	739	514
05	Ohio	261	10
05	Wisconsin	775	276
06	Arkansas	249	9
06	Louisiana	125	52
06	New Mexico	137	41
06	Oklahoma	166	11
06	Texas	460	43
07	Iowa	115	1
07	Kansas	243	7
07	Missouri	118	0
07	Nebraska	177	21
08	Colorado	217	22
08	Montana	746	37
08	North Dakota	118	60
08	South Dakota	251	103
08	Utah	708	131
08	Wyoming	132	30
09	Arizona	80	5
09	California	195	21
09	Hawaii	33	0
09	Nevada	71	6
10	Idaho	287	17
10	Oregon	567	29
10	Washington	236	30

Appendix B: NUTDB Nutrient Mapper Feature Layers

Table B-1. Display Tool Map Feature and Image Layers (The content items with special significance in providing an infrastructure for analyses involving nutrient monitoring station locational or water quality information are flagged with an asterisk.)

Item	Feature Name	Item	Feature Name
1	Cities	14	NHD Drains
2	States	15	NHD Waterbodies
3	Watersheds (HUC8 Sub-basins)	16	Zip Codes
4	* Nutrient NHD-Indexed Points	17	* Aggregated Ecoregions
5	* PCS NHD-Indexed Points	18	* Level III Ecoregions
6	* Dams	19	Counties
7	* TMDLs/303(d)		
8	* Territorial Sea		**Image Layer Name**
9	* Karst	20	USGS DRG
10	* EMAP Coastal Provinces	21	USGS DOQ
11	* Alluvium	22	Elevation (NED)
12	Highways	23	* Landcover (NLCD)
13	Major Roads	24	Landcover and Elevation

A label layer is included.
Abbreviations:
HUC8 = 8-digit cataloging unit (USGS)
TMDL = Total maximum daily load
EMAP = Environmental Monitoring and Assessment Program
NED = National Elevation Data
NHD = National Hydrography Dataset
NLCD = National Land Cover Data
PCS = Permit Compliance System
USGS DRG = U.S. Geological Survey Digital Raster Graphics
USGS DOQ = U.S. Geological Survey Digital Orthophoto Quadrangle

NUTDB Nutrient Mapper Data Layers

1) **Cities:** The Cities layer is from the U.S. Geological Survey (USGS) Map Layers Warehouse. This map layer includes the locations of more than 28,000 towns and cities in the United States. For more information, see http://www.nationalatlas.gov/citiesm.html

2) **States:** The States layer is from the U.S. Census Bureau's Census 2000 TIGER/Line. For more information, see http://www.census.gov/geo/www/tiger/index.html

3) **Watersheds:** This feature is a set of the USGS 8-digit Cataloging Unit (HUC8) sub-basins used in USEPA's Watershed Assessment, Tracking, and Environmental Results System (WATERS). The United States is divided and subdivided into successively smaller hydrologic units, which are classified into four levels: regions, subregions, accounting units, and cataloging units (or sub-basins). The hydrologic units are arranged within each other, from the smallest (cataloging units) to the largest (regions). Each hydrologic unit is identified by a unique hydrologic unit code (HUC) consisting of two to eight digits based on the four levels of classification in the hydrologic unit system. For more information, see http://water.usgs.gov/GIS/huc.html

4) **Nutrient NHD-Indexed Points:** NUTDB georeferenced stations that have been mapped to NHD.

5) **Permit Compliance System (PCS) NHD-Indexed Points:** PCS discharge points that have been mapped to NHD. Also called Pipe Report Outfalls, these points were developed for the Office of Wastewater Management.

6) **Dams:** The National Inventory of Dams (NID) database, assembled by the U.S. Army Corps of Engineers for the late 1990s, contains information, including latitude and longitude, for more than 76,000 dams throughout the United States and its territories. The MDVADC_Dams layer is a subset of the national database, with data for the Virginia, Maryland, and District of Columbia areas only. Significant changes have been made to the current version of the NHD data, including the addition of new dam records and the removal of breached dams and duplicate dam records. Several new fields have been added to the NHD, including dam former name, dam designer, core, foundation, and state-regulated designation. This new information is deemed necessary to assess dam characteristics and to more effectively and appropriately allocate federal resources for dam safety programs. Information on the NHD, as well as facilities to download NHD data tables, is available at http://crunch.tec.army.mil/nid/webpages/nid.cfm

7) **Total Daily Maximum Loads (TMDLs)/§303(d) Listings:** USEPA's Office of Water has created events for the Section 303(d) impaired waters. The section 303(d) (impaired waters) list is a prioritized list of waters that do not meet water quality standards. Each state is required to establish TMDLs for these waters. Impaired water events include river segments, lakes, and estuaries designated under Section 303(d) of the Clean Water Act. This layer is a subset of the pre-2002 303(d) impaired waters list containing impairment codes dealing with nutrients and eutrophication. These materials reflect archived GIS materials and data tables from the Office of Water's National TMDL Tracking System (NTTS) database. As of November 2004, USEPA was still validating the GIS locational information for the most current set (2002) of 303(d) listed waters. (In the display tool, NHD-georeferenced segments are displayed only for reaches with impairment codes that have an obvious relationship with nutrient pollutants such as phosphorus or nitrogen or impairments related to problems stemming from excess nutrients such as algal blooms or elevated chlorophyll-a levels. The selection of these particular codes is provisional and could be further refined.)

8) **Territorial Sea:** Coverage from Department of the Interior's Mineral Management Service (MMS) shows the approximate extent of the territorial sea, which extends in most states from 5 nautical miles from the U.S. Coast out to the limits of the outer continental shelf. The coverage for the conterminous United States is part of the USEPA Office of Coastal Protection Division's Mercury in Marine Life (MML) database and geographic information system (GIS) mapping system.

9) **Karst:** The karst layer shows areas with predominant sub-surface karst lithology. The coverage is from USGS and is not considered to be of sufficient spatial resolution to be used for applications requiring a high degree of accuracy. USGS has provided this layer for inclusion in the USEPA RAD. USGS is in the process of developing a new karst coverage. Additional background information is available at http://water.usgs.gov/ogw/karst/kigconference/rco_proposedmap.htm

10) **EMAP Coastal Provinces:** The Environmental Monitoring and Assessment Program (EMAP) led an initiative to integrate previous EMAP coastal data collections. The result was the National Coastal Assessment (NCA). Because there is not a consistent definition of an estuary or other near-coastal waterbodies, the EMAP NCA layer can assist in relating monitoring sites to well-defined polygons in estuarine and near coastal settings. This coverage also provides assistance in determining the USEPA-defined coastal (or biogeographical) province in which a monitoring site in estuarine, marine, or tidal river locations would be located. An USEPA Office of Research and

Development Web-enabled mapping system that uses this coverage is available at http://www.epa.gov/emap/nca/html/data/index.html

11) **Alluvium:** This layer was derived from the USGS coverage of surficial geology. The "Alluvium" attribute was used to select the desired data from this layer to depict the alluvial beds of the main river basins in the United States. The complete surficial geology coverage is available as part of the RAD.

12) **Highways:** The Highways layer is extracted from the Department of Transportation's (DOT's) Office of Geographic Information Services Major Roads layer. More information is available at http://www.transtats.bts.gov/MappingCenter.asp

13) **Major Roads:** This layer is extracted from the Major Roads layer provided by the DOT Office of Geographic Information Services. More information is available at http://www.transtats.bts.gov/MappingCenter.asp

14) **NHD Drains:** The NHD is a geographic data set that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells. Within the NHD, these features are combined to form reaches, which provide the framework for linking water-related data to the NHD surface water drainage network. These linkages, which are essentially stream addresses based on unique reach codes, enable the analysis and display of these water-related data in upstream and downstream order, among other uses. The NHD is based on USGS Digital Line Graph (DLG) hydrography data integrated with reach-related information from the USEPA Reach File Version 3 (RF3) data set. Detailed metadata are available at http://134.67.99.56/NHDMapper/help/standards-data5.html

15) **NHD Waterbodies:** The NHD is a geographic data set that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells. Within the NHD, these features are combined to form reaches, which provide the framework for linking water-related data to the NHD surface water drainage network. These linkages, which are essentially stream addresses based on unique reach codes, enable the analysis and display of these water-related data in upstream and downstream order, among other uses. The NHD is based on USGS Digital Line Graph (DLG) hydrography data integrated with reach-related information from the USEPA Reach File Version 3 (RF3) data set. Detailed metadata are available at http://134.67.99.56/NHDMapper/help/ standards-data5.html

16) **Zip Codes:** The Zip code data are from Geographic Data Technology (GDT). Please refer to the GDT Web site at http://www.geographic.com/home/index.cfm for more information on the Zip code boundaries.

17) **USEPA Office of Water Aggregated Nutrient Ecoregions**: This set of 14 aggregated ecoregions (revised by Omernik in 1998) cover the conterminous U.S. and are aggregations of the Level III ecoregions. Information is available at: http://www.epa.gov/waterscience/standards/ecomap.html.

18) **Level III Ecoregions:** There are 84 sub-ecoregions in this layer. This GIS coverage was made available by the USEPA OST/HECD and is used to assign nutrient monitoring stations to their appropriate Level III ecoregion.

19) **Counties:** The Counties layer is from the USGS Map Layers Warehouse. This map layer portrays the county boundaries of the United States. For more information, see http://www.nationalatlas.gov/countiesm.html

NUTDB Nutrient Mapper Image Layers

1) **USGS DRG:** A digital raster graphic (DRG) is a scanned image of a USGS topographic map. The image is georeferenced to the surface of the earth and fit to the Universal Transverse Mercator projection. For more information on the development of this layer, see http://topomaps.usgs.gov/drg/. This image layer is based on USGS data and is accessed through the Microsoft TerraServer, available at http://terraserver.microsoft.com/

2) **USGS DOQ:** A digital orthophoto quadrangle (DOQ) is a computer-generated image of an aerial photograph. The DOQ can be used to measure distances accurately because it has been orthorectified, meaning it has been altered so that it has the geometric properties of a map. For more information on the development of this layer, see http://geography.wr.usgs.gov/doq/index.html This image layer is based on USGS data and is accessed through the Microsoft TerraServer, available at http://terraserver.microsoft.com/.

3) **Elevation (NED):** National Elevation Data (NED) is a seamless raster coverage of 1:24,000 resolution Digital Elevation Model (DEM) data for the entire United States. For more information on the development of this layer, see http://ned.usgs.gov/. For information on accessing these USGS products as Web-services as part of the National Map Seamless Data Distribution System (SDDS), see http://seamless.usgs.gov/

4) **Landcover (NLCD):** Derived from the early to mid-1990s Landsat Thematic Mapper satellite data, the National Land Cover Data (NLCD) is a 21-class land cover classification scheme applied consistently over the United States. The spatial resolution of the data is 30 meters and mapped in the Albers Conic Equal Area projection, NAD 83. For more information on the development of this USGS layer, see http://landcover.usgs.gov/natllandcover.asp For information on accessing these USGS products as Web services as part of the National Map SDDS, see http://seamless.usgs.gov/

5) **Landcover and Elevation:** This layer is a combination of the Landcover and Elevation layers. The Landcover data have been "draped" over the top of the elevation data to present the land cover in relation to the elevation. For information on accessing these USGS products as Web services as part of the National Map SDDS, see http://seamless.usgs.gov/

End Notes

[i] http://www.epa.gov/waterscience/criteria/nutrient/database/index.html
[ii] Additional information on the techniques involved in georeferencing entities to the NHD and the EPA RAD can be found at http://www.epa.gov/waters/about/geography .html.
[iii] http://www.epa.gov/waters/webrit/
[iv] At this writing, the Oracle Spatial procedure cannot be applied to states, commonwealths, and territories where the needed version of the 1:100K NHD is not yet deployed in EPA's RAD (examples are Alaska, the U.S. Virgin Islands and Guam).
[v] The prototype currently uses Oracle 9*i*. It is envisioned that the design will become compatible with the new Oracle Database 10g Grid Computing product line. Oracle 10g capabilities are a cornerstone of EPA Enterprise Architecture business rules, which in the future will establish secure access to all EPA systems and provide a new range of "portal" tools to take advantage of web-based application concepts making full use of modular development approaches and a growing range of Web services. In the prototype display tool, the expanded set of statistical functions available in Oracle 9i are used to generate the summary statistics. In the future, the statistical features in Oracle or in web-enabled SAS procedures (using the new SAS version 9) will be applied to provide tables of statistical results or special graphical outputs for the statistical information (e.g., Tukey box-and-whisker diagrams or graphs showing cumulative distribution functions).
[vi] More information on the TerraServer is available at http://terraserver.microsoft.com/

References

Rineer, J., M. Plastino, A. M. Miller, J. Sinnott. 2003. Developing and Integrating Web Services for Georeferencing to the NHD. 23rd Annual ESRI International User Conference, July 7–11, San Diego, CA. Available at http://gis.esri.com/library/userconf/proc03/index.html.

USEPAa, Office of Water, Office of Science and Technology. April 2000. Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs EPA-822-B00-001. Washington, DC 20460. Available at http://www.epa.gov/waterscience/criteria/nutrient/guidance/lakes/index.html.

USEPAb, Office of Water, Office of Science and Technology. July 2000. Nutrient Criteria Technical Guidance Manual: Rivers and Streams. EPA-822-B-00-002. Washington, DC 20460. Available at http://www.epa.gov/waterscience/criteria/nutrient/guidance/rivers/index.html.

Author Information

Ifeyinwa Davis
U.S. Environmental Protection Agency, National Nutrient Criteria Program, (MC 4304T)
Ariel Rios Building, 1200 Pennsylvania Avenue, NW, Washington DC 20460
Ph. (202) 566-1096
Fx. (202) 566-1140
davis.ifeyinwa@epa.gov

William Cooter, James Rineer, James Sinnott, Randy Dodd
RTI International
PO Box 12194, Research Triangle Park, NC 27709-2194
Ph. (919) 316-3728
Fx. (919) 541-7155
sid@rti.org

Bill Arnold
Great Lakes Environmental Center
739 Hastings, Traverse City, MI 49686
Ph: (231) 941-2230
Fx: (231) 941-2240
barnold@glec-tc.com