USDA Service Center Agencies
Geospatial Data Management Team
Data Management Plan For
National Hydrography Dataset (NHD)
High Resolution (24K)

Ken Becker
Ravi Kaushika
Randy Frosh

September 2007

I. Purpose and Scope (business case)

A. Purpose

The U.S. Geological Survey has developed a National Hydrography Dataset (NHD). The NHD is a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells. Within the NHD, surface water features are combined to form "reaches," which provide the framework for linking water-related data to the NHD surface water drainage network. These linkages enable the analysis and display of these water-related data in upstream and downstream order.

The NHD is based upon the content of USGS Digital Line Graph (DLG) hydrography data integrated with reach-related information from the EPA Reach File Version 3 (RF3). The NHD supersedes DLG and RF3 by incorporating them, not by replacing them. Users of DLG or RF3 will find the National Hydrography Dataset both familiar and greatly expanded and refined.

The NHD is a national framework for assigning reach addresses to water-related entities, such as industrial discharges, drinking water supplies, fish habitat areas, wild and scenic rivers. Reach addresses establish the locations of these entities relative to one another within the NHD surface water drainage network, much like addresses on streets. Once linked to the NHD by their reach addresses, the upstream/downstream relationships of these water-related entities--and any associated information about them--can be analyzed using software tools ranging from spreadsheets to geographic information systems (GIS). GIS can also be used to combine NHD-based network analysis with other data layers, such as soils, land use and population, to help understand and display their respective effects upon one another. Furthermore, because the NHD provides a nationally consistent framework for addressing and analysis, water-related information linked to reach addresses by one organization (national, state, local) can be shared with other organizations and easily integrated into many different types of applications to the benefit of all.

While initially based on 1:100,000-scale data, the NHD is designed to incorporate and encourage the development of higher resolution data required by many users.
The specifications for the data are: Geographic coordinate system Horizontal datum of NAD83. The NHD is a living dataset that is updated bimonthly to incorporate the "best available" data. The NHD also incorporates the National Spatial Data Infrastructure framework criteria established by the Federal Geographic Data Committee.

The dataset is provided in essentially two resolutions at this time. The first is the “medium” resolution based on the 100K scale data; the second is the “high” resolution based on the 24K scale data. As higher resolution data becomes available from state and local governments and other sources, a third category will become more and more common – “local” resolution data.

The NHD home, http://nhd.usgs.gov/ provides a great deal of information about the NHD.

B. Scope

The scope is the United States, Hawaii and the Virgin Islands, and Puerto Rico. Not all areas are completed at the high resolution scale as of the date of this plan.

II. Acquisition

A. Data Source

1. Producer Information

a. Name

USGS

b. Location of Headquarters

U.S. Geological Survey
Bldg. 810 – Denver Federal Center
Box 25046
Denver, Co. 80225

c. Internet Address

http://nhd.usgs.gov/

2. Publisher Information

a. Name

USGS

b. Location of Headquarters

U.S. Geological Survey
Bldg. 810 – Denver Federal Center
Box 25046
Denver, Co. 80225

c. Internet Address

http://nhd.usgs.gov/

3. Acquisition Information

a. Delivery Media

DVD

b. Download URL

http://nhdgeo.usgs.gov/viewer.htm

c. Projected Data Availability Schedule

Currently available

B. Standards Information

1. Geospatial Data Standard

a. Standard Name and Steward Information

The National Hydrography Dataset (NHD) is built upon the most current applicable standards for geospatial data and metadata. In particular, NHD's foundation is an assembly of Digital Hydrography files from USGS topo maps and other sources that conform to the USGS's National Mapping Program standards.

b. Standard Version

None

c. Standard URL

http://rockyweb.cr.usgs.gov/nmpstds/nhdstds.html

2. Metadata Standard

a. Standard Name and Steward Information

Metadata are compliant with:
Federal Geographic Data Committee (FGDC)
Content Standard for Digital Geographic Metadata FGDC
STD-001-1998 Version 2 revised June 1998

NHD is documented using the Federal Geographic Data Committee (FGDC) content standard for geospatial metadata, and will be kept in conformance with a future ISO metadata standard. NED is a searchable collection through the National Spatial Data Infrastructure (NSDI) and its network of distributed clearinghouse nodes. NED will be offered using the Spatial Data Transfer Standard and in particular, the raster profile, an FGDC endorsed (FGDC-STD-002.5) profile of the ANSI and FGDC endorsed SDTS (ANSI NCITS 320:1998, FGDC-STD-002, FIPS 173-1). NED is provided in formats that are fully compliant with Executive Order 12906, OMB Circular A130, and OMB Circular A-119.

b. Description of Metadata Captured

The Federal Geographic Data Committee's content standard for digital geospatial metadata (http://www.fgdc.gov/metadata/geospatial-metadata-standards) is used to document NED data. Metadata about the individual source datasets used to assemble NED is presented in a spatially-referenced form. The polygonal footprint of each portion used from a source dataset is retained during NED assembly to provide the spatial context. The attributes of each source dataset, such as original resolution, production method, and date entered into NED, are linked to this polygonal footprint. All of these source polygons together form a seamless National coverage. Through this spatially-referenced metadata, the information is made available to the user regarding the source data used for any area of NED. For example, a NED user might use the spatially-referenced metadata to identify the portions of a study area that were produced by obsolete production methods.

For a template (spatial coverage information added when data is extracted) see the National Hydrography Dataset FGDC compliant metadata:
http://nhdgeo.usgs.gov/metadata/nhd_high.htm

c. Metadata Accuracy and Completeness Assessment

The metadata is very complete.

C. Acquired Data Structure

1. Geospatial Data Format

a. Format

Vector

b. Format Name

ESRI Arc SDE Export

c. Data Extent

The data extent is the coterminous United States, Hawaii the Virgin Islands, and Puerto Rico. High resolution (24K) data is not yet available for the entire area however.

d. Horizontal and Vertical Resolution

None stated but probably no more than one meter.

e. Absolute Horizontal and Vertical Accuracy

Statements of horizontal positional accuracy are based on accuracy statements made for U.S. Geological Survey topographic quadrangle maps. These maps were compiled to meet National Map Accuracy Standards.

f. Nominal Scale

1:24,000

g. Horizontal and Vertical Datum

The horizontal datum for all areas is NAD83. The vertical datum for all areas is NAVD29.

h. Projection

Geographic

i. Coordinate Units

Decimal Degrees.

j. Average Data Set Size

The full dataset, including lower 48, HI, AK, the Virgin Islands and Puerto Rico is over --- GB and getting bigger with each bimonthly update of new data.

k. Symbology

None

2. Attribute Data Format

a. Format Name

ArcSDE export files.

b. Database Size

N/A

3. Data Model

a. Geospatial Data Structure

Data is delivered in ArcSDE internal ESRI proprietary format.

b. Attribute Data Structure

ESRI ArcSDE

c. Database Table Definition

Varies by layer.
See http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/draft/dlg-f/nhd/NHDH0799.PDF

for definition of the fields in each layer.

d. Data Relationship Definition

Varies by layer.
See http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/draft/dlg-f/nhd/NHDH0799.PDF

for definition of the fields in each layer.

e. Data Dictionary

A data dictionary (NHDinGEO Feature Codes (FCodes) by layer) is located at:

http://nhd.usgs.gov/NHDinGEO_FCodes_by_layer.pdf

D. Policies

1. Restrictions

a. Use Constraints

None. Acknowledgment of the originating agencies would be appreciated in products derived from these data.

b. Access Constraints

None

c. Certification Issues

None

2. Maintenance

a. Temporal Information

The NHD is designed to accommodate both the higher resolution data that many users need, and the 1:100,000 scale data. The higher resolution data is incorporated into the NHD through the participation of users at the national, State, and local

levels. The common identifiers for the features are the basis for tracking and sharing deletions, additions, and modifications of features during maintenance. They are used to communicate and share corrections among organizations. The NHD will improve the integration of hydrographically related data in support of the varied applications of a growing national user community, and it will also enable shared maintenance and enhancement.

b. Average Update Cycle

The NHD is updated continually. A visual index is located at http://nhdgeo.usgs.gov/viewer.htm. Click on NHD Status in the right hand side of the display, then click the "High" Button. In the upper right, click "Legend".

E. Acquisition Cost

1. Cooperative Agreement

a. Description of Agreement

There is an agreement with the USGS. The NHD database will be sent to NCGC quarterly. There is no coup agreement.

b. Status of Agreement

A coup agreement is desired.

2. Cost to Acquire Data

None

As of the date of this document, the USGS is providing the entire database to NCGC on DVD's, free of charge.

III. Integration

A. Value Added Process

1. Benefit to the Service Center

The database is shipped from USGS and loaded on a NCGC DBMS server, which facilitates storage, access and backup as a DBMS database.
A process must be run to assign HUC_8 codes to the three layers. HUC_8 codes can be assigned, if a REACHCODE is present, by calc'ing the left 8 digits of the REACHCODE = to the HUC_8 value.
First, the process as described below is run on all 3 layers. Then the above described field calculation is run on all 3 layers. This ensures that wherever there IS a REACHCODE present, the HUC_8 number matches the REACHCODE information. Where there is no REACHCODE, the process as described below has already populated the HUC_8 values. Subsequent to this processing, all the themes can then be ordered by county, state or minimum bounding rectangle, and delivered by HUC_8.

2. Process Model

a. Flow Diagram

See Process Description

b. Process Description

1. Setting up SQL Server

a. Install Microsoft's SQL Server on a Windows operating system server and apply the latest service packs for Windows operating system and for the version of SQL Server. Instructions to install SQL server can be found at: http://msdn2.microsoft.com/en-us/library/aa197926(SQL.80).aspx – usually infrastructure support will install the software for the end users.

b. Create a new database in SQL Server 2000 using Enterprise Manager as shown in: http://msdn2.microsoft.com/en-us/library/aa176819(SQL.80).aspx.

c. Install ArcSDE on the same server using install DVD or CD media provided.

d. Perform post install as shown in: http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=29320.

e. Enter services name and port number in service file located in <Windows>\system32\drivers\etc\ folder.

2. Data Load Process Overview

a. The data is delivered as SDE layers on DVD's.

b. Three layers are extracted from the DVD's for serving on the Data Gateway:
NHDFLOWLINEHIGH
NHDAREAREACHHIGH
NHDWATERBODYHIGH

c. Since the data is served by HUC_8, this field is created and populated for each of the 3 layers.

d. The following instructions are for populating these layers with the identifiers (HUC_8).

3. Data Loading Steps

a. Copy desired layers (NHDFLOWLINEHIGH, NHDAREAREACHHIGH, and NHDWATERBODYHIGH) from DVD to a folder on a server that has Spatial Database Engine (SDE). Figure 1 shows the files copied to NHD folder in the c:\ drive.

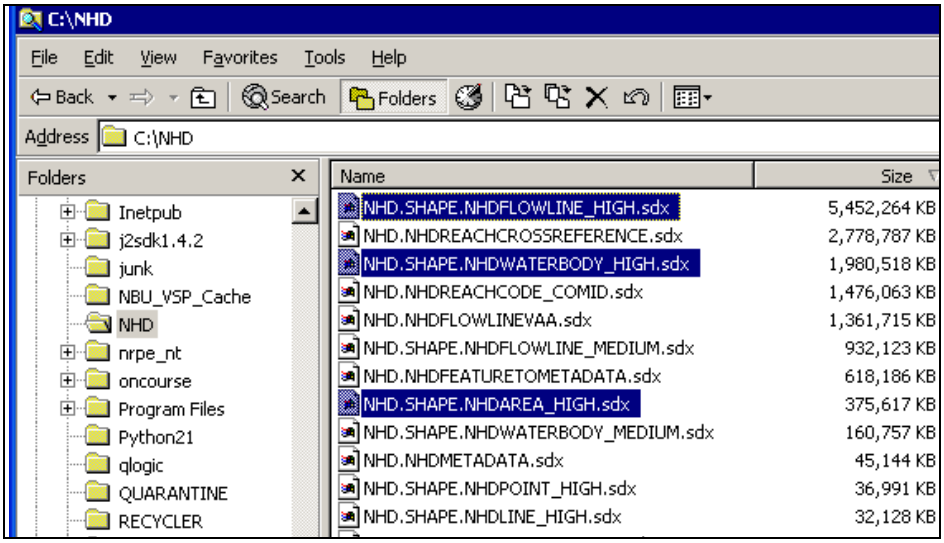

Figure 1 Three NHD layers to be loaded into SDE

b. Open a command prompt and navigate to the folder where the .SDX export files are present (in this case c:\NHD).

c. Run the SDE Import command as indicated with the options in Figure 2. The name of export file was made to be generic to indicate that this command needs to be run for each export file.

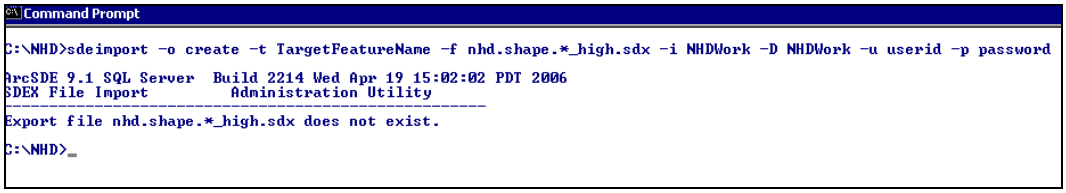

Figure 2 Sdeimport command to import it as a feature layer in SDE

d. Verify data has been imported into the feature layer by testing at random.

e. Download reference layer for populating NHD_HUC8 layers for HUC_8 codes.

4. Populate Feature Classes with Index values

Since the data is served by HUC_8, this field is created and populated for each of the 3 layers. See instructions below.

a. Add HUC_8 column to each layer. More information as to how to add field can be found at: http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Adding%20and%20deleting%20fields%20in%20shapefiles.

b. Ensure that target layers are not registered as version in Spatial Database Engine (SDE). If it is versioned, click on unregister as version.

c. Right click on each target layer, open attribute table, choose options button, and choose 'Add Field' command to add the identifier field HUC_8. Set the new column to be text and a width of at least 250 characters (Figure 3).

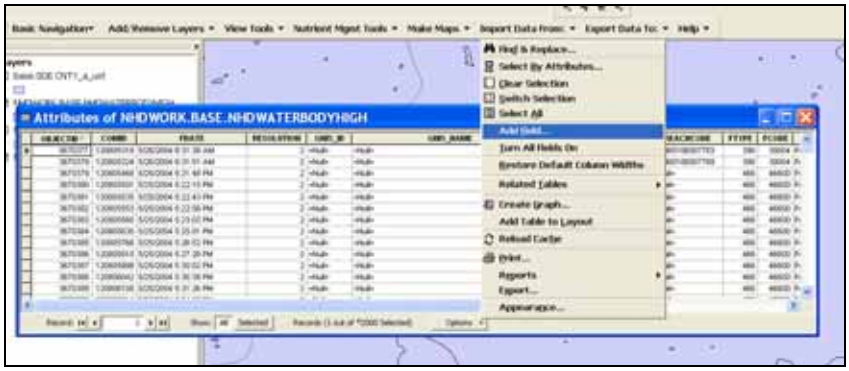

Figure 3 Adding identifier field to target layer (NHD water body layer)

d. To begin, click on 'IDS for GDW' command button to open 'Populate Layers with Identifiers 4 Data Gateway' gateway window (Figure).

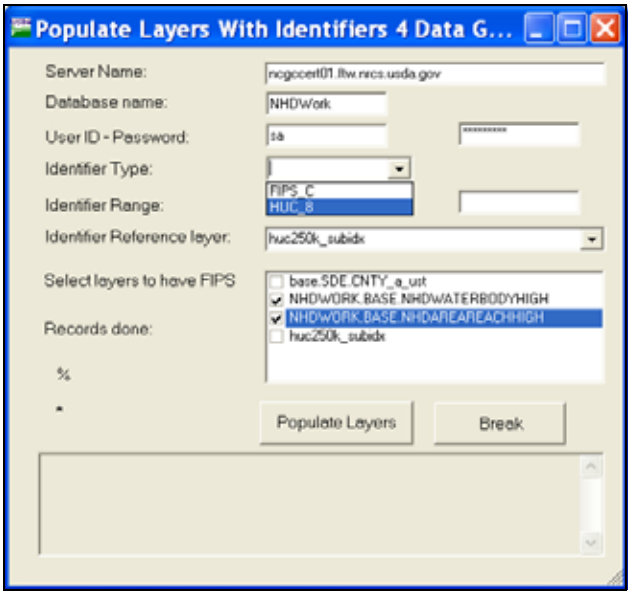

Figure 6 Populating the layers with HUC_8

e. Enter information in 'Populate Layers with Identifiers 4 Data Gateway' as indicated. Identifier range text boxes are provided to provide the ability to populate target layers into different batches of FIPS or HUC_8 codes.

f. Select the identifier reference layer from the drop down. Examples of identifier reference layer are county layer with FIPS code, NHD reference layer from USGS with HUC_8 codes and the like.

g. Select one or more target layers that require the identifiers to be populated. It is possible to populate more than one target layer in a sequential order with an identifier code of choice.

h. Click on 'Populate Layers' button to begin the population process. The process could take a long time depending on the number of records. Random counts on the approximate number of records populated per second were to the tune of 300 to 325. This could vary depending on network connectivity.

Following the above steps could result in almost all records getting populated with identifiers such as HUC8 codes.
For reasons beyond the control of the application developed, some records do not get populated. Number of records populated is usually less than 1 percent of the total number of records in target layer. Procedure to populate this small fraction of records is explained in the next section.

Manual Steps

Steps to populate HUC_8 code are explained in this section. The same procedure can be followed for populating target layers with other identifying codes.

a. Select features by drawing a rectangle around the records of interest. The features will be highlighted in the map control window (Figure 7).

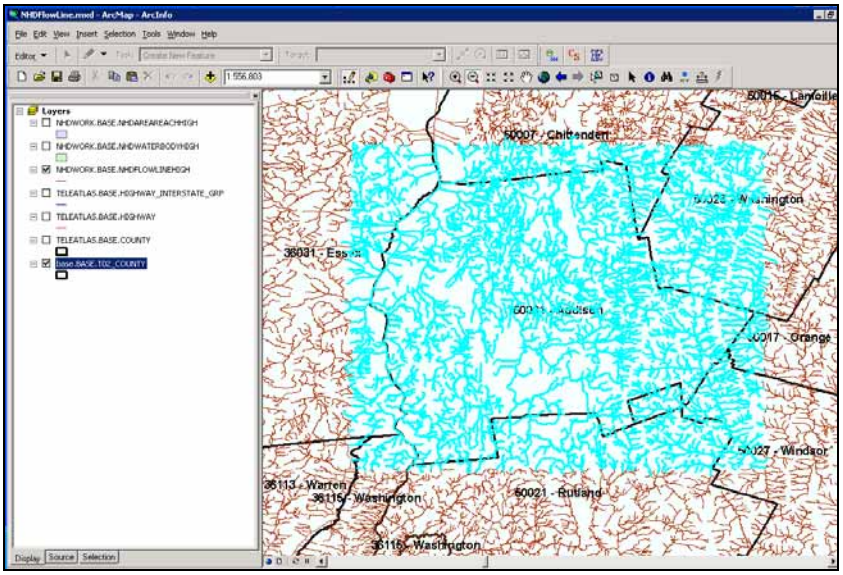

Figure 7 Highlighted features that need to be populated with HUC_8 code

b. Select the layer of interest, right click on its name and choose 'Open Attribute Window' option to examine its attributes (Figure 8).

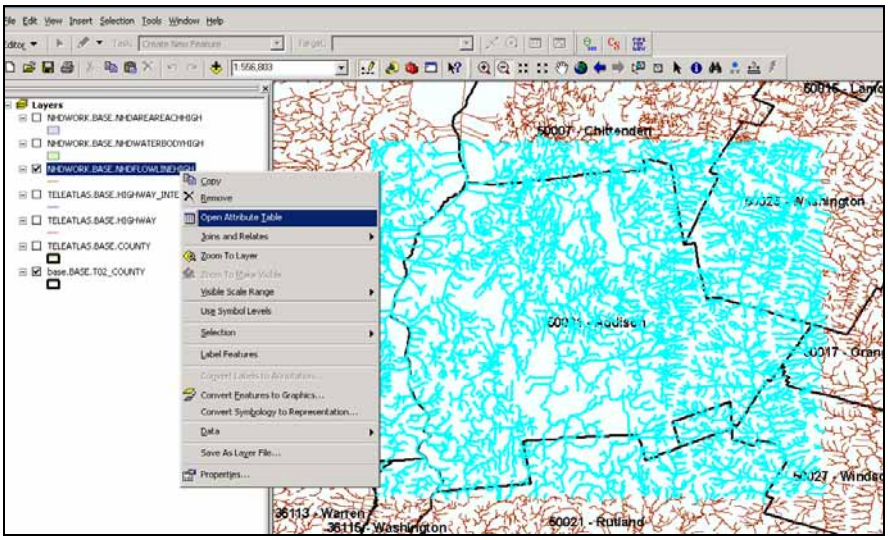

Figure 8 Opening attribute table to make changes

c. Right click on HUC_8 code column to identify records with HUC_8 code by sorting them in descending order (Figure 99).

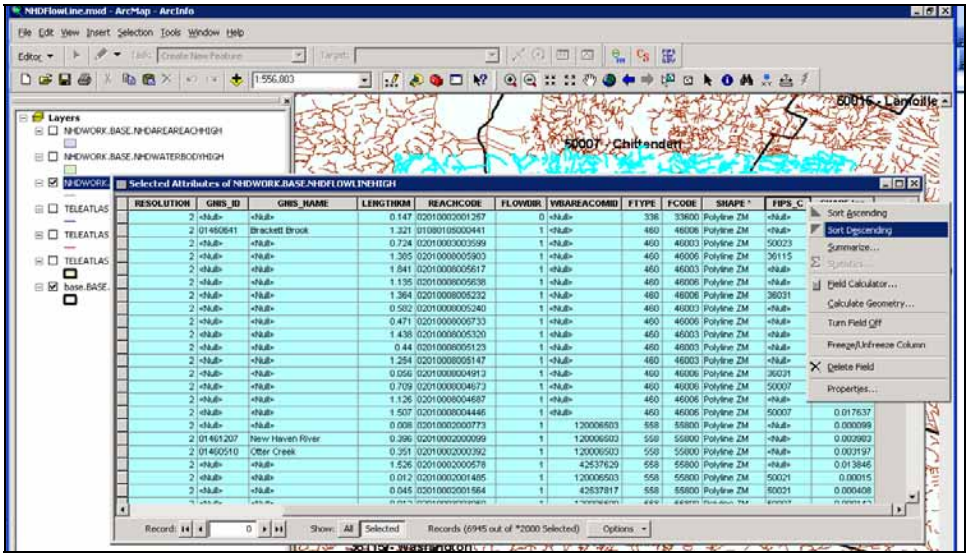

Figure 9 Sorting selected records to identify non-null values in FIPS column

d. Select all records (Figure 40) that have HUC_8 code (yellow colored records) and right click on the records and choose 'Unselect highlighted' (Figure 5).

Figure 40 Highlighting and selecting records that have HUC_8 values (yellow color records)

Figure 51 Unselecting highlighted records that have good HUC_8 values (yellow colored records)

e. Highlight the HUC_8 column and select field calculator option to insert appropriate FIPS codes (Figure 62).

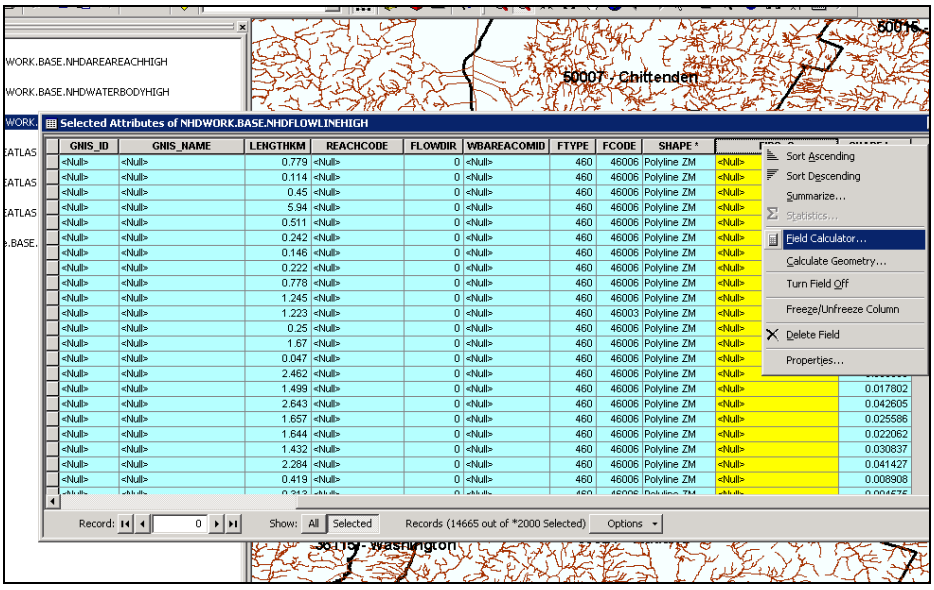

Figure 62 Setting HUC_8 value using field calculator

f. In the field calculator window, enter the appropriate HUC_8 code and click on ok to begin computing the values (Figure 73).

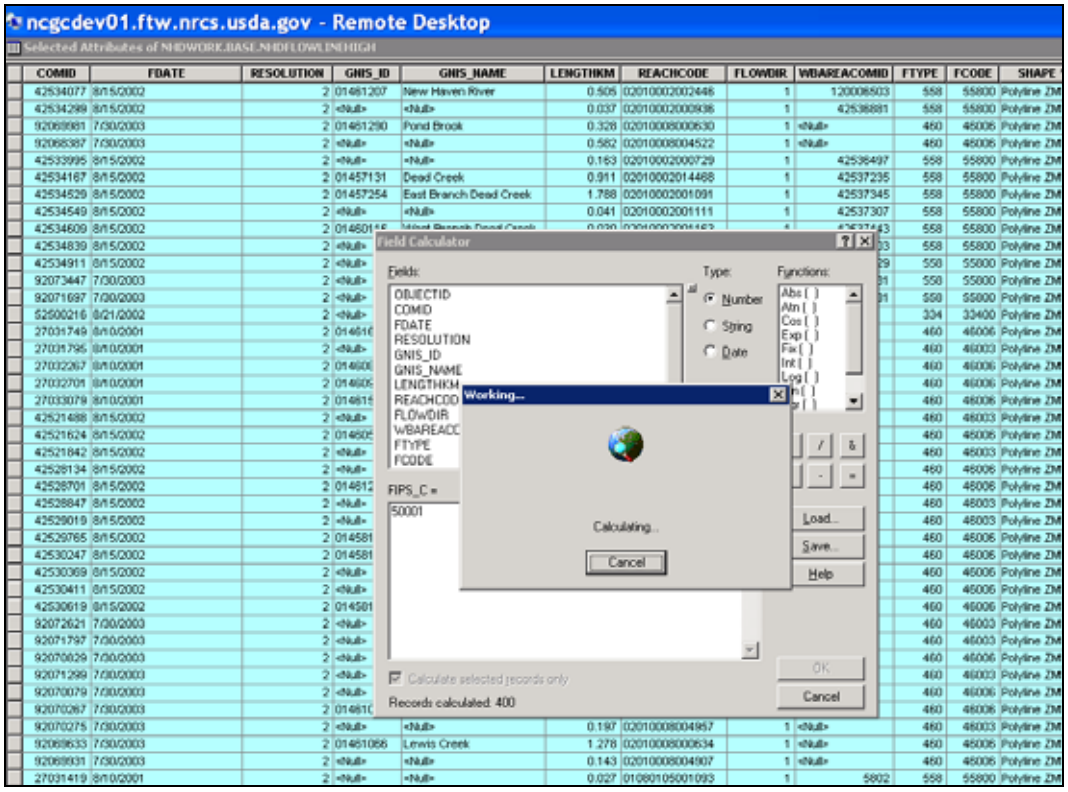

Figure 73 Calculating (populating) records with appropriate HUC_8 code

This step can be repeated to populate all the missing HUC_8 values.

5. Field Calculation on HUC_8 field.

a. Once the HUC_8 field is populated by the above method(s), if there is already a REACHCODE value present for the feature, we want the HUC_8 value to be equal to the left 8 digits of the REACHCODE. Calculate the left 8 digits of the REACHCODE = to the HUC_8 value for the NHDFLOWLINEHIGH and NHDWATERBODYHIGH layers. The NHDAREAREACHHIGH layer does not have a REACHCODE field.

6. Gateway Preparation and Loading

a. Write or revise the Product Description file (link from "Status Maps" page)

b. Generate shape files for each product with the DataServices_Driver. Normally the button 'All states using current product' is used to generate the shape files. Because NHD is so large and because NHD crosses county boundaries and would cause duplicate processing, this button is not used. Instead put in * for county and * for state. Also put in NHD24K for the Order ID and Item ID then press the 'Web' button.

c. Run CatalogFP_Maker for each product and generate the catalog shape files for each product.

d. Create the Status Maps (link from "Status Maps" page) for each product.

e. Run MakePreviews for each product to generate the preview images for step 2 of the gateway ordering process. There is no metadata for NHD24K so that button is not used. The DataServices_Driver generates shapes for streams, area reach and water bodies. CatalogFP_Maker totals the size of all three shape files. However, it puts the stream map name in the record for previews. Once CatalogFP_Maker is complete and prior to running MakePreviews all the water body and area reach maps should be removed from the destination directory so that MakePreviews doesn't waste time make .jpg files that will never be used.

f. Notify gateway Fort Collins team to load the catalogs, status maps and news.

3. Technical Issues
   a. Tiling

      Variable as needed during an extract.

   b. Compression

      LZ77

   c. Scale

      1:24,000

   d. Tonal Matching

      None

   e. Edge-matching

      None

4. Quality Control
   a. Procedures

      Visual quality check

   b. Acceptance Criteria

      Visual quality check

5. Data Steward
   a. Name and Organization

      Currently, the data steward for the integrated data is:
      National Cartography and Geospatial Center
      Natural Resources Conservation Service
      US Department of Agriculture
      501 Felix Street, Building 23
      P. O. Box 6567
      Fort Worth, Texas 76115-0567 USA

   b. Responsibilities

      Storage and access of the data.

B. Integrated Data Structure

1. Geospatial Data Format
   a. Format (raster, vector, etc.)

      ArcSDE Data Layers in Oracle Database

   b. Format Name

      ESRI SDE raster

   c. Data Extent

Same as source data

d. Horizontal and Vertical Resolution

Same as source data

e. Absolute Horizontal and Vertical Accuracy

Same as source data

f. Nominal Scale

1:24,000

g. Horizontal and Vertical Datum

Same as source data

h. Projection

Geographic

i. Coordinate Units

Degrees

j. Symbology

None

2. Attribute Data Format

a. Format Name

N/A Raster data

b. Database Size

xxxGB

3. Data Model

a. Geospatial Data Structure

ESRI SDE

b. Attribute Data Structure

N/A raster data

c. Database Table Definition

N/A raster data

d. Data Relationship Definition

N/A raster data

e. Data Dictionary

N/A raster data

C. Resource Requirements

1. Hardware and Software

This is unknown at this time.

2. Staffing

This is unknown at this time.

D. Integration Cost

1. Hardware and Software

This is unknown at this time.

2. Staffing

This is unknown at this time.

IV. Delivery

A. Specifications

1. Directory Structure

a. Folder Theme Data is Stored In

F:/geodata

2. File Naming Convention

http://www.itc.nrcs.usda.gov/scdm/scgdm.htm

a. List of Theme Files and The File Naming Convention

nhd24kst_l_<xxxxxxxx>	1:24,000 USGS/EPA National Hydrography Dataset **streams** line data by 8 digit sub basin
nhd24kar_a_<xxxxxxxx>	1:24,000 USGS/EPA National Hydrography Dataset polygon **area reach** data by 8 digit sub basin
nhd24kwb_a_<xxxxxxxx>	1:24,000 USGS/EPA National Hydrography Dataset polygon **water body** data by 8 digit sub basin

B. User Information

1. Accuracy Assessment

a. Alignment with Other Theme Geospatial Data

The data is captured at scales of 1:24,000. This elevation data should be considered not sufficiently detailed for the purpose of analysis at large scales. Alignment with the other data layers will not be perfect due to the fact that the data is captured at different scales and at different dates from other data.

b. Content

This elevation data should be considered not sufficiently detailed for the purpose of analysis at large scales.

2. Appropriate Uses of the Geospatial Data

a. Display Scale

The original data source scale or smaller, usually 1:24,000.

b. Plot Scale

The original data source scale or smaller, usually 1:24,000.

c. Area Calculations

Area Calculations are as accurate as the source data and capture scale and the algorithm used by ESRI software.

d. Decision Making

The data is as accurate as the source data and capture scale and the algorithm used by ESRI software

C. Maintenance and Updating

1. Recommendations and Guidelines

a. Original data location and structure

The integrated database is at NCGC and the data is delivered to the Service Center.

b. Update Cycle

The updates from the USGS should be applied periodically. The current thought is quarterly would be sufficient.

c. Availability

When the source data is updated.

d. Change Control

This is to be determined.